A Study of Covid-19 Outbreak Data Analysis using Machine Learning Methods: A Case Study of India

Ramjeet Singh Yadav
Department of Business Management and Entrepreneurship,
Dr. RammanoharLohia Avadh University
Ayodhya- 224001, Uttar Pradesh, India
ramjeetsinghy@gmail.com

***Abstract*—This study aims to use machine learning methods to analyze the Covid-19 outbreak data in India and provide insights for decision-making. The methods used include regression analysis and Susceptible, Infectious, and Recover/Removedmodel. Here we have analyzed the data of Covid-19 of India using different models of regression analysis. Data were collected from the website of health ministry of India from 2 March 2020 to 12 April 2020. Apart from this, the Covid-19 dataset as on 7 May 2020 has been used for the initial calculation of SIR model. The results of the analysis can help in forecasting the spread of the virus, identifying high-risk areas, prioritizing control measures, and early diagnosis and treatment. The use of machine learning methods in the analysis of Covid-19 outbreak data can provide valuable insights for effective control and mitigation of its impact.**

Keyword: Covid-19, Machine Learning, SIR Model, Reproductive Number, Infection Rate

I. INTRODUCTION

Covid-19 outbreak is a very fast contagious disease. The coronavirus disease is a spreadable infection that has spread in the form of a terrible disease all over the world today[1-2]. The Covid-19 outbreak is a severe infectious disease similar to the Spanish flu. The Spanish flu was a severe pandemic spread throughout the world in between 1918 to 1920 [3-5]. About 1, 75,000 people died in India alone. According to the World Health Organization, on 23-06-2021, a total of 178,701,170 cases of Covid-19 were registered and a total of 3,877,316 deaths occurred. So, here we can say that the Covid-19 pandemic is spreading very fast from 1st December 2020 till now. This disease has come twice in India. The first phase was from January 2020 to March 2020 and the second phase was from April 2021 to June 2021. The highest numbers of cases of Covid-19 pandemic were found in India, USA, Brazil and UK. The Covid-19 pandemic being a dreaded infectious disease spreads very rapidly from person to person. To date, more than 500 variants of this disease have been found. Some of these viruses are highly contagious and some are less contagious. The Covid-19 outbreak is a member of the SARS-CoV-2 family. SARS-CoV-2 is also called severe acute respiratory syndrome Coronavirus-2 [6].

First of all, China invented the corona virus in 2002, after which Saudi Arabia also studied a lot about the corona virus in 2012 [7]. There are symptoms of Covid-19 outbreak like fever, cough, shortness of breath or difficulty breathing, chills or repeated shaking with chills, fatigue, muscle pain, headache, sore throat, new loss of, smell or taste, congestion or runny nose, nausea or vomiting and diarrhea [8].] It is responsible for life-threatening diseases such as Middle East Respiratory Syndrome (MERS) and severe acute respiratory syndrome (SARS) [8]. The first case of Covid-19 outbreak was found in Wuhan city of China [2]. According to news channels and some newspaper reports, the first case of Covid-19 outbreak was found in the first week of November 2020 but China kept hiding this fact. This feat of China has to be paid by the countries of the whole world. The study by Jiang and colleagues found that the mortality rate for the Covid-19 outbreak is near 7.5% [9]. He also found in his study that the death rate for people above 80 years of age was close to 14.8%, while the death rate for people aged 70 to 80 years was near 8.0% [9]. This study also found that the people, who are above 50 years of age and those who are already suffering from underlying diseases like diabetes, Parkinson's disease and heart, etc., will prove to be very fatal for Covid-19 outbreak [9]. Symptoms of the disease start showing between 2 to 14 days in humans suffering from Covid-19 [10]. In the early stages of the Covid-19 outbreak, it was found by scientists that this virus does not spread in the air. But in a study by Neeltje van Doremalen and his colleagues, it was found that the virus of Covid-19 outbreak stays in the air for 2 to 3 hours [11].

Today, many scientists and data analysts of the world are doing data analysis in different domains with the help of machine learning methods and also getting very good results.Over the years, machine learning has proven to be capable in analyzing and forecasting healthcare data [12-

15].Elaine and associates have delivered a methodical way to forecast the dynamics of influenza pandemics using various models of machine learning [16]. They have also looked at a lot of research papers on prediction, such as regression analysis and mass action. These people also studied most of the research papers of various models (Bayesian network, SEIR model, ARIMA forecasting, prediction rules, regression models, deterministic mass action models, prediction rules and deterministic models) of machine learning which are helpful in forecasting the epidemic [16]. The SEIR and ARIMA epidemic models are also known as "susceptible (S), exposed (E), infected (I), and resistant (R) and Auto-Regressive Integrated Moving Average" models. The research done by various researchers on the data of Covid-19 outbreak shows that there has very limited study and exploratory analysis on the data [17-18]. Wu and his colleagues have reported a study that till date no country in the world has developed any medicine to completely cure the Covid-19 outbreak and to reduce its effect [19]. These people had talked about no vaccination a year ago, but today many countries of the world are going on very fast in their work. The analysis of data mining researchers and data scientists has been done by collecting technology and related data to explain the characteristics of Covid-19 and the role of this virus [20-23].So, the above review shows how we can get rid of Covid-19 outbreak type diseases in future by the methods of machine learning. This type of study also suggests that the Covid-19 outbreak will be able to properly treat diseases like coronavirus, development of vaccines, and strengthen the infrastructure of the hospital.

It has been found by various researchers that initially, an exponential curve was found in most of the epidemics and gradually its curve becomes flat [24]. In this present paper, it has been considered that every person in India follows a face mask, social distancing and lockdown. The nationwide lockdown for 21 days was announced by the "Prime Minister of India, Shri Narendra Modi" on 24 March, 2020. Studies by various researchers suggest that a 21 days lockdown is insufficient to prevent the Covid-19 outbreak [25]. Even after this, the cases of Covid-19 in India did not decrease at that time because most of the people in India do not listen to anyone. The lockdown was also announced in the second phase of Covid-19 in India, but after that daily belonging coronavirus came to around 400000.

II. MATERIALS AND METHODS

Hence, the above study suggests that more studies are needed for more evidence of the Covid-19 outbreak. Hence, we have proposed two methods of machine learning for analyzing the dataset of Covid-19 outbreak in India. These two machine learning methods are: Regression analysis and SIR Model. Both regression analysis and SIR model are machine learning based models which are very efficiently analyzing data in any domain. Both regression analysis and SIR models are machine learning based models that analyze data from any domain very efficiently. Regression analysis is used in classification and future prediction. The SIR model is an epidemic model which is used to analyze data of any infectious disease. In this proposed study we have used various types of regression models such as "exponential, quadratic, third degree, fourth degree, fifth degree and sixth degree polynomials" for analyzing the Covid-19 outbreak datasets. Apart from this, we have also used the SIR model for predicting the peak and end of the coronavirus disease of India.

A. Regression Analysis

Regression analysis is a statistical method for solving classification problems and prediction for future use. It identifies the link between two or more variables. Regression analysis deals with the problem through the approximation of the output variable which is based on the input variable.Here let us assume the equation of a straight line which is connected through the X as input and Y as the output. The following is the equation for the linear relationship between X and Y:

$$Y = bX + a \tag{1}$$

The intercept on the y-axis is represented by a, while the slope of the line is represented by b. The least square approach, artificial neural networks, evolutionary algorithms, and other relevant learning methods can all be used to learn these parameters.

B. SIR Model

We applied the Susceptible, Infectious, and Recover/Removed (SIR) epidemic model to the data analysis of the Covid-19 outbreak in this proposed research work. Kermack and McKendrick, two outstanding scientists, devised the SIR epidemic model in 1927 [26]. The proposed SIR model also predicts future maximum growth, peak and end of the epidemic. TheSIR model is a family of

differential equations. These differential equations make up the suggested SIR model which is given below:

$$S'(t) = -rSI \quad (2)$$

$$I'(t) = rSI - aI \quad (3)$$

$$R'(t) = aI \quad (4)$$

The infection rate and recovery rate of the epidemic are represented by the parameters r and a.We believed that the population of India will remain constant during this investigation in our proposed study. Here we have taken a tested dataset of Covid-19 outbreak on 7th April 2020. We have proposed an adaptation of the SIR model that does virus development for the Covid-19 outbreak in India. Fig. 1 depicts the description of my suggested SIR model.

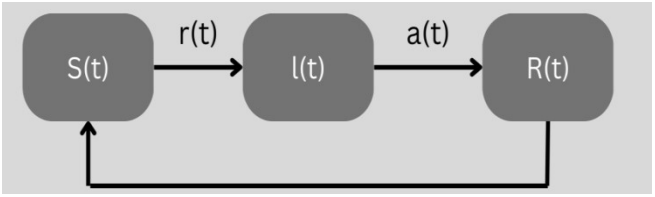

Fig. 1. Proposed SIR model

According to numerous studies, the incubation time of the coronavirus disease in India is roughly 14 days [12]. The differential equations (2), (3) and (4) are also can be written as follows:

$$\frac{dS}{dt} = -rSI \quad (5)$$

$$\frac{dI}{dt} = rSI - aI \quad (6)$$

$$\frac{dR}{dt} = aI \quad (7)$$

Adding above three differential equations (5), (6) and (7), we the following useful expression:

$$dS + dI + dR = 0 \quad (8)$$

$$S' + I' + R' = N \quad (9)$$

The constant N is the constant of integration. This constant of integration measures the population size at the beginning and ending of the epidemic. The constant of integration has been used at all stages of Covid-19 outbreak.In this proposed paper, we have taken the initial values of proposed SIR model which are given below:

$$S(0) = S_0, I(0) = I_0 \; and \; R(0) = R_0 \quad (10)$$

Here we have calculated the recovered population of Covid-19 outbreak from the following equation:

$$R(t) = N - (S(t) + I(t)) \quad (11)$$

We have discovered some useful observations about the Covid-19 outbreak in this suggested paper, which are listed below:

Case 1: If $S(0) < \frac{a}{r}$. This indicates that the infection of Covid-19 outbreak will be epidemic.

Case 2: If $S(0) > \frac{a}{r}$. This indicates that the infection of Covid-19 outbreak will decrease. After some time, it will be zero.

Reproductive Number:The reproductive number is defined by following below expression:

$$R_n = \frac{S_0 r}{a} \quad (12)$$

The two Covid-19 epidemic cases with the highest reproductive numbers are listed below:

Case-1: If $R_n < 1$: this means that the Covid-19 outbreak will be brought to a halt.

Case-2: If $R_n > 1$: this indicates that the Covid-19 outbreak is still ongoing.

C. Phase Plan of SIR Model

The proposed SIR model requires the solution of differential equations. For each day, we counted the total number of examined populations as susceptible cases and infected populations. Let K is the population size of susceptible cases. This population size *(K)* is approximately equal to the initial population(S_0). The calculation is given below:

$$S_0 = K, I_0 = 0 \; \alpha\nu\delta \; R_n = \frac{rK}{a} \quad (13)$$

If $I(t) = 0$ as $t \to \infty$ and $S_0 < \frac{a}{r}$ then $V(S_0, I_0) = V(S_0)$

$$K - \frac{a}{r} ln \; ln \; (S_0) \; = S_\infty - \frac{a}{r} ln \; (S_\infty)$$

If the infective case will be zero then S_∞ is the susceptible population. By simplifying the above expression then we have gotten the following expression:

$$\frac{r}{a} = \frac{ln\left[\frac{S_0}{S_\infty}\right]}{K - S_\infty} \quad (14)$$

The inequality $0 < S_\infty < K$ indicates that the past population of the country outflows to Covid-19 outbreak. The estimation of parameters *r* and *a* is time consuming and very difficult due to these parameters depends on study of disease, social and behavioral factors of the country. Serological studies conducted before and during the Covid-19 outbreak are used to estimate S_∞ and S_0.The following formula is used to calculate the basic reproductive number:

$$R_n = \frac{rK}{a}$$

The calculation of peak of Covid-19 outbreak is as follows:

$$I_{max} = I_0 + S - \frac{a}{r} + -\frac{a}{r} ln(S_0) + \frac{a}{r} ln\left(\frac{a}{r}\right) \quad (15)$$

Many numerical methods, including Runge-Kutta and Euler methods, can be used to solve the differential equation of the proposed SIR model. For solving SIR model based differential equations, we employed the Euler method. We used the MATLAB software to solve the differential equation using the above initial conditions values of S_0, I_0, R_0, *a* and *r* in this suggested study. Table 1 shows the experimental results of the SIR model. The Euler method was used to perform numerical calculations and data analysis for the Covid-19 epidemic in India.The following is a description of Euler's method:

$$\frac{dy}{dx} = f(x, y)$$

The following expression gives the solution to above differential equation:

$$y_{n+1} = y_n + \Delta t f(x_n, y_n) \quad (16)$$

The function $f(x_n, y_n)$ and a small quantity (Δt) represent the slope of curve and step size in the time domain respectively. In this proposed paper, there is a need to calculate quantities S, I and R. Thus, the above differential equation can be written as follows:

$$S(n+1) = S(n) - rI(n)S(n)\Delta t \quad (17)$$

$$I(n+1) = I(n) + [rI(n)S(n) - aI(n)]\Delta t \quad (18)$$

$$R(n+1) = R(n) + aI(n)\Delta t \quad (19)$$

III. EXPERIMENTAL RESULT AND DISCUSSIONS

A. Experimental Results and Discussion (Regression Analysis)

In this paper, we have used a regression analysis method to forecast the number of Covid-19 patients in the next one or two weeks. For experimental purposes, we have taken complete datasets of India's Covid-19 epidemic from the health ministry of India and other sources like Kaggle and World Health Organization (WHO) website [1]. The data retrieved from the website of the health ministry of India between 2 March 2020 and 12 April 2020 has been analyzed using different models of regression analysis. We have done all the experimental work by using MATLAB software. After doing the experimental work, we have got the various equations based on the regression analysis. These equations are given below:

Exponential polynomial: $Y = 18.74 * e^{0.14x}$

Quadratic Polynomial: $Y = 8.157X^2 - 214.76X + 1013.4$

Third Degree polynomial: $Y = 144.48X^3 + 597.77X^2 + 865.66X + 618.82$

Fourth Degree polynomial: $Y = 144.48X^4 + 597.77X^3 + 865.66X^2 + 618.82X + 272.43$

Fifth Degree polynomial: $Y = -52.17 * X^5 + 144.48 * X^4 + 766.97 * X^3 + 865.66 * X^2 + 512.09 * X + 272.43$

Sixth Degree polynomial:$Y = -90.92 * X^6 \pm 52.17 * X^5 + 505.65 * X^4 + 766.97 * X^3 + 515.23 * X^2 + 513.09 * X + 320.96.$

The table 1 depict the calculated values of sum of square error (SSE), R^2, degree of freedom error and adjusted R^2. These parameters reflect the validity of the regression model.

TABLE I. CALCULATED VALUES OF SUM OF SQUARE ERROR, R^2, DEGREE OF FREEDOM ERROR AND ADJUSTED R^2

S.N.	Proposed Methods (Regression Analysis)	Sum of Square Error	R^2	Degree of Freedom Error	Adjusted R^2
1	Exponential Polynomial	845000	0.9951	40	0.9950
2	Quadratic Polynomial	9209100	0.9463	39	0.9436
3	Third Degree Polynomial	6466400	0.9962	38	0.9959
4	Fourth Degree	2777700	0.9984	37	0.9982

	Polynomial				
5	fifth Degree Polynomial	2426900	0.9986	36	0.9984
6	Sixth Degree Polynomial	1656800	0.9990	35	0.9989

The results of fitted sixth degree polynomial regression method are shown in fig.2 to the confirmed cases of Covid-19 data set. The SSE of various regression analysis models are shown in table 1. The SSE of various proposed regression models play a significant character in the data analysis of any outbreak. The goodness of fit is measured by the SSE of the regression line. The various regression models give the satisfactory results while quadratic polynomials give the unsatisfactory result to the data analysis of Covid-19 outbreak (table 1).In this proposed paper we have also determined the value of R^2 and adjusted R^2 to the proposed various regression models for the measure of goodness of the models. The calculated values of Sum of Square Errors (SSR), R^2, Degree of Freedom for Error (DFE) and adjusted R^2 are shown in table 3.3. The value of Sum of Squared Errors (SSR), R^2, Degree of Freedom for Error (DFE) and adjusted R^2, the sixth-degree polynomial is the lowest in comparison to other proposed regression models. Therefore, here we have found out that the sixth-degree polynomial gives the better result to other regression models.

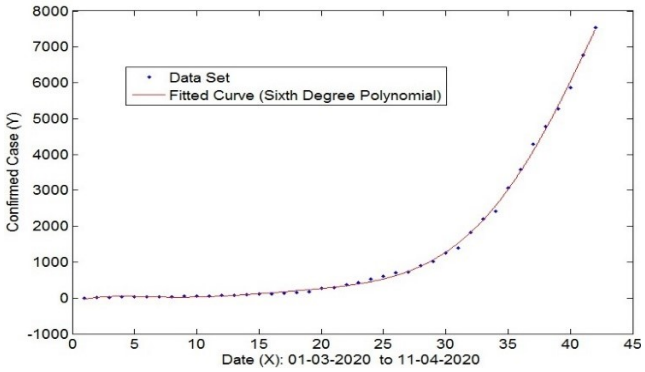

Fig. 2. Fitted Curve of Sixth Degree Polynomial

TABLE II. EXPERIMENTAL VALUE OF 6TH DEGREE POLYNOMIAL

Date	Number of Covid-19 Case (Desired Value)	Experimental value of 6th degree polynomial
02-03-2020	4	-32.70
03-03-2020	6	18.90
04-03-2020	7	39.91
05-03-2020	29	49.51
06-03-2020	31	40.45
07-03-2020	32	35.70
08-03-2020	35	30.45
09-03-2020	40	24.60
10-03-2020	47	24.62
11-03-2020	59	45.54
12-03-2020	61	42.80
13-03-2020	75	54.50
14-03-2020	92	73.60
04-04-2020	2415	2485.00
05-04-2020	3073	3085.00
06-04-2020	3578	3548.00
07-04-2020	4282	4185.00
08-04-2020	4790	4810.00
09-04-2020	5275	5295.00
10-04-2020	5866	5977.00
11-04-2020	6762	6852.00
12-04-2020	7530	7544.00

The experimental results of various regression models are shown in table 2 and 3 for the training dataset and testing dataset, respectively. Here we have taken Covid-19 data set in between 12th –April-2020 to 19th- April-2020 for testing purposes. The comparison of actual result (confirmed case) and predicted result (result of sixth degree polynomial) are shown in fig. 3. Here we have found that the experimental result of the 6th degree polynomial is very close to the actual result (number of confirmed cases).

B. Expiremental Results and Discussion (SIR Model)

In this paper, for experimental point of view we have used Covid-19 dataset from India on 7th May 2020. The simulated result of SIR mode is shown in figure 4. At the initial level, I have taken S_0 as the tested population, I_0 as

the total number infected cases and R_0 as the total number of recovered or removed cases on the 7th May 2020 for Covid-19 outbreak.

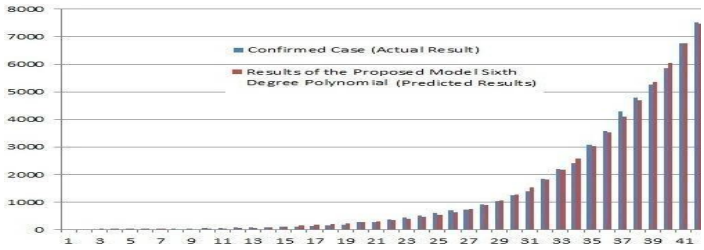

Fig. 3. Comparison of Actual Result and the Predicted Results

TABLE III. TESTING DATA SET AND EXPERIMENTAL VALUE OF PROPOSED REGRESSION ANALYSIS METHODS (6TH DEGREE POLYNOMIAL)

Date	Number of Covid-19 Case (Desired Value)	Experimental value of 6th degree polynomial
13-04-2020	8455	8405
14-04-2020	9433	9489
15-04-2020	10815	9960
16-04-2020	11933	11600
17-04-2020	12759	12953
18-04-2020	13835	13975
19-04-2020	17792	17490

In other word, Total Number susceptible population $(S_0) = 13.57410$, total number of infected case $(I_0) = 0.56340$ and total number of recovered/removed cases $(R_0) = 0.18430$. All S_0, I_0 and R_0 have been taken as in lac.The greatest number of infection cases during the Covid-19 epidemic in India is estimated to be on May 26, 2020, according to fig.7 and table 4. The peak of the Covid-19 outbreak in India will occur on this date. The recovery rate and infection have been calculated using the following expression:

$$r = \frac{Infected\ Population\ (p)}{Susceptible\ Population\ (q)} = \frac{42836}{1357413} \Rightarrow r = 0.03156$$

Fourteen days is the development time of coronavirus disease. It is also known as incubation time of coronavirus disease. It can be expressed as the: $\frac{1}{a} = 14 = 0.0714$ and the time interval is Δt: 0.1407. The Euler's method has been used to solve three differential equation of SIR model and analyzed the Covid-19 epidemic data in India. At the initial stage of Euler's method, we have calculated the generational values of Susceptible (S_0), Infectives (I_0) and Recover or Remove (R_0) population using the values $\Delta t = 0.1507, r = 0.03156, S_0 = 14.68521, I_0 = 0.57451, R_0 = 0.19532, a = 0.0714$. For the calculation purpose, we have used the equation (1), (2) and (3). The generational values calculations are given below:$S_1 = 14.63562$, similarly, we can calculate the infective and recovery/removal case. The calculations are: $I_1 = 0.61801$ and $R_1 = 0.25712$. After multiplying 100000 in above calculated values of S_1, I_1 and R_1 we get the actual value of susceptible, infectives and recovery/removal cases of Covid-19. The actual values of susceptible, infective and recover/removal case of Covid-19 are given below:$S_1 = 1463562, I_1 = 61801, R_1 = 25712$. These values have also been matched from figure 4. This is the iterative process. The iteration will stop when the process reaches the convergence criteria. The greatest number of infection cases during the Covid-19 epidemic in India is estimated to be on May 26, 2020 (figure 4). The peak of the Covid-19 outbreak in India will occur on May 26, 2020. The maximum infective Covid-19 cases can be also calculated with help below equation. The calculations are: $\frac{r}{a} = \frac{ln\left[\frac{S_0}{S_\infty}\right]}{K - S_\infty}$, $r = 0.03156, K = S_0 = 14.68521$, $I_0 = 0.57451, R_0 = 0.19532, a = 0.0714$

Putting the values S_0 and S_∞ in above equation, we get

$\frac{r}{a} = \frac{ln\left[\frac{14.68521}{0.027}\right]}{14.68521 - 0.0270}$, $\frac{r}{a} = \frac{ln[543.897]}{14.65821}$, $\Rightarrow \frac{r}{a} = \frac{6.29876}{14.65821} = 0.4297$, $\Rightarrow \frac{a}{r} = \frac{1}{0.4297} = 2.3272, I_{max} = I_0 + S_0 - \frac{a}{r} + -\frac{a}{r} ln(S_0) + \frac{a}{r} ln\left(\frac{a}{r}\right), I_{max} = 8.99802$

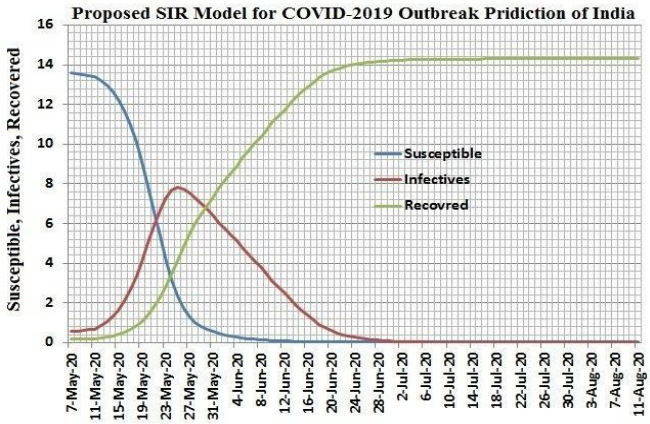

Fig. 4. The infective and recovered Covid-19 cases

In this proposed study, we have taken 100000 as the normalization factor. This normalization factor converts the training and test data between the range of [0, 1] by dividing each feature value by 100000. Here we have multiplied by 100000 in above calculated value of maximum number of infected cases (I_{max}), calculated as $I_{max} = 8.99802 \times 100000 = 899802$. This is the peak value of Covid-19 of India. Apart from this, we have observed that the Covid-19 outbreak will decline constantly. Here we have also calculated the reproductive number during the Covid-19 outbreak. In this proposed paper we have calculated the reproductive number of initial, peak and the end phase. The calculation of reproductive number of India at the initial phase, peak phase and end phase are as follows:

Case 1: Initial Phase: $R_n = \frac{rS_0}{a} = \frac{0.03156 \times 14.68521}{0.0714} = 6$

Case 2: Peak Phase: $R_n = 14.68521 * 0.4297 = 6.3102 \left(\frac{r}{a} = 0.4297, S_0 = 14.68521\right)$

Case-3: End Phase: $R_n = 0.027 * 0.4297 = 0.0116019 \left(\frac{r}{a} = 0.4297\right)$

From the above calculated value of reproductive number, we have been found out that the reproductive number is greater than one the outbreak of Covid-19 continuously increasing at peak of epidemic (maximum level epidemic) (Case 1 and Case 2) and the calculated value of reproductive number is less than one then the outbreak of Covid-19 in India will be stop (Case 3). At present there is no definite reproductive number available for Covid-19 as it is a new virus. The experimental value of SIR model and observations are: (1) the peak of outbreak Covid-19) is 26th-May-2020 to 7th-April-2020. After this, the outbreak of Covid-2019 will continue gradually. (2) The Covid-19 pandemic will continue gradually till the first and second week of August-2020, after which the outbreak of the pandemic will end. (3) The value of reproductive number at the initial phase and peak phase of Covid-19 outbreak is approximately 6.

IV. CONCLUSION

For the analysis of the coronavirus outbreak in India, we have presented machine learning methodologies in this proposed paper. The machine learning based models give better results for the prediction of the next one week or two weeks. From the previous experimental results and discussion of regression analysis, we have found out that the proposed machine learning based models predict the best results for the next ten days. Apart from this the experimental results of regression analysis-based models are very close to the confirmed case regarding the training datasets. Table 1 show that the calculated value of SSE of sixth degree polynomials is lower than the other methods. Hence, sixth degree polynomials regression gives better results to both datasets (training and testing datasets). In this proposed study, we have found that the SIR model predicts the peak and end of coronavirus in India is 25th May 2020 or by the end of May 2020. The coronavirus outbreak in India is expected to stop in the first week of August 2020, according to this analysis. We discovered that the Covid-19 outbreak will begin slowly after the end of May 2020, and by the second week of August 2020, the outbreak will be nearing its end. From the observation and discussion of our proposed study, we have found that the machine learning based regression model will prove to be very useful for the Government of India and doctors in managing the coronavirus outbreak. The number of coronavirus cases will be automatically predicted using the proposed machine learning model on a weekly and bi-weekly basis. In the future we will develop SIER (susceptible, exposed, infected, and recovered) and gradient descent learning based regression analysis-based software for monitoring Covid-19 patients. The SIER epidemic model is an extended form of SIR model. This model will be very useful for Indian government and doctors in monitoring of Covid-19 patients.

REFERENCES

[1] World Health Organization,"Coronavirus disease 2019 (COVID-19): situation report," 2020,pp. 67-72.

[2] Stoecklin, B. Sibylle, P. Rolland, Y. Silue, A. Mailles, C. Campese, A. Simondon andM. Mechain,"First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures,"Euro Surveillance, 25, 2020,pp. 6:1-7.

[3] WHO, "Pandemic Influenza Risk Management WHO Interim Guidance," World Health Organization, 2021, pp. 19-23.

[4] P. Spreeuwenberg, M. Kroneman and J. Paget, "Reassessing the Global Mortality Burden of the 1918 Influenza Pandemic," American Journal of Epidemiology, Oxford University,Vol. 187 (12), 2018, pp. 2561–2567.

[5] M.S. Rosenwald, "History's deadliest pandemics. Ancient Rome to modern America," The Washington Post, 2020.

[6] C.C. Lai, T.T. Shih, W.C. Ko, H.I. Tang and P.R. Hsueh,"Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges,"International Journal of Antimicrobial Agents, Vol. 55 (3), 2020, pp. 1-9.

[7] J. Zheng,"SARS-CoV-2: An Emerging Coronavirus that Causes a Global Threat,"International Journal of Biological Sciences, Vol. 16, 2020, pp. 1678-1685.

[8] T. Struyf, J.J. Deeks, J. Dinnes, Y. Takwoingi, C. Davenport, M.M.G. Leeflang, R. Spijker, L. Hooft, D. Emperador, J. Domen, S.R.A. Horn and A.V.D. Bruel,"Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19," Cochrane Database of Systematic Reviews published by John Wiley and Sons, Ltd., 2021, 2. Art. No.: CD013665. DOI: 10.1002/14651858.CD013665.pub2.

[9] Z. Wu andJ.M. McGoogan,"Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention 2020.", 2020. DOI: 10.1001/jama.2020.2648

[10] P. Daniel, A.M. Oran, J. Eric andM.D. Topol, "Prevalence of Asymptomatic SARS-CoV-2 Infection: A Narrative Review,"Annals of Internal Medicine, Vol. 173(5), 2020, pp. 362-367. DOI: 10.7326/M20-3012.

[11] N.V. Doremalen, T. Bushmaker, D.H. Morris, M.G. Holbrook, A. Gamble, B.N. Williamson, A. Tamin, J.I. Harcourt, N.J. Thornburg, S.I. Gerber, J.O. Lloyd-Smith, L. Angeles, Bethesda, E.D. Wit and V.M. Munste, "Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1,"The New England Journal of Medicine, Vol. (382) 2020, pp.1564-1567: DOI: 10.1056/NEJMc2004973.

[12] F. Jiang, L. Deng, L. Zhang, Y. Cai, C.W. Cheung and Z. Xia,"Review of the clinical characteristics of coronavirus disease 2019 (COVID-19),"Journal of General Internal Medicine,Vol. 35, 2020, pp. 1-5. PMCID: PMC7088708 DOI: 10.1007/s11606-020-05762-w.

[13] Q.H. Ye, L.X. Qin,M. Forgues, P. He, J.W. Kim, A.C. Peng, R. Simon, Y. Li, A.I. Robles, Y. Chen and Z.C. Ma, "Predicting hepatitis B virus–positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning," Nature medicine, Vol. 9 (4), 2003, pp. 416-423.

[14] M.V. Mai and M. Krauthammer, "Controlling testing volume for respiratory viruses using machine learning and text mining,"AMIA Annual Symposium Proceedings, 2016, pp.1910-1919. PMID: 28269950.

[15] G. Purcaro, C.A. Rees, W.F. Wieland-Alter, M.J. Schneider, X. Wang, P.H. Stefanuto, P.F. Wright, R.I. Enelow and J.E. Hill,"Volatile fingerprinting of human respiratory viruses from cell culture," Journal of breath research, Vol. 12 (2) 2018, pp 1-15. 026015. PMCID: PMC5912890 DOI: 10.1088/1752-7163/aa9eef.

[16] E.O. Nsoesie, J.S. Brownstein, N. Ramakrishnan and M.V. Marathe, "A systematic review of studies on forecasting the dynamics of influenza outbreaks," Influenza and other respiratory viruses,Vol. 8 (3) 2014, pp. 309-316.

[17] B. Pirouz, S.S. Haghshenas, S.S. Haghshenas andP. Piro, "Investigating a Serious Challenge in the Sustainable Development Process: Analysis of Confirmed cases of COVID-19 Through a Binary Classification Using Artificial Intelligence and Regression Analysis," Sustainability, Vol. 12 (6), 2020, pp. 1-21.https://doi.org/10.3390/su12062427.

[18] G.D. More, M. Dunowska, E. Acke and N.J. Cave, "A serological survey of canine respiratory coronavirus in New Zealand," New Zealand Veterinary Journal, Vol. 68 (1), 2020, pp. 54-59.

[19] C. Wu, X Chen, Y. Cai, X. Zhou, S. Xu, H. Huang, L. Zhang, X. Zhou, C. Du,Y. Zhang and J. Song, "Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China," JAMA Internal Medicine, Vol. 180 (7), 2020, pp. 934-943. doi:10.1001/jamainternmed.2020.0994

[20] Deb, Soudeep andM.A. Majumdar, "A time series method to analyze incidence pattern and estimate reproduction number of COVID-19 2020," ARXiv preprint ARXiv: 2003, 2020, pp. 1-15. https://arxiv.org/pdf/2003.10655.pdf.

[21] S. Mandal, T. Bhatnagar, N. Arinaminpathy, A. Agarwal, A. Chowdhury, M. Murhekar, R.R. Gangakhedkar, S. Sarkar, "Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: A mathematical model-based approach," The Indian journal of medical research, Vol. 151 (2), 2020, pp.190-199.

[22] E. Dong, H. Du and Gardner, "An interactive web-based dashboard to track COVID-19 in real time,"The Lancet Infectious Diseases, Vol. 20, 2020, pp. 533-534. https://doi.org/10.1016/S1473-3099 (20)30120-1.

[23] R. Gupta and S.K. Pal, "Trend Analysis and forecasting of COVID-19 outbreak in India," Preprint archive in medRxiv 2020. https://doi.org/10.1101/2020.03.26.20044511.

[24] J. Ma, J. Dushoff, B.M. Bolker and D.J. Earn, "Estimating initial epidemic growth rates," Bulletin of mathematical biology,Vol. 76 (1), 2014, pp. 245–260.

[25] R. Singh andR Adhikari,"Age-structured impact of social distancing on the COVID-19 epidemic in India," ARXiv 2020. https://doi.org/10.48550/arXiv.2003.12055.