Journal of Pharmaceutical Research International

33(51A): 305-312, 2021; Article no.JPRI.72549
ISSN: 2456-9119
(Past name: British Journal of Pharmaceutical Research, Past ISSN: 2231-2919, NLM ID: 101631759)

Predictive Analysis of Real-Time Strategy using Face book's Prophet Model on Covid-19 Dataset of India

Pankaj Kumar [a*], Renuka Sharma [b] and S. K. Singh [c]

[a] Department of Computer Science & Engineering, LIET, Greater Noida, India.
[b] Department of Computer Science & Engineering, NIET, Greater Noida, India.
[c] Department of Computer Science & Engineering, GCET, Greater Noida, India.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JPRI/2021/v33i51A33496
Editor(s):
(1) Ana Cláudia Coelho, University of Trás-os-Montes and Alto Douro, Portugal.
(2) Aurora Martínez Romero, Juarez University, Mexico.
(3) Ali Nokhodchi, University of Sussex, UK.
Reviewers:
(1) Impana Appaji, VTU, India.
(2) Kadam Anilkumar Jagannathrao, SP Pune University, India.
(3) Balaji V, Vellore Institute of Technology, India.
(4) Vaishali A. Gaikwad, KJSIEIT, University of Mumbai, India.
(5) M. Sri Bala, Lakireddy Balireddy College of Engineering, India.
Complete Peer review History, details of the editor(s), Reviewers and additional Reviewers are available in this link: https://www.sdiarticle5.com/review-history/72549

Original Research Article

Received 01 September 2021
Accepted 04 October 2021
Published 23 November 2021

ABSTRACT

The global epidemic of the novel coronavirus (COVID-19) called SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) has infected millions and killed millions. The prevalence of the virus is of paramount importance in identifying future infections and preparing healthcare facilities to avoid death. Accurately predicting the spread of COVID-19 is a challenging analytical and practical task for the research community. We can learn to use predictive analytics to predict the positive outcomes of these risks. These predictive analytics can look at the risks of past successes and failures.

In this paper, the Facebook prophet model discusses the number of large-scale cases and deaths in India based on daily time-series data from 30 January 2020 to 30 April 2021, for forecasting and visualization. The covid-19 pandemic could end prematurely if social distancing and safety measures are required to stabilize and control is required to achieve treatment

**Corresponding author: E-mail: unpankaj@gmail.com;*

in India. This paper suggests that the Prophet Model is more effective in predicting COVID-19 cases. The forecast results will help the government plan strategies to prevent the spread of the coronavirus.

Keywords: Prophet; India; Covid-19; coronavirus; face book; prediction; machine learning.

1. INTRODUCTION

Forecasting is a technique that uses historical data as inputs to make informed estimates that are predictive in determining the direction of future trends. It is an important and common data science task in everyday life. Having prior knowledge of any event can help us tremendously in the formulation of its goals, policies and planning. However, producing high quality and reliable forecasts comes with challenges of its own. Forecasting is a complex phenomenon both for humans and for machines [1,2,3]. It also requires very experienced time series analysts, which actually are quite rare.

The world is resisting Covid-19, the world's most dangerous disease, first reported in Wuhan, China, on 31 December 2019. The World Health Organization (WHO) has declared an epidemic due to a steady rise in positive cases and deaths [4,5]. COVID-19 is currently having a widespread impact on the world, affecting an estimated 10 million people in over 100 countries.

India, the second-most populous and infected country in the world, implements a strictly national government policy of social distancing, nationwide lockdown, and bans all forms of travel among the people. Later on, India implemented the containment zone model to prevent the spread of the disease [6,7]. However, due to unforeseen circumstances, people returned to their hometowns, which added fuel to the fire of spreading the disease. The first positive case of SARS-CoV-2 in India was reported in Kerala on 30 January 2020. As a result, the number of cases has increased significantly. According to the Indian Council of Medical Research (ICMR), as of 30 November 2020, there were 9,463,254 cases of COVID-19 [8,9].

COVID-19 has no special treatment and is rapidly becoming popular. Healthcare is important for future cases. Several algorithms (such as Linear Regression, Exponential Smoothing, ARIMA, SARIMA, GARCH, and LSTM, etc.) could be used for future cases. Although linear regression and exponential smoothing can handle multiple time series components, these models are sensitive to anomalous processing and have narrow confidence intervals. The ARIMA model combines automatic regression and moving averages with realistic confidence intervals. However, this algorithm requires a large amount of data to predict future cases [10,11].

This paper uses the Prophet model to predict the number of confirmed cases and deaths in India. Input datasets, including confirmed cases and deaths, are stored on a separate sheet.

The Prophet algorithm is open source and was developed by the Facebook AI team to analyze time-series forecasts. The Prophet is based on an additive model in which the nonlinear model is equipped with hours, days, weeks, months, years, or regular seasonality [12]. The Prophet architecture has a unique data structure for efficient management of time series and seasons [13]. The mathematical logic underlying the Prophet model includes three main elements, such as trends, seasonality, holidays, and an unlimited time series model with an additional fourth noise/error element. Since it follows the principle of additive regression, the Prophet is well suited as a component of linear and non-linear functions of time [14].

1.1 Literature Review

This section describes the prediction algorithms applied to time series datasets. Time-series are defined as sequence of data. COVID-19 is a time-series data. A large amount of data came out in the age of modern science. Designing, implementing, and running computational tools is essential to understanding large amounts of data. It is difficult for researchers to reach consensus on which method to apply to which problem. For this reason, several predictive algorithms have been used in the last decade to improve future data.

Prediction is the science of estimating the future of a particular variable. As datasets grow larger and more complex, and computational analysis becomes more frequent, scientific researchers can find better-fitted data.

T. Ozturk et al. [15] propose a model for automatic detection of COVID-19 using the original chest X-ray. The proposed model is

designed for accurate diagnosis of binary classification.

Implementation of educational methods that have formed many complex and deep networks by M. Rahimzadeh et.al. The training method is based on two open source datasets and divides x-rays into three categories: normal, pneumonia, and COVID-19.

Martı´nez-Alvarez F, et.al. [16] proposed a new metaheuristic method that can simulate the spread of coronavirus and infectious diseases in healthy people. Starting with the main infection (zero patients), the coronavirus quickly infects new victims, killing many or spreading the infection.

The cellular automata are believed to provide an excellent platform for sequential gene algorithms that can efficiently estimate dynamic parameters in the accurate data-driven transmission simulation. The concept is proposed by S. Ghosh et.al. [17].

A. Tomer et.al. [18] used data-driven scoring methods (such as long short-term memory and curve fitting) to describe the number of COVID-19 cases in India 30 days ago, with social isolation and COVID- 19 Prevented epidemics. Authors predict the outcome of precautionary measures. Prediction is done using various parameters (number of positive cases, number of cured cases, etc.).

Jayesh S et al. [19] will analyze the case of Covid-19 and take effective action by the Indian government of Kerala in response to the outbreak of Kerala.

2. METHODS

2.1 Covid-19 Dataset

In this paper, this study used COVID-19 data as an experiment to analyze the effectiveness of the predictive approach. A dataset of daily COVID-19 cases in India was collected from Worldometer (https://www.worldometer.info/coronavirus/country/india/) [20,21]. The COVID-19 data used in this study include daily confirmation, recovery, and death from March 2020 to April 2021. We split the sample dataset into two data files and started building the model. The first 80% of the sample is used for training and the remaining 20% is used for the test suite to avoid overtraining. The basic set of test data was only used to test the validity of the constructed classification model. This approach provides an idea of the success of the actual model. The methodology is shown in Fig. 1.

2.2 Facebook Prophet

Accurate predictions of the future can be extremely valuable. Prediction is a common data science task that we can use to set goals. Machine learning for prediction is a branch of computer science that shapes historical data using artificial neural networks, deep learning, decision trees, and Bayesian networks [18]. The idea behind the algorithm is to select the appropriate training model based on the properties of previous data and use it to predict future observations. The Prophet's method is applied to the prediction of COVID-19 in India. The non-linear tendencies of the Prophet affect annual, weekly, and daily holidays. The correct function of the Prophet is not only to predict the future but also to fill in the missing values and detect anomalies.

Prophet or “Facebook Prophet” is a procedure/tool, and an open-source library for univariate time-series reliable prediction developed by Facebook in 2017. It uses a decomposable time-series with three main model components: trend, seasonality, and holidays [22]. They are combined in the following equation:

$$\text{Prophet Prediction} = \text{Trend} + \text{Seasonality} + \text{Holiday} + \text{error}$$

$$P(t) = T(t) + S(t) + H(t) + E(t)$$

Here $T(t)$ is the trend function, which models non-periodic (i.e. linear or logistic) changes in the value of the time series. $S(t)$ is the Seasonality function, which model the periodic changes like daily, weekly, or yearly seasonality. $H(t)$ is the Holiday function, which occurs on irregular schedules over a day or a period of days. $E(t)$ is the error terms, which represent the any idiosyncratic changes which are not accommodated by the model.

It is Similar to a generalized additive model (GAM), with time as a regressor; Prophet fits several linear and non-linear functions of time as components [19]. The GAM formulation has the advantage that it decomposes easily and accommodates new components as necessary, for instance when a new source of seasonality is identified.

Prophet is “framing the forecasting problem as a curve-fitting exercise” rather than looking explicitly at the time-based dependence of each observation. It does not allow non-Gaussian noise distribution, autocorrelation on residual and stochastic trend [23].

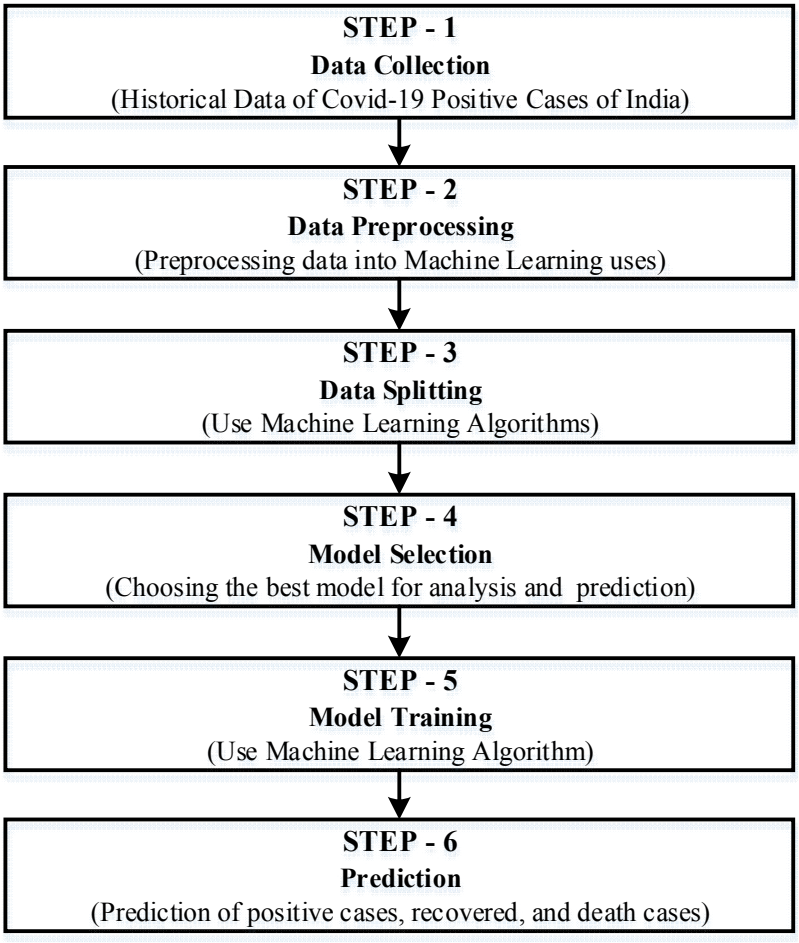

Fig. 1. Proposed methodology

Advantages of Facebook's Prophet are more powerful and often effective. It is simple to use, interpret, and highly tunable open-source. It Scales well in both R and Python. We use Python. Disadvantages are it works primarily on one time series data and requires data to be specified in a specific format. Its performance varies by data set. All limitations that come with additive models.

COVID-19 datasets is a time-series dataset that have strong seasonal effects and several seasons of historical data. Prophet is used in Facebook for reliable forecast to set the goal. The performance of Prophet is better than any other approach in the majority of cases. Therefore, we choose Prophet for predictive analysis of COVID-19 time-series dataset. It is robust to missing data and shifts in the trend, and typically handles outliers well.

3. RESULTS, ANALYSIS, AND DISCUSSION

In this section, we study the spread of COVID-19 in India as there are thousands of cases reported each day. For validation and analysis of the proposed model, data pertaining to India has been used with the Python Programming.

Prediction of future events is always challenging in data modelling. For the final stage of validation of the methodology, the predictive power of the model has been tested. As the impacts of this Covid-19 pandemic becomes far reaching as the socioeconomic contexts vary, a considerably accurate prediction about the dynamics of the infection spread can be crucial and useful in many ways.

The Python libraries (i.e. matplotlib, prophet, seaborn, and datetime) is used to predictive analysis of time series Covid-19 data.

Fig. 2 shows the variation of the daily covid-19 cases as confirmed and corresponding Table 1 shows the date wise predicted data. The cases has so much fluctuation. As one variable increases, so does the other. The graph can be used to describe different Covid-19 cases based on events over time.

Fig. 3 shows the variation of the daily covid-19 cases as death and corresponding Table 2 shows the date wise predicted data. The cases has so much fluctuation. As one variable increases, so does the other. The graph can be used to describe different Covid-19 death cases based on events over time.

Fig. 4 shows the variation of the daily covid-19 cases as recovered and corresponding Table 3 shows the date wise predicted data. The cases has so much fluctuation. As one variable increases, so does the other. The graph can be used to describe different Covid-19 recovered cases based on events over time.

Table 1. Predicted total no. of cases

Sl.	ds	yhat	yhat_lower	yhat_upper
459	03/05/2021	1.51E+07	1.40E+07	1.62E+07
460	04/05/2021	1.52E+07	1.41E+07	1.63E+07
461	05/05/2021	1.53E+07	1.42E+07	1.64E+07
462	06/05/2021	1.53E+07	1.42E+07	1.64E+07
463	07/05/2021	1.54E+07	1.43E+07	1.65E+07

Table 2. Predicted total no. of death cases

Sl.	ds	yhat	yhat_lower	yhat_upper
460	04/05/2021	181712.16	174846.431	189150.592
461	05/05/2021	182142.466	175294.589	189381.425
462	06/05/2021	182582.686	175318.363	189879.575
463	07/05/2021	183012.103	175484.682	190477.718
464	08/05/2021	182914.479	176038.248	190618.263

Table 3. Predicted total no. of recovered cases

Sl.	ds	yhat	yhat_lower	yhat_upper
459	03/05/2021	1.32E+07	1.26E+07	1.38E+07
460	04/05/2021	1.33E+07	1.26E+07	1.39E+07
461	05/05/2021	1.33E+07	1.27E+07	1.39E+07
462	06/05/2021	1.33E+07	1.27E+07	1.39E+07
463	07/05/2021	1.34E+07	1.27E+07	1.40E+07

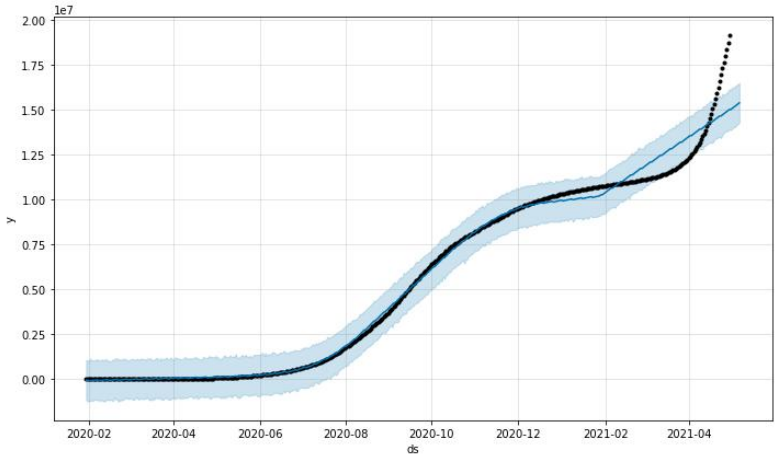

Fig. 2. Graph of Predicted total no. of cases

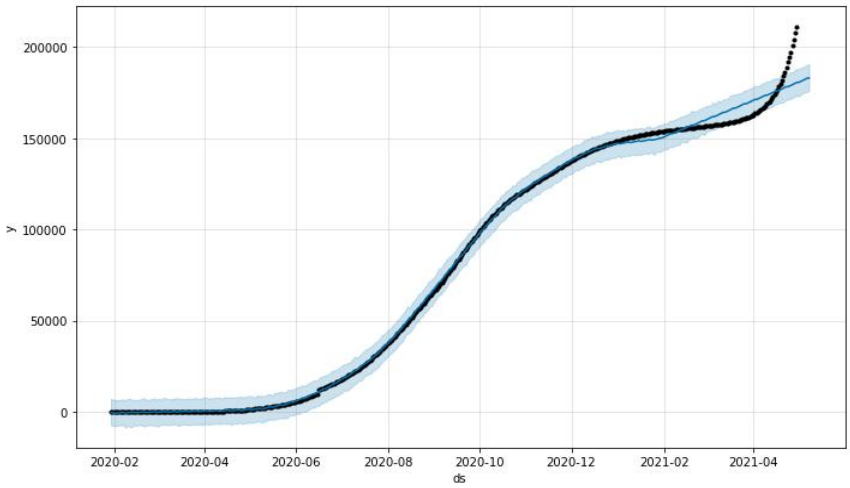

Fig. 3. Graph of Predicted total no. of death cases

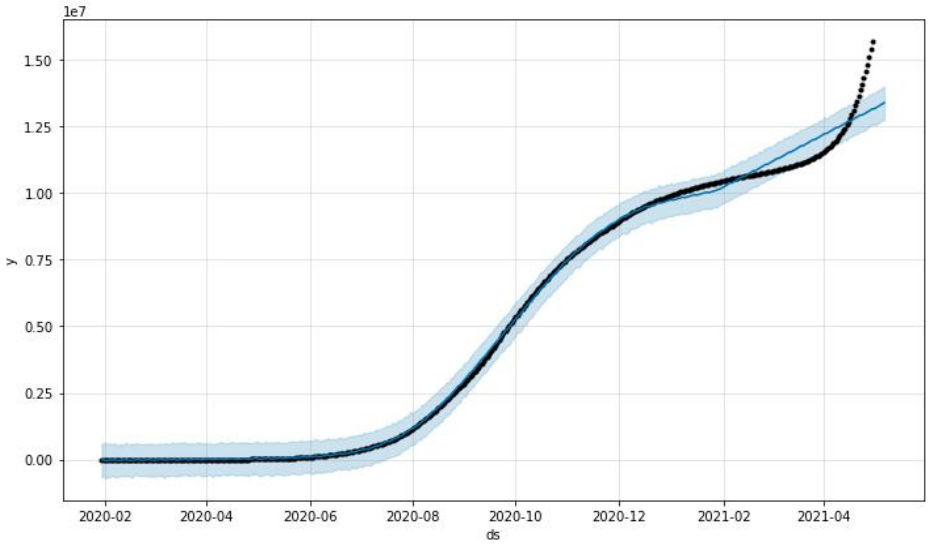

Fig. 4. Graph of Predicted total no. of recovered cases

4. CONCLUSION

India is one of the high population counties in world. To stop Covid -19 is biggest challenges to county like India. But today's date (26 August, 2020); India spread of Corona viruses to extent. From the prediction times values tells that India would reach more than 32 lakhs with 64,213 deaths till 1 September 2020. Now after the all lock down, cases are increasing day by day. India has increasing day by day testing, which helps to get more information about Covid-19 spread. Also India is working on small trials on plasma therapy. India has developed aarogya setu to create more awareness about Covid-19. Due to lockdown spread to Covid-19 controlled and cases are less as compare to other countries. After the lockdown, biggest is challenge to maintain social distancing in society. But actually, medicine is required to completely stop the spread of Covid-19. This projection help government to make strategies against Covid-19.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. KumarP, Singh SK, Deo Choudhary S. Reliability prediction analysis of aspect-oriented application using soft computing techniques Mater. Today Proc. 2021;45: 2660–2665.
   DOI: 10.1016/j.matpr.2020.11.518.
2. Ardabili SF, et al. COVID-19 Outbreak Prediction with Machine Learning," Algorithms. 2020;13(10):249.
   DOI: 10.3390/a13100249.
3. Arora P, Kumar H, Panigrahi BK. Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. Chaos, Solitons and Fractals. 2020;139.
   DOI: 10.1016/j.chaos.2020.110017.
4. Kumar P, Singh SK, Chishti SN. A comprehensive investigation of quality of AOP-based Small-scale projects using aspect-oriented software quality (AOSQ) model. Proc. - IEEE 2018 Int. Conf. Adv. Comput. Commun. Control Networking, ICACCCN 2018;497–500. DOI: 10.1109/ICACCCN.2018.8748592.
5. Car Z, Baressi Šegota S, Anđelić N, Lorencin I, Mrzljak V. Modeling the Spread of COVID-19 Infection Using a Multilayer Perceptron," Comput. Math. Methods Med;2020.
   DOI: 10.1155/2020/5714714.
6. Kumar P, Singh SK. An emerging approach to intelligent techniques—soft computing and its application. Intell. Syst. Ref. Libr. 2020;180:171–182.
   DOI: 10.1007/978-3-030-39119-5_13.
7. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time," Lancet Infect. Dis. 2020;20(5):533–534.
   DOI: 10.1016/S1473-3099(20)30120-1.
8. Kumar P, Singh SK. A systematic assessment of aspect-oriented software development (AOSD) using JHotDraw application. Proceeding - IEEE Int. Conf. Comput. Commun. Autom. ICCCA 2016. 2017;779–784.
   DOI: 10.1109/CCAA.2016.7813840.
9. Kumar P, Singh SK. A comprehensive evaluation of aspect-oriented software quality (AOSQ) model using analytic hierarchy process (AHP) technique," Proc. - 2016 Int. Conf. Adv. Comput. Commun. Autom. (Fall), ICACCA 2016;2016.
   DOI: 10.1109/ICACCAF.2016.7748957.
10. Kumar PSK. Defect prediction model for aop-based software development using hybrid fuzzy c-means with genetic algorithm and k-nearest neighbors classifier. Int. J. Appl. Inf. Syst., 2016;11(2):26–30.
    DOI: 10.5120/ijais2016451579.
11. Kumar P, Dixit S, SK. Performance of aspect-oriented software quality modelling using artificial neural network technique. Int. J. Comput. Appl. 2019;182(36):6–10.
    DOI: 10.5120/ijca2019918352.
12. Kumar P, Bajpai B, Gupta DO, Jain DC, Vimal S. Image recognition of COVID-19 using DarkCovidNet architecture based on convolutional neural network. World J. Eng., vol. ahead-of-p, no. ahead-of-print, 2021,
    DOI: 10.1108/wje-12-2020-0655.
13. Rahimzadeh M, Attar A. A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2," Informatics Med. Unlocked. 2020;19:100360.
    DOI: 10.1016/j.imu.2020.100360.
14. Kumar P, Singh SK. A Framework for Assessing the Evolvability Characteristics along with Sub-characteristics in AOSQ Model Using Fuzzy Logic Tool," in IEEE 3rd International Conference on Computing, Communication and Automation (ICCCA-2017), Greater Noida, (Uttar Pradesh), India. 2017;339–344.
    DOI: 10.1109/CCAA.2017.8229839.
15. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput. Biol. Med. 2020; 121:103792.
    DOI: 10.1016/j.compbiomed.2020.103792.
16. Martínez-Álvarez F, et al. Coronavirus Optimization Algorithm: A Bioinspired Metaheuristic Based on the COVID-19 Propagation Model. Big data. 2020;8(4):308–322.
    DOI: 10.1089/big.2020.0051.
17. Ghosh S, Bhattacharya S. A data-driven understanding of COVID-19 dynamics using sequential genetic algorithm based probabilistic cellular automata. Appl. Soft Comput. J. 2020;96:106692.
    DOI: 10.1016/j.asoc.2020.106692.

18. Tomar A, Gupta N. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. Sci. Total Environ. 2020;728: 138762. DOI: 10.1016/j.scitotenv.2020.138762.
19. Sreedharan JSS. Analysing the Covid-19 Cases in Kerala: a Visual Exploratory Data Analysis Approach. SN Compr. Clin. Med. 2020;2(9):1337–1348.
DOI: 10.1007/s42399-020-00451-5.
20. Worldometer:
Available:https://www.worldometers.info/coronavirus/country/india/ [Last visited: 30/04/2021].
21. Domingos DS, de Oliveira JFL, de Mattos Neto PSG. An intelligent hybridization of ARIMA with machine learning models for time series forecasting. Knowledge-Based Syst. 2019;175:72–86.
DOI: 10.1016/j.knosys.2019.03.011.
22. Kumar P. Aspect-Oriented Software Quality Model: The AOSQ Model," Adv. Comput. An Int. J. 2012;3(2):105–118.
DOI: 10.5121/acij.2012.3212.
23. Kumar P, Sharma R, Choudhary SD, Singh SK, Kapse VM. Predictive analysis of novel coronavirus using machine learning model - a graph mining approach. J. Math. Comput. Sci., 2021;11(3):3647–3662.
DOI: 10.28919/jmcs/5775.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/72549