Journal of Interdisciplinary Mathematics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjim20

Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario

Amit Kumar Gupta , Vijander Singh , Priya Mathur & Carlos M. Travieso-Gonzalez

To cite this article: Amit Kumar Gupta , Vijander Singh , Priya Mathur & Carlos M. Travieso-Gonzalez (2020): Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario, Journal of Interdisciplinary Mathematics, DOI: 10.1080/09720502.2020.1833458

To link to this article: https://doi.org/10.1080/09720502.2020.1833458

Published online: 25 Nov 2020.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tjim20

Journal of Interdisciplinary Mathematics
ISSN: 0972-0502 (Print), ISSN: 2169-012X (Online)
DOI : 10.1080/09720502.2020.1833458

Taylor & Francis
Taylor & Francis Group

Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario

Amit Kumar Gupta
Amity School of Engineering & Technology
Amity University Jaipur
Jaipur 302006
Rajasthan
India

Vijander Singh*
Department of Computer Science and Engineering
Manipal University Jaipur
Jaipur 303007
Rajasthan
India

Priya Mathur
Department of Mathematics
Poornima Institute of Engineering & Technology
Jaipur 302022
Rajasthan
India

Carlos M. Travieso-Gonzalez
Department of Signals & Communication (IDeTIC)
University of Las Palmas de Gran Canaria
Las Palmas
Spain

---

*E-mail: **vijan2005@gmail.com**

© Taru Publications

Abstract

The whole world is embroiling the pandemic situation caused by COVID-19, which is spreading across all countries. As of mid-May, COVID-19 continues to increase the number of people affected and the number of deaths in each country. Each country's administrations concerned are making endless efforts to maintain public health, mental health and to regulate the rate of illness of COVID-19. Analysis of COVID-19 data using the machine learning paradigm is becoming a major interest of the researcher in these situations. Several researchers analyzed data from COVID-19 to predict infection, death, cured persons in the future, which may lead to the planning of each country's regulatory authority to maintain the public health of its people. The machine learning algorithm provides more accurate results when the data size is large due to the lower number of data sets available to COVID-19, making the most accurate predictions a challenging task to implement the machine learning algorithm. This paper was essentially designed to predict the active rate, the death rate, and the cured rate in India by analyzing the data of COVID-19. There are three models of machine learning Support Vector Machine (SVM), Prophet Forecasting Model, and Linear Regression Model for predicting active rate, death rate and cured rate. Prophet Forecasting Model has been shown to be the best predictive method for predicting active rate, death rate and cured rate compared to SVM and Linear Regression when the vast uncertain and small data sets.

Subject Classification: *(2010) 68T01.*

Keywords: *COVID-19, Support vector machine, Prophet forecasting model, Linear regression model, Machine learning, Forecasting, Time series analysis.*

1. Introduction

The world is facing the Novel Coronavirus (COVID-19) pandemic that outbroke in December 2019 in Wuhan , China. As on May 13, 2020 the active cases in the world has grown to 43.07 lakh, total deaths have reached 2.90 lakh. The pandemic is already spreading in all states of India, and the figure is continually rising.

In this paper, the authors analyzed the COVID-19 data from India using models, namely the support vector machine (SVM), the prophet prediction model and the linear regression model. The objectives of this analysis are: (a) to compile a time series COVID-19 pandemic data set for the forecasting of the active rate, death rate and cured rate (b) comparative analysis of the forecasting results produced by the support vector machine, linear regression model and prophet model, among which the prophet model provides better accuracy (c) to predict the future public health scenario.

In this paper the sections 2 discussed the work of COVID-19 and other related pandemic work. Methods and materials have been explained in Section 3. Section 4 shows the train process and the test of the models, the results are discussed in section 5 and section 6 concludes the paper and identifies possible future scope.

2. Related Work

Li Q et. Al. identified that novel Coronavirus (COVID-19) pandemic was initiated from Wuhan, China in December 2019 which has been spread out whole world. Authors reported that COVID-19 has diverse effects on the public health due to no early detection and a person infected by COVID-19 infect several persons within a small time [1].

Liu K et.al.'s clinical study on different age group patients of COVID-19 compared the clinical data and suggested that the morality of elderly patients is larger than the teen and medium-aged patients. They have also emphasized that the elder patients have higher proportions than teen and medium-aged patients in terms of PSI score IV and V [2].

Cobb JS et.al., identified the effect of COVID-19 on cumulative growth trend of shelter in place (SIP) orders of US. They have used statistical analysis and random forest approach to analyze the cumulative growth trend of COVID-19. They have concluded that the growth trend of COVID-19 is effectively reduced after SIP order of US government and countries with higher density getting more benefit from the SIP order [3].

S.J. Fong et.al define a hybrid methodology for time series analysis to forecast the COVID-19 cases. The hybrid model is combination of composite Monte-Carlo (CMC) and GROOM methodology [4]. The result of CMC improved by using deep learning neural network and set of fuzzy rules. It has been experimentally proved that the suggested model provide better results for the forecasting the possibilities on the highly uncertain and lower size data set [5].

Tanujit Chakraborty et. al performed risk analysis and real time forecasting on COVID-19 data set of France, Canada, South Korea, India and UK. Research has been conducted in two phases. First phase covers hybrid model of ARIMA [6] and Wavelet-based forecasting model [7] for the real time forecasting. In second phase they have used the regression tree model to find the dependency parameter for morality rates [8].

Wei Yu et. al have identify the SVM as best classifier in terms of analysis of uncertain data. They have experimentally analyzed the data set of diabetes and pre-diabetes of US population [9].

A variety of studies related to the pandemic of COVID-19 has been carried out considering the time series analysis to forecast future results. Several time series methods [10-14] have been used including ARIMA

[15], forecasting the seasonality and trend [16], and Epidemic forecasting [17-19] for pandemic prediction.

Vaibhav Bhatnagar et. al have analyzed different data of human being like gender, age, communication type, current status and travel history for the identification the factor for COVID-19 spread. This paper concludes that only age is not a factor of spreading the COVID-19. The Indian people dataset taken for analysis [20].

Vijander Singh et.al have analyzed the COVID-19 data and uses the two models support vector machine (SVM), linear regression to find out the better model for predicting the mortality rate. It concludes that SVM is better predictor to predict the mortality rate over uncertain data of COVID-19 [21].

3. Methods and Material

The research has been performed over three machine learning algorithms SVM, linear regression and prophet methods. The methodology adopted for this analysis described in following sections.

3.1 *Data Collections*

The dataset has been extracted from "https://raw.githubusercontent.com/umangkejriwal112 2/MachineLearning/master/Data%20Sets/covid_19_clean_complete.csv**".** The data set covers date wise confirmed, death, recovered from date January 20, 2020 to May 07, 2020. Authors have calculated the day wise active cases (ac), active rate (ar), death rate (dr), and cured rate (cr) by the following equations where cc is confirmed cases, dc is death cases and rc is recovered cases for i = first day (January 20, 2020) to last day (May 07, 2020):

$$ac_i = cc_i - dc_i - rc_i$$

$$ar_i = \frac{cc_i}{ac_i} * 100$$

$$dr_i = \frac{cc_i}{dc_i} * 100$$

$$cr_i = \frac{cc_i}{rc_i} * 100$$

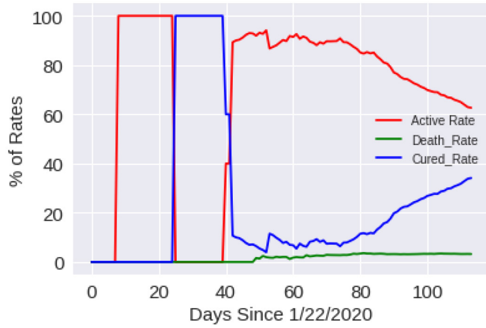

Figure 1

Active Rate, Death Rate and Cured Rate due to COVID-19 in India

Now the data set of India has following columns Date, Confirmed, Death, Recovered, Active, Active rate, Death rate and Cured rate. The figure-1 shows the graph of active rate, death rate and cured rate day wise in India.

3.2 *Support Vector Machine (SVM)*

The SVM [22] is a supervised learning model which applied on labeled data set. It produces a series of labelled input-output mapping functions and features details. The SVM can be used as a regression method or classification method **[17]**.

3.2.1 SVM Classification (SVC)

The nonlinear classification kernel function which generates space for input data to grow more distinguishable associated with the original input space. Maximum-margin hyperplanes are built. The resulting model relies only on a subset of training data within the class boundaries.

3.2.2 SVM Regression (SVR)

The model generated by the Support Vector Regression do not consider any training data that is too similar to the predicted values. In the linear regression the cost function given in equation 1 is aim to minimized.

$$\text{MIN} \sum_{i=1}^{n} (y_i - x_i c_i)^2 \quad (1)$$

Where y_i is the label, c_i is the coefficient, and x_i is the features of labeled data set. SVR defines the hyperplane in higher dimensional to perfectly fitting of input data. SVR also provide to end users to choose the acceptable error in the model. SVR have objectives to minimize the coefficients not the cost function or squared error. Error is marked in to the constrained where the identified margin is set to greater than or equal to the absolute error, known as maximum error, $\in$ (denoted by epsilon). The tuning is performed over the $\in$ (epsilon) to acquire the accuracy of the model. The objective of SVR is defined by following equation

$$\text{MIN}\,\frac{1}{2}\|c\|^2 \tag{2}$$

Constraint

$$|y_i - x_i c_i| \leq \in \tag{3}$$

3.3 *Linear Regression Model*

The LR [23] model simply finds the best fit line across the set of data input (features and labels). The LR draws a regression line through which all data points have least distance from it **[24]**. Mathematically the LR can be expressed in the form of equation of line, A simple linear regression line equation can be written as:

$$y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i, \qquad i = 1, 2 \ldots n \tag{4}$$

The equation (4) can be expressed individually by following equations

$$y_1 = \alpha_0 + \alpha_1 x_1 + \varepsilon_1 \tag{5}$$

$$y_2 = \alpha_0 + \alpha_1 x_2 + \varepsilon_2 \tag{6}$$

$$\vdots$$

$$y_n = \alpha_0 + \alpha_1 x_n + \varepsilon_n \tag{7}$$

All the equation will be added to get the equation

$$\sum y_n = n\alpha_0 + \alpha_1 \sum x_i + \sum \varepsilon_i \tag{8}$$

In linear regression, the sum of the residuals is zero. Removing the $\Sigma\varepsilon_i = 0$ the following equation will get

$$\sum y_n = n\alpha_0 + \alpha_1 \sum x_i \tag{9}$$

The LR uses the least square test to reduce the amount of the squared distance from the regression line. If C represents the sum of the squared residuals:

$$C = (\alpha_0 + \alpha_1 x_1 - y_1)^2 + (\alpha_0 + \alpha_1 x_2 - y_2)^2 + \ldots + (\alpha_0 + \alpha_1 x_n - y_n)^2 \quad (10)$$

This is a quadratic polynomial problem. To minimize C, we take the partial derivatives with respect to α_1 and set the results to 0 and we get:

$$\sum x_i y_i = \sum x_i \alpha_0 + \alpha_1 \sum x_i^2 \quad (11)$$

Solving equations (1) and (2), we can then get the value of α_0 and α_1

3.4 *Prophet Forecasting Model*

Prophet model [24] is designed by facebook for forecasting time series analysis based on simple linear equation which fit the non-linear trends by adding the daily, weekly, and yearly seasonality by considering holiday effect. Prophet basically fit the non-linear and linear functions of time as regressor **[19]**. The model equation of Prophet Model can be expressed as follows:

$$k(t) = tr(t) + se(t) + ho(t) + id(t) \quad (12)$$

Where tr (t) : **trend** that can be defined non-periodic changes in terms of growth, se(t): **seasonality** that can be defined periodic changes that can be measured in form of weekly, monthly or annual, ho(t): **Holiday** defines effects of holidays (possibly unpredictable schedules ≥ 1 day(s)), id(t): **Individual** define individual changes not adjusted by the model.

Now lets understand mathematics behind these parameters is shown in below section.

3.4.1 Trend

In Prophet Model the trends can be illustrated by two fundamentals which depends upon the growth nature if the growth is logistic then the **Saturating growth model** is defined for the measuring the trends and if growth is linear then the **Piecewise linear model** is defined for the measuring the trends.

a. Saturating growth model (if growth is logistic):

When the data is saturated in nature, the patterns can be determined by the equation (13) i.e. wrestling restrictions such as: cubed images, computing capacity, number of users, Internet connection. Typical simulation of these nonlinear, saturating phenomena is effectively achieved.

$$tr(t) = \frac{CC}{1+\exp(-gr(t-d)} \tag{13}$$

Where: CC is the carrying capability, gr is the growth rate, d is an offset parameter

The CC cannot be constant because the data features have varies in numbers like the number of persons use social media is increases or decreases. So the changing of the CC with a time-dependent capacity CC(t). To identify behavior changes of the model it need to identify the change points where the growth rate has been changed.

Let's suppose the number of change points are A at time a_j, where j = 1, …, A.

If δj defines the changing in the rate at time a_j then the rate for time t will be summation of the is then the base rate (gr), and δj in the condition of $j : t < a_j$, i.e. given by following equation.

$$gr + \sum_{j:t<a_j} \delta_j \tag{14}$$

It can be more precisely defined by describing a vector using following equation.

$$\mathrm{b(t)} \in \{0,1\}^A \tag{15}$$

It can be better understand by below mentioned equation:

$$b_j(t) = \begin{cases} 1, & \text{if } t \geq a_j \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

The rate at time t is can be expressed by $\mathbf{gr+b(t)}^T\delta$. When the rate **gr** is changed due to δj , the offset parameter **d** need to change to link the endpoints of the sections. Thus accurate changes in d at change point j is evaluated by following equation:

$$\pi_j = \left(a_j - d - \sum_{i<j} \pi_i\right)\left(1 - \frac{gr + \sum_{i<j} \delta_i}{gr + \sum_{i\leq j} \delta_i}\right) \tag{17}$$

Finally, the piecewise growth is derived by the below mentioned equation:

$$tr(t) = \frac{CC(t)}{1+\exp(-(gr+b(t)^{\mathrm{T}}\delta)(t-(\mathbf{d}+b(t)^{\mathrm{T}}\pi)))} \tag{18}$$

b. Piecewise linear model (if Growth is Linear)

The forecasting dataset which does not have the saturated growth, the Piecewise constant rate of growth evaluate the most accurate forecasting often used by the Prophet. The linear trend is modeled and there is no requirement of adjusting of basic growth rate and offset. It can be calculated by following equation

$$gr(t) = (gr+b(t)^{\mathrm{T}}\delta)t+(\mathrm{d}+b(t)^{\mathrm{T}}\pi) \tag{19}$$

Where: **gr** is the rate of growth, δ is changing in the rate, **d** represent the displacement.

3.4.2 Seasonality

The seasonal component se(t) defines and add the effects of changes periodically in daily, weekly and annual seasonality of the data set. The prophet method uses the Fourier series analysis to define an agile model for periodic effects. Suppose that R determines the standard time span that the time series would have. It can be expressed R = 365.25 for annual data or R = 7 for weekly data, when time is measured in days. The se(t) could be expressed by below mentioned equation:

$$se(t) = \sum_{i=1}^{J}\left(a_n \cos\left(\frac{2\pi it}{R}\right)+b_n \sin\left(\frac{2\pi it}{R}\right)\right) \tag{20}$$

Seasonality fitting involves the calculation of 2J parameters $\alpha = [a1, b1, \ldots, aJ, bJ]^{\mathrm{T}}$. It is achieved by forming a matrix of seasonal vectors for each value of t in our past and predicted results.

3.4.3 Holiday

In Prophet it is also evaluated that the holiday also has impact on the time series analysis of the datasets for forecasting. The parameter ho(t) is also consider on asymmetrical schedules for predictive happenings of the year. Let Pi is the collection of all historical and forthcoming dates for that holiday. It includes a pointer function which shows that time t is during holiday i. we also add every holiday a value £i i.e. equivalent updation in the forecast. That can be expressed as similar procedure as seasonality by obtaining a matrix of regressor

$$Y(t) = [1(t \in P_1), \ldots\ldots\ldots, 1(t \in P_L)] \quad (21)$$

$$ho(t) = Y(t)\, £ \quad (22)$$

4. Train and Test Models

This paper uses SVM, LR and Prophet for forecasting the active rate, death rate and cured rate on the data set COVID-19 pandemic in India. The study found that Prophet method is more accurate in forecasting the time series analysis due its trend measurement characteristics which may either linear or logistic for determine the growth curve. The seasonal module uses the Fourier transformation series for analysis of data based on yearly, weekly, or daily basis. Prophet model automatically identify the change points in data which affects the trends [25]. SVM may shows more accuracy in the area of classification, time series forecasting, data mining and regression. The nonlinear SVM regressor generates a high dimensional features space for input data to develop more distinguishable associated to the original input space for time series forecasting. Maximum-margin hyperplanes are then created [26- 29]. LR can also be used for the time series forecasting. The LR predicts in two ways, one is casual method (it is used when the data set does not have the timing constraint) and second is time-series analysis (which fits the points with the regression line of historical performance of dataset). The LR method does not use the seasonality or other component which may affect the data on the parameter of time [30]. Data have been split in training and testing data set for the purpose of train and validate model. All three models have applied to predict the active rate, death rate and cured rate for future of three weeks. The SVM tuned for the parameters input space which transforms it to a higher dimensional space. Prophet model has also cross-validated in the span of 50 days to check the results of the model. The figure 2, 3 and 4 shows the graph between the actual test data and the prediction made over the test data using SVM train model. The figure 5, 6 and 7 shows the graph between the actual test data and the prediction made over the test data using LR train Model. The figure 8, 9 and 10 shows the graph for predicted and actual datasets using Prophet.

Figure 2

Test the active rate using SVM trained Model

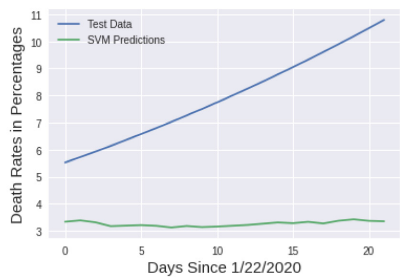

Figure 3

Test the Death rate using SVM trained Model

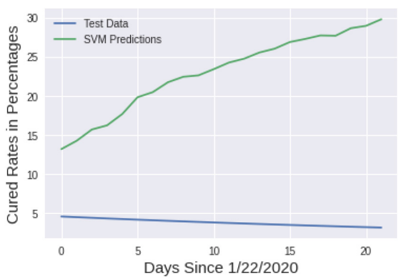

Figure 4

Test the cured rate using SVM trained Model

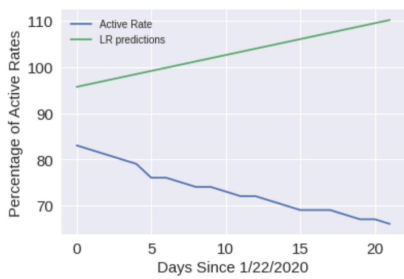

Figure 5

Test the Active rate using LR trained Model

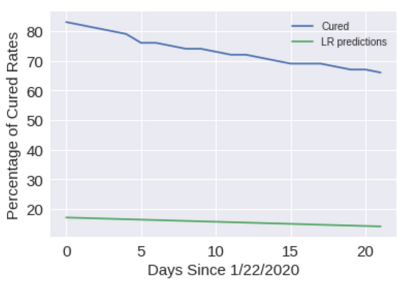

Figure 6

Test the Death rate using LR trained Model

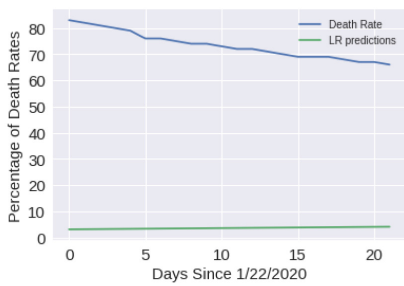

Figure 7

Test and cured rate using LR trained model

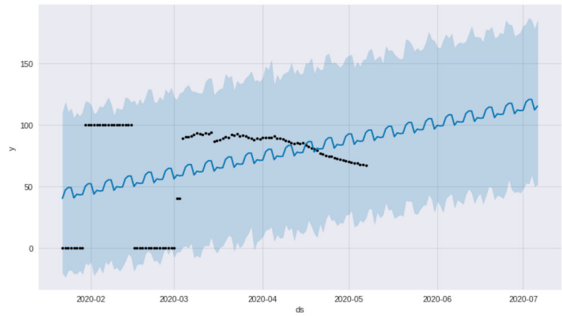

Figure 8

Forecasting the Active rate using Prophet Model

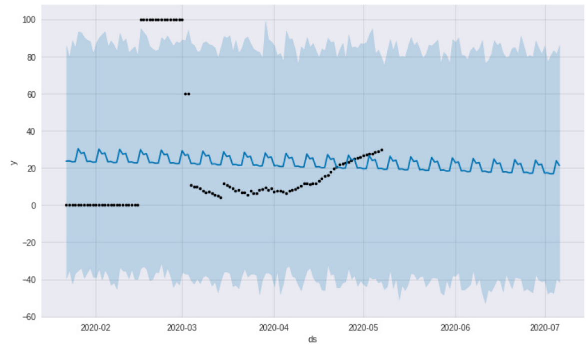

Figure 9

Forecasting the cured rate using Prophet Model

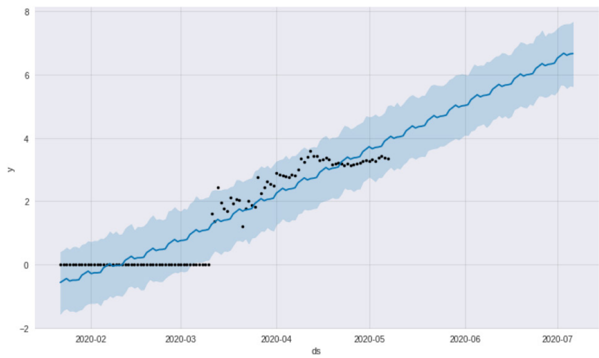

Figure 10

Forecasting the Death rate using Prophet Model

5. Results and Discussion

The forecasting of predicted active rate, death rate and cured rate of COVID-19 has been calculated by applying the SVM, LR and Prophet methods. The experimental setup of all three models concludes that the Prophet model gives the most accurate forecasting the predicted COVID-19 active rate, death rate and cured rate in comparison to other two models. The following figures are the graphs between the predicted and actual values of active rate, death rate and cured rate by SVM, LR and Prophet model.

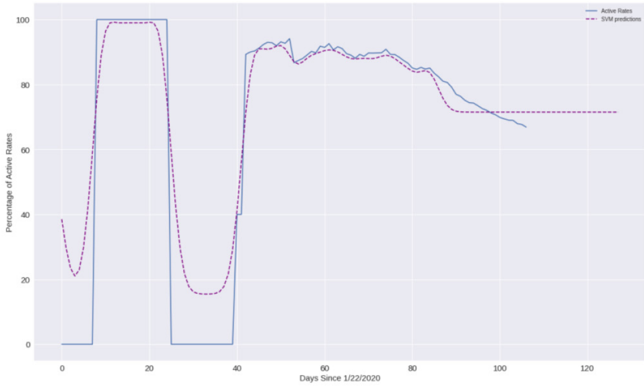

Figure 11

Actual & Predicted Active Rate (SVM Model)

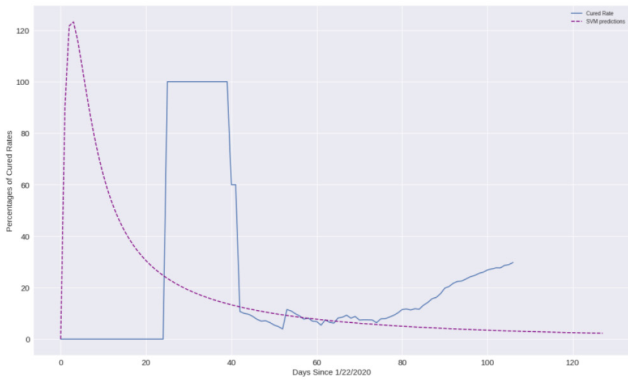

Figure 12

Actual & Predicted Cured Rate (SVM Model)

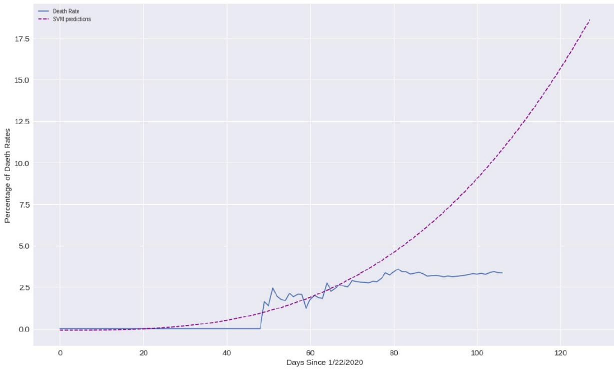

Figure 13

Actual & Predicted Death Rate (SVM Model)

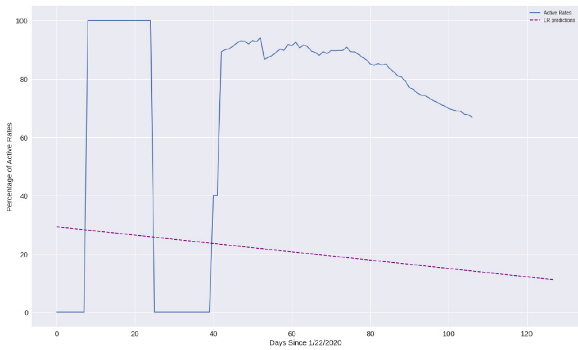

Figure 14

Actual & Predicted Active Rate (LR Model)

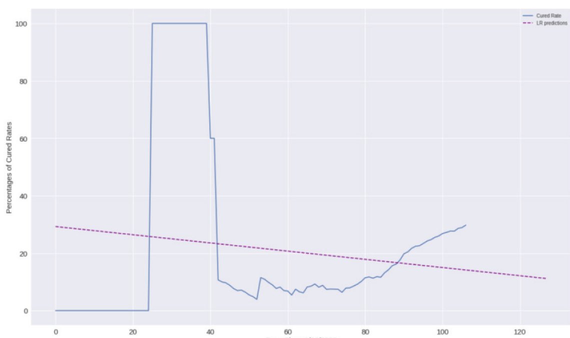

Figure 15

Actual & Predicted Cured Rate (LR Model)

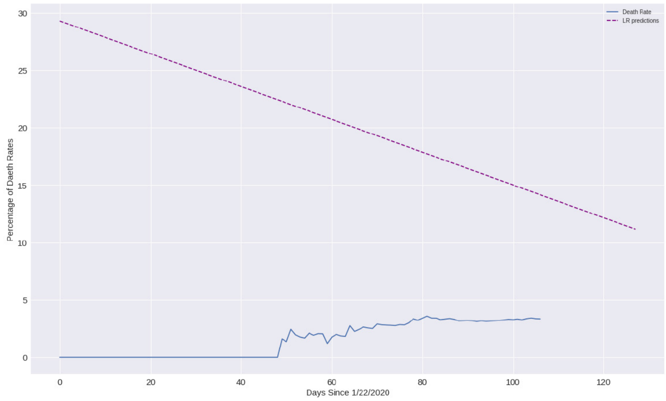

Figure 16

Actual & Predicted Death Rate (LR Model)

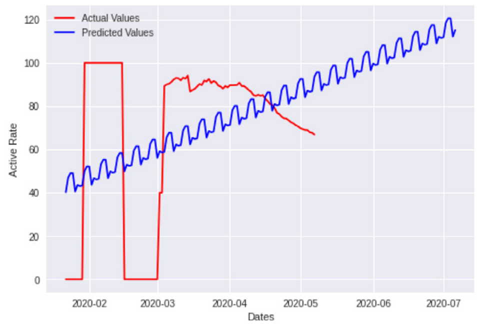

Figure 17

Actual & Predicted Active Rate (Prophet Model)

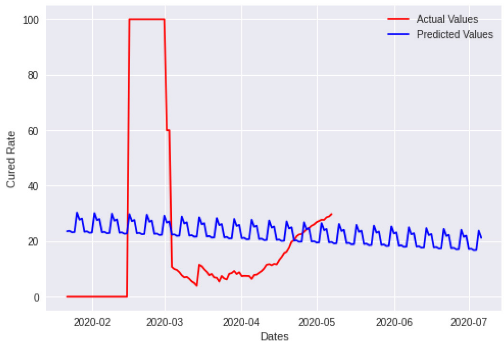

Figure 18

Actual & Predicted Cured Rate (Prophet Model)

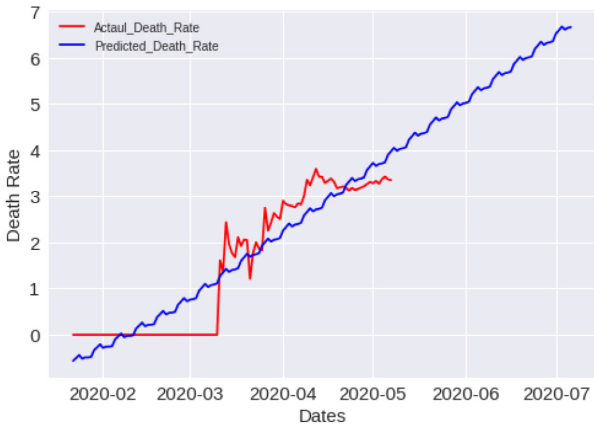

Figure 19

Actual & Predicted Death Rate (Prophet Model)

5.1 *Mean Absolute Error (MAE), Mean Square Error (MSE) & Root Mean Square Error (RMSE) of SVM, LR and Prophet Models:*

In this paper we have also calculated the "Mean Absolute Error (MAE)", "Mean Square Error (MSE)" & "Root Mean Square Error (RMSE)" of proposed models (SVM, LR and Prophet Model) on the active rate, death rate and cured rate. The errors are calculated on the basis of test data of 20% in SVM and LR model. The Prophet model has been cross validated in the horizon of 50 days check points. The table 1 shows the MAE, MSE and RMSE for SVM, LR and Prophet Model i.e. calculated over the data set novel COVID-19. By the analysis of these errors it is concluded that it is concluded that Prophet Model provides the optimized solution for forecasting than other two models.

Table 1

Showing the MAE, MSE and RMSE for SVM, LR and Prophet Model

Model	Parameter	**MAE**	**MSE**	**RMSE**
SVM	Active Rate	15.73	266.82	16.33463
	Cured Rate	366.26	139229.8	373.1351
	Death Rate	3.9	17.12	4.137632

Contd...

LR	Active Rate	33.6	1196.42	34.5893
	Cured Rate	12.95	191.2	13.82751
	Death Rate	0.62	0.466	0.682642
Prophet	Active Rate	31.55511	1037.042	32.20313
	Cured Rate	52.11855	2738.635	52.33197
	Death Rate	1.300317	1.701083	1.304256

6. **Conclusion and Future Scope**

In this research the SVM, LR and Prophet Model has been experimentally tested for the time series analysis of forecasting the predicted active, death and cured rate on novel COVID-19 pandemic data set of Indian population. The main concerned of this research is to compare the SVM, LR and Prophet Model for forecasting the time series analysis on highly uncertain and short data of novel COVID-19 pandemic data set of Indian population. This research suggested that Prophet Method gives the better performance than the SVM and LR model for the time series data forecasting. The Prophet model measures trends of the dataset by analyzing data nature either linear or logistic for determine the growth curve. The seasonal module uses the Fourier transformation series for analysis of data based on yearly, weekly, or daily basis. Prophet model automatically identify the change points in data which affects the trends. These features are not available in SVM and LR method. However, SVM also provide the significant results in comparison to LR Model. SVM generates a high dimensional features space for input data to develop more distinguishable association to the original input space for time series forecasting. Maximum-margin hyperplanes are then created. SVM also generate improved performance values to forecast the predicted COVID-19 data set using kernel function. The LR model gives the worst result in comparisons to other two models. The LR method does not use the seasonality or other component which may affect the data on the parameter of time. From whole analysis it is also suggested that in India more severe restriction and regulating steps are recommended to control the speed of COVID-19. Authors believe that the suggested work will help to the regulating authority of any country to predict the future based on recommended Prophet model for what will be scenario of COVID-19 in coming days. They can make impressive arrangements of medical emergency services, quarantine, social distancing, and Lockdown period to recover from the pandemic situation raised due to novel COVID-19.

References

[1] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, et al., Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia, *New England Journal of Medicine* (2020).

[2] K. Liu, Y. Chen, R. Lin, K. Han, Clinical features of covid-19 in elderly patients: A comparison with young and middle-aged patients, *Journal of Infection* (2020).

[3] J. Cobb, M. Seale, Examining the effect of social distancing on the compound growth rate of sars-cov-2 at the county level (united states) using statistical analyses and a random forest machine learning model, Public Health (2020).

[4] S. J. Fong, G. Li, N. Dey, R. G. Crespo, E. Herrera-Viedma, Composite monte carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction, Applied Soft Computing (2020) 106282.

[5] A. Mubayi, C. K. Zaleta, M. Martcheva, C. Castillo-Ch´avez, A cost-based comparison of quarantine strategies for new emerging diseases, *Mathematical Biosciences & Engineering* 7 (3) (2010) 687.

[6] C. Chatfield, H. Xing, The analysis of time series: an introduction with R, CRC press, 2019.

[7] D. B. Percival, A. T. Walden, Wavelet methods for time series analysis, Vol. 4, Cambridge university press, 2000.

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

[9] W. Yu, T. Liu, R. Valdez, M. Gwinn, M. J. Khoury, Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes, *BMC medical informatics and decision making* 10 (1) (2010) 16.

[10] V. Kurbalija, M. Radovanovi´c, M. Ivanovi´c, D. Schmidt, G. L. von Trzebiatowski, H.-D. Burkhard, C. Hinrichs, Time-series analysis in the medical domain: A study of tacrolimus administration and influence on kidney graft function, *Computers in biology and medicine* 50 (2014) 19–31.

[11] M. C. Thomson, A. M. Molesworth, M. H. Djingarey, K. Yameogo, F. Belanger, L. E. Cuevas, Potential of environmental models to predict

meningitis epidemics in africa, *Tropical Medicine & International Health* 11 (6) (2006) 781–788.

[12] L. Zhang, L. Wang, Y. Zheng, K. Wang, X. Zhang, Y. Zheng, Time prediction models for echinococcosis based on gray system theory and epidemic dynamics, *International journal of environmental research and public health* 14 (3) (2017) 262.

[13] H. Ren, J. Li, Z.-A. Yuan, J.-Y. Hu, Y. Yu, Y.-H. Lu, The development of a combined mathematical model to forecast the incidence of hepatitis e in shanghai, china, *BMC infectious diseases* 13 (1) (2013) 421.

[14] X. Zhang, Y. Liu, M. Yang, T. Zhang, A. A. Young, X. Li, Comparative study of four time series methods in forecasting typhoid fever incidence in china, PloS one 8 (5) (2013).

[15] Y.-w. Wang, Z.-z. Shen, Y. Jiang, Comparison of arima and gm (1, 1) models for prediction of hepatitis b in china, PloS one 13 (9) (2018).

[16] Q. Liu, Z. Li, Y. Ji, L. Martinez, U. H. Zia, A. Javaid, W. Lu, J. Wang, Forecasting the seasonality and trend of pulmonary tuberculosis in jiangsu province of china using advanced statistical time-series analyses, *Infection and drug resistance* 12 (2019) 2311.

[17] C. Orbann, L. Sattenspiel, E. Miller, J. Dimka, Defining epidemics in computer simulation models: How do definitions influence conclusions, Epidemics 19 (2017) 24–32.

[18] E. O. Nsoesie, R. J. Beckman, S. Shashaani, K. S. Nagaraj, M. V. Marathe, A simulation optimization approach to epidemic forecasting, PloS one 8 (6) (2013).

[19] Q. Li, W. Feng, Y.-H. Quan, Trend and forecasting of the covid-19 outbreak in china, *Journal of Infection* 80 (4) (2020) 469–496.

[20] V. Bhatnagar, R. C. Poonia, P. Nagar, S. Kumar, L. Sing, Vijander Raja, P. Dass, Descriptive analysis of covid-19 patients in the context of india, *Journal of Interdisciplinary Mathematics* (2020). doi:https://doi.org/10.1080/09720502.2020.1761635.

[21] V. Singh, R. C. Poonia, S. Kumar, P. Dass, P. Agarwal, V. Bhatnagar, L. Raja, Prediction of covid-19 corona virus pandemic based on time series data using support vector machine, *Journal of Discrete Mathematical Sciences & Cryptography* (2020). DOI : 10.1080/09720529.2020.1784535

[22] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (3) (1995) 273–297.

[23] X. Yan, X. Su, Linear regression analysis: theory and computing, World Scientific, 2009.

[24] S. J. Taylor, B. Letham, Forecasting at scale, The American Statistician 72 (1) (2018) 37–45.

[25] S. J. Taylor, B. Letham, Prophet: forecasting at scale, Facebook research (2017).

[26] R. Samsudin, A. Shabri, P. Saad, A comparison of time series forecasting using support vector machine and artificial neural network model, *Journal of applied sciences* 10 (11) (2010) 950–958.

[27] L.-J. Cao, F. E. H. Tay, Support vector machine with adaptive parameters in financial time series forecasting, I*EEE Transactions on neural networks* 14 (6) (2003) 1506–1518.

[28] G. W. Flake, S. Lawrence, Efficient svm regression training with smo, Machine Learning 46 (1-3) (2002) 271–290.

[29] C. Zhao, H. Zhang, X. Zhang, M. Liu, Z. Hu, B. Fan, Application of support vector machine (svm) for prediction toxic activity of different data sets, *Toxicology* 217 (2-3) (2006) 105–119.

[30] A. Gellert, Linear Regression Forecasting Method by Companies (2009 (accessed February 3, 2014)). URL Chron.com,http://smallbusiness.chron.com/linear-regression-forecasting-method-companies-73112.html.