https://doi.org/10.48048/tis.2022.1919

COVID-19 Outbreak Prediction using Additive Time Series Forecasting Model

Karuna Gull[1,*], Suvarna Kanakaraddi[2] and Ashok Chikaraddi[3]

[1]*Department of ComputerScience and Engineering, SG Balekundri Institute of Technology, Belagavi, Karnataka, India*
[2]*School of ComputerScience and Engineering, KLE Technological University, Hubballi, Karnataka, India*
[3]*Computer Applications, KLE Technological University, Hubballi, Karnataka, India*

([*]Corresponding author's e-mail: karunagull74@gmail.com)

Received: 10 January 2022, Revised: 27 February 2022, Accepted: 10 March 2022, Published: 3 November 2022

Abstract

It is no secret that COVID-19 is a hot topic these days. Its spread has engulfed the world. People all throughout the world have suffered because of it. A unique coronavirus epidemic that has swept throughout the globe is examined and analysed in this article. COVID-19 outbreaks in various places are analysed using machine learning models, which are visualised using charts, tables, graphs and predictions depending on the available data. For prediction models, the time series forecasting package (PROPHET) is utilised as part of machine learning. The work done may aid in the development of some novel concepts and ways that can be utilised as recommendations to prevent the spread of COVOD-19.

Keywords: COVID-19, MoHFW, PROPHET, ARIMA model, Machine learning

Introduction

Coronavirus 2019 (COVID-19) is a transmittable disease caused by coronavirus 2 that causes severe lung infections (SARS-CoV-2). Most people who are sick with COVID-19 will have moderate to severe symptoms and will recover without any treatment. The COVID-19 virus is primarily transmitted by droplets that are formed when a diseased person coughs, sneezes or urinates. These droplets are very heavy to float in the air and then fall to the ground quickly or to other objects. If you are very close to someone with COVID-19, you can be infected by inhaling the virus, or by touching a dirty area and then touching your eyes, nose or mouth.

In the month of December 2019, Covid 19 was created in Wuhan city, China [1], and since then it has spread worldwide, leading to an ongoing epidemic. The original instance of the COVID-19 epidemic in India was recorded on 30th January 2020, when a Wuhan student arrived in Kerala. On March 9, 2020, the first cases of COVID-19 were confirmed in the Indian state of Karnataka. The Ministry of Health and Family Welfare (MoHFW) said among the total of 332,424 cases, 169,798 recovered (including one migration), and 9,520 have died in the country as of 15 June 2021. On June 10, India's repatriation skipped cases that were active for the first time, obtaining the total infections by 49 %.

As the coronavirus outbreak spreads throughout the country, the question that we as Indians are attempting to answer is, will India be able to combat this pandemic, or will we experience another Italy/South Korea. The authors stumbled upon a simple problem statement while looking for answers to these problems. As a result, the proposed system is concerned with the creation and implementation of a model to anticipate the COVID-19 outbreak [2] utilizing a time series forecasting tool (PROPHET) and Machine Learning ideas [3,4].

As an extension of the suggested system, it forecasts how the virus would spread across different countries and regions [5], as well as how the infection will expand in the next 7 days. So that the government and people of India can be prepared to take or execute proactive control measures to limit the impact of coronavirus disease in 2019 (COVID-19).

Data analysis models

The objective of the time series investigation is to produce reliable and accurate statistics, which are used to predict future prices. With past and present investigation, time sequence models attempt to predict upcoming values. Timeline analysis is distinguished by the fact that sequential detection is not usually self-determining and that the analysis must use methods to detect existing patterns in the series.

Accordingly, among the strategies of the time series, the ARIMA model [6] is used to forecast the future state of coronavirus (COVID-19).

Autoregressive Integrated Moving Average (ARIMA) modeling

The most widely used time series model is [4] because it calculates temporal trends, specific modifications and random breaks. It can be used with any type of data, including non-stationary data (where there is no systematic change in mean or variance and the time difference has been removed) [7]. In practice the time series is static, hence [8] recommend removing any non-stationary data sources before removing the time series. Applying regular differencing and log transformation to the source time series is the most popular method for creating stationary time series. When a series is differentiated d times, it is said to follow an autoregressive integrated moving average process, denoted by ARIMA (p, d and q), and can be labeled as follows:

$$\emptyset(\beta)\nabla^{d}X_{t} = \theta(\beta)W_{t} \tag{1}$$

Autoregressive operator can be written as: $\emptyset(B) = 1 - \emptyset_1 B - \emptyset_2 B^1 - \dots . -\emptyset_p B^p$

Moving average operator $\left(\theta(B)\right)$: $\theta(B) = 1 + \theta_1 B + \theta_2 B + \dots . +\theta_q B$

Differentiating operator $(\nabla^d) = (1-\beta)^d$, is the expression of d^th^ consecutive differentiating in order to make series stationery. W_t denotes a Gaussian white nose series with a mean of 0 and a variance of $(б^2{}_{w})$.

This study investigates [9] the context of the use of a linear parametric model structure prediction approach. The Autoregressive (AR) model is the most basic and extensively used model structure. The current output z (t) in AR model is expressed by the preceding values and parameters t - p, as defined in (2), where t is the time and p is the parameter order.

$$Z_t = -\alpha_1 Z_{t-1} - \alpha_2 Z_{t-2} - \dots - \alpha_k Z_{t-p} + \varepsilon \tag{2}$$

where αZ^{-1} is given by

$$\alpha(Z^{-1}) = 1 - \alpha_1 Z^{-1} + \dots + -\alpha_k Z^{-p}$$
$$Z_t = \varepsilon + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} - \dots + \beta_k \varepsilon_{t-q}$$
$$\beta(Z^{-1}) = 1 - \beta_1 Z^{-1} + \dots + -\beta_k Z^{-q}$$

By combining both MA and AR we obtain a more complementary approach called the ARMA model described in (3).

$$\mathrm{Z_t} = -\alpha_1 \mathrm{Z_{t-1}} - \alpha_2 \mathrm{Z_{t-2}} - \dots - \alpha_k \mathrm{Z_{t-p}} + \varepsilon + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} - \dots + \beta_k \varepsilon_{t-q} \tag{3}$$

The Autoregressive Integrated Moving Average (ARIMA) is an additional model that incorporates differences at least once. In this paper, an ARIMA formula model is proposed (4). As previously stated, this model has many examples of effective predictions in the literature that do not matter.

$$\mathrm{Z_t} = (1 - \mathrm{Z_{t-d}}) - \alpha_1 \mathrm{Z_{t-1}} - \dots - \alpha_p \mathrm{Z_{t-p}} + \varepsilon + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} - \dots + \beta_k \varepsilon_{t-q} \tag{4}$$

Three orders are used to identify the ARIMA model: p of AR, q of MA, and number of variables d Parameters.

The data were exposed using the Augmented Dickey Fuller (ADF) test prior to estimating the ARIMA model parameters, which is a negative H0 hypothesis that the time series is static. The findings of the ADF experiment demonstrated that time series data were not static ($p > 0.05$). Following the application of the first difference, d (0), the *p*-value obtained was less than the significance level ($p = 0.05$), and the ADF values were fewer than any critical values, suggesting that the erroneous hypothesis was rejected. Chosen in accordance with the Akaike Information (AIC) value principles supplied by (5).

$$\mathrm{A}\left|\mathrm{C_{p,q}} = \frac{-2\ln(maximized\ likehood)+2r}{\mathrm{n}} \approx \ln(\hat{\sigma}_a^2)\right) + r\frac{2}{n} + constant \tag{5}$$

where n is the number of data observations, r = p + q + 1, and $\hat{\sigma}_a^2$ is a high probability forecast.

We experimented with several p and q parameter values ranging from 0 to 5, whereas d was set to 1 based on the ADF test. We discovered the best ARIMA model for providing the most affordable AIC ARIMA costs (2,1,1).

Predictive modeling

The dynamic Susceptible Infection and Recover (SIR)) model [2,9] was used to make this prediction, which is a well-validated model that is highly effective in studying the transmission of infectious diseases because the vital role of exposure or communication was a major component in disease transmission. Because all SEIR characteristics remained stable over time, this model is intended to employ robust statistics to assess the original situation and identify appropriate conditions.

Testing under potentially harmful settings, such as interaction with a person infected with COVID-9, can discover new instances. The model begins when the first affected person detects and spreads the sickness to others. It could be 1 or more people, depending on a good person's community work. COVID-19. A person infected with COVID-19 can spread the disease to others either individually or, in many circumstances, simultaneously. A multitude of factors, such as health problems, personal hygiene and environmental health precautions, might have an impact on contacts.

The number of infected cases is approximated based on the number of cases received. The referral rate can be determined using data from the Department of Public Health. The characteristics of COVID-19 disease, such as incubation and infection periods, were also investigated. Based on associated information such as health data, recovery rates and fatality rates were also estimated. As the first study, the causal loop was translated into a computer picture utilizing a STELLA Software license to construct an equation, such as pricing, infection, and recovery. Until far, the model has been focused on the number of cases infected as a result of intervention measures.

Furthermore, the core method of detection is utilized to know the previously acquired instances in order to anticipate infected cases in three stages: a 2, 3, or 5 day journey average. The pattern or inclination of the discovered examples is utilized to foretell future events [10]. The RMSE (Root Mean Square Error) method was used to determine the equality of each of the models listed above.

Exponential model

It is believed that most epidemics are rampant during the outbreak. I(t) is the number of cases "found to be infected", and time measured in days is denoted by t [11].

$$\mathrm{I(t)} = \mathrm{I_0 e^{rt}} \tag{6}$$

$$\frac{\mathrm{dl(t)}}{\mathrm{dt}} = \mathrm{rl(t)} = \mathrm{rl_0 e^{rt}} \tag{7}$$

When r denotes the rate of increase, I_0 denotes the constant estimated by data fitting [10].

The writer attempted to customize a descriptive model based on parameters (y 14 39.64 * exp (0.1887 * x)) as used by [11] to predict cases of infection and death, which was applicable elsewhere at the beginning of the spread, but the descriptive model does not work long-term speculation after its first phase epidemic. To facilitate modeling of infectious illness statistics, the most frequent types of Susceptible Infectious Recovered (SIR) models are used.

$$\frac{dS(t)}{dt} = -\frac{\beta}{N} S(t)l(t) \tag{8}$$

$$\frac{dl(t)}{dt} = \beta S(t)l(t) - \gamma l(t) \tag{9}$$

$$\frac{dR(t)}{dt} = \gamma l(t) \tag{10}$$

where S(t) is the quantity of persons influenced, I(t) is the quantity of infected people, and R(t) is the quantity of individuals recovered; b is the transfer rate per infected person, g is the recovery rate, N is the quantity of people, and

$$N = S(t) + 1(t) + R(t) \tag{11}$$

The basic reproduction number is as follows: $R_0 = -\frac{\beta}{\gamma}(1 - \frac{I_0}{N})$. SIR models used to create predictions by [11]. These types can include a high number of parameters and assumptions, increasing the possibility of confounding assumptions with earlier assumptions, failing to examine the equality of visual data, and leading to a wide range of forecast space [12].

Logistic model

Belgian mathematician Pierre Verhulst created the Logistic model (1838). An exponential growth followed by a gradual slowing and a scenario is depicted by the Logistic model.

$$\frac{dC(t)}{dt} = rC(t)(1 - \frac{C(t)}{K}) \tag{12}$$

$$C(t) = \frac{KC_0}{C_0 + (K - C_0)e^{-rt}}$$

where C(t) is the total number of infections, r is the highest growth rate, K is the highest population growth rate, and is called the carrying capacity. C_0 is C(t) where t ¼ 0.

Crow-AMSAA model

Introduction of Crow-AMSAA

GE Motors Division's James T. Duane conducted a fiduciary analysis by looking at the product system's increasing failure rate during the development test. A cumulative failure occurs when a log-paper is compared to the development period [13]. Larry Crow created the Army Material Systems Analysis Activity (AMSAA) model in 1974, which was a major step forward in Duane's career (AMSAA). The Duane model could be expressed as a non-standard Poisson (NHPP) model with the Weibull function, according to [14,15].

The total number of cases of infection or death N (t) can be stated using the Crow-AMSAA model as follows.

$$N(t) = \lambda t^{\beta} \tag{13}$$

where the period is measured in days, l and b are certain times, will be defined later.

Accumulated event logarithm N (t) compared to logarithm t time, measured in line plan days. By taking the natural logarithms of the Eq. (14).

$$\text{Ln } N(t) = \text{Ln }(\lambda) + \beta \text{Ln}(t) \tag{14}$$

The model intensity function

$$p(t) = \frac{dN(t)}{dt} = \lambda\beta t^{\beta-1} \tag{15}$$

The event rate of cumulative equation separated by t it is

$$C(t) = \lambda t^{\beta-1} \tag{16}$$

The collection of events N(t) ¼ ltb, where r (t) is known as the probability ratio, reveals strengthening activity (ROC). The scale parameter, l, grips on the y (N) t axis in Eq. (17), where t ¼ 1, (ln (1) ¼ 0); the slope b is understood similarly to the Weibull scheme, If b is greater than 1, the transfer rate increases and becomes much faster; if b is less than 1, the transfer rate decreases and becomes much slower; if b equals −1, the process is known as the Homogenous Poisson Process (HPP); if b does not equal 1, the process is known as the Non Homogenous Poisson Process (NHPP)

The Weibull Distribution was established and it is frequently used in the field of reliability engineering for failure data analysis. The sort of failure is reflected in the slope of the Weibull b structure. The CA model is also known as the "Bull Power System" (WPP). Slope interpretation b is comparable to Weibull analysis. Weibull employs failure time, whereas CA employs accumulation times. The Weibull distribution only allows for 1 failure mode at a time, but the CA distribution allows for state mixes.

The piece-wise Crow-AMSAA

Piece wise Crow-AMSAA is applied for each segment of the data. Take point Ci (i ¼ 1,… k-1) (in days), k is the number of segments, the piece wise Crow-AMSAA [13] will be:

$$N_1(t) = \lambda_1 t^{\beta 1} \quad (t \le C_1)$$

$$N_2(t) = \lambda_2 t^{\beta 2} \quad (C_1 < t \le C_2)$$

.
.
.

$$N_{k-1}(t) = \lambda_k t^{\beta k-1} \quad (C_{k-2} < t \le C_{k-1})$$

$$N_k(t) = \lambda_k t^{\beta k} \quad (C_{k-1} < t) \tag{17}$$

The 2 curves interpret at time Ci,

$$N_i(C_i) = N_{i+1}(C_i)$$

And the model parameters has the relationship

$$\lambda_{i+1} = \lambda_i C^{\beta i - \beta i+1}$$

For the system starts from time 0, and has one change point C, the log-likelihood function is as follows [16].

$$L = N\,Ln(\lambda_1) + N_1\,Ln(\beta_1) + N_2\,Ln(\beta_2) + N_2(\beta_1 - \beta_2)Ln(C) + (\beta_1 - 1)\sum_{n=1}^{N_1} Ln(t_i) + (\beta_2 - 1)\sum_{1=N_1+1}^{N} Ln(t_i) \tag{18}$$

The solution for the model parameters are:

$$\lambda_1 = \frac{N}{C^{\beta_1} - \beta_2\, T^{\beta_2}}$$

$$\beta_1 = \frac{N_1}{N_1 Ln(C) - \sum_{i=1}^{N_1} Ln(t_i)}$$

$$\beta_1 = \frac{N_1}{NLn(T) - \sum_{i=N_1+1}^{N} Ln(t_i) - N_1 Ln(C)}$$

where N is the entire quantity of infected or dead people. N_1 is the number of cases of deaths or deaths in category 1, N_2 is the quantity of cases of deaths or deaths in category 2. T is the end time in days. Solved the C_i iteratively [16] using the heuristic method and plotted the sicked cases or death vs. time (days) in Crow-AMSAA logarithm scale. 2.) From the plot, identify the range for C_i, denoted as [C_i _min, C_i _max]. 3.) Set C_{ij} ¼ C_i _min þ jDC_i, calculate the MLE solution for l^i, b^i, b ^ iþ1using þ C_{ij}. 4.) Calculate the log likelihood value using l^i, b^i, biþ^1, 5.) Set j ¼ j þ 1 and repeat step 3 and 4 until C_i reach C_i _max. The solution of l^i , b^i , biþ^1and the value of C_{ij} that provide the largest likelihood value will be the ML solution. Consider the general regression model [4],

$$y = X\beta + \varepsilon \tag{19}$$

where y is an n×1 vector of response variable, X is a known n×p full rank matrix of predictor or explanatory variables, β is an p×1 vector of unknown regression parameters, ε is an n×1 vector of errors such that E(ε) = 0 and V (ϵ) = σ^2, is an n×n identity matrix. The ordinary least squares estimator (OLS) of β in (1) is defined as:

$$\beta = (S)^{-1}X'y,$$

where $S = X'X$

XoX is a design matrix. The measurement performance is excellent if there is no violation of any of the concepts of the classic Linear regression [17,18]. Predictions include: Irregularity of the error term, unrelated error name, variable prediction order and so on. Variables of interest include confirmed cases such as response variable, travel history and contact as the regressors. The regression model is as follows:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i \tag{20}$$

where y represents the COVID-19 certified cases, the X1 represents the pre-closure travel history and the X2 represents the number of contacts made with the COVID-19 patient. Designed a correlation analysis to investigate the relationship between regressors and response variable.

Regression models

Machine Learning (ML) algorithms [6,19] include regression model analysis [20]. There are several varieties of regression, including linear and nonlinear forms, as well as Multiple Linear Regression. Nearly few of these types use parametric or non-parametric mathematical acquisition approaches. In epidemiologic research, the regression process is a form of modeling tool used to quantify the association between groups of variables. Regression analysis approaches are a collection of machine learning (ML) methods that permit us to predict continuous variable results (Y) based on 1 or more variables (X). It requires a direct relationship between the results and the forecast's variability. Several regression analysis techniques have been used to forecast the accumulated unconfirmed COVID-19 within the next year (15 days).

Exponential regression model

It is utilized to form pandemic habits when it begins to slow down and accelerate swiftly without being detected, or when the decline begins quickly and the speed lowers to near 0. This model is represented by the number:

$$y = a_1 e^{\alpha X} \tag{21}$$

where α_1 and α_2 are called the parameters of regression analysis.

Model of polynomial regression the polynomial terminal alters a linear curve model while remaining a simple model. In such instances, polynomial quadratic, third-degree, fourth-degree, fifth-fifth-degree and sixth-degree models are used. Equation: provides an n^{th} order polynomial for single variables.

$$y = a_1.X + a_2.X^2 + a_3.X^3 + a_3.X^3 \ldots + a_n.X^n + \varepsilon$$

where (n = 2, …, 6) represents the degree of the models. The coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ $a_1, a_2, \ldots, a_n$ are called the parameters of regression analysis.

Logit growth regression model

A logit model, often known as a (logistic model), is a mathematical technique derived from machine learning. The logit model is a regression model that is often used in epidemiology mathematical models to calculate the rate of spread of this epidemic.

The model experiences rapid growth at the start of the epidemic, followed by a constant increase, and finally a decreased growth rate. Following Eq. (22) depicts the logit model as follows:

The natural growth equation:

$$\frac{1}{C}\frac{dI}{dt} = C_r\left(1 - \frac{C}{K}\right) \tag{22}$$

where A = K-C0C0A = K-C0C0 and t = 0 = t = , assuming A, k > C_0 K > C0.

$$C = \frac{Ke^{C_r t}}{A^{C_r t} + A} \tag{23}$$

Hence, if C is an accumulated number of cases, Cr defined as the rate of infection cases, K is the final epidemic size, t is the time, dCdt is the growth rate reaches its maximum $dC_2dt_2 = 0$ $dCdt_2 = 0$ $dCdt_2 = 0$. To fit the maximum number of confirmed cases (peak number of cases) of the infected population C_{Peak} and coefficient, t_{Peak} and $\frac{dC}{dt_{Peak}}$ are defined by the formulas.

$$t_{Peak} = \frac{\ln A}{C_f} \tag{24}$$

$$t_{Peak} = \frac{\ln A}{C_f} \tag{25}$$

If $C_1, C_2, \ldots, C_f$ represent the number of cases at times $t_1, t_2, \ldots, t_{final}$ then the final size predictions of the epidemic based on these data are $K_1, K_2, \ldots, K_f$ the predicted final epidemic size presented by the Eq. (26) by [21].

$$K = \frac{K_{f+1} \ K_{f-1} - K_f^2}{K_{f+1} - 2K_F - K_{f-1}} \tag{26}$$

Because of the model's nonlinearity, the logit model described in equation comprises 3 coefficients: K, C_r and A, which need be evaluated using regression analysis. Officials all across the world are using several COVID-19 epidemic prediction models to help them make informed decisions and implement effective control measures. Simple epidemiological and statistical models have garnered significant attention from authorities and are popular in the media among the standard models for COVID-19 global pandemic prediction. Due to a high level of uncertainty and a lack of critical data, standard models have proven insufficient accuracy for long-term prediction. Despite various attempts in the literature to address this issue, current models must enhance their essential generalization and resilience capacities.

It is vital to have access to accurate outbreak prediction models in order to gain insight into the likely spread and repercussions of infectious illnesses. Prediction models are used by governments and other legislative bodies to propose new policies and evaluate the success of existing programmes. This system compares machine learning and soft computing methodologies for predicting the COVID-19 epidemic. The model is written in Python and is based on the data provided in order to evaluate the model's impact thus far, to investigate the COVID-19 pandemic across different sites, and to estimate the number of confirmed, recovered and death cases.

Proposed system

COVID-19 Outbreak prediction model to examine the current situation in India and Karnataka, compare trends to Italy and South Korea, and forecast confirmed, recovered, and death cases using Prophet. To obtain an accurate prediction value, the prediction is done successfully with the help of Time series forecasting tools.

The various data samples considered for the planned COVID-19 Outbreak prediction effort are given as Covid cases in India such as covid_19_clean_complete, Indian Coordinates, per_day_cases, time_series_covid19_confirmed_global,time_series_covid19_recovered_global and time_series_covid19 _deaths_global. The time series forecasting tool prophet is then used to forecast the virus's spread over the next 7 days. The use of a time series forecasting tool can estimate the accuracy of the given data to up to 95 %. Forecasting tools handle errors carefully, for example, if the values in the dataset are in unusual characters, it throws an exception.

Architecture of COVID-19 predictor

Import the COVID-19 dataset and extract it. Then it must be preprocessed. Proceed with the preprocessed data segmentation into training and testing. For the evaluation procedure, apply an ML Algorithms to the constructed model. Then, using the designed model, repeat the processes for predicting the testing dataset (**Figure 1**).

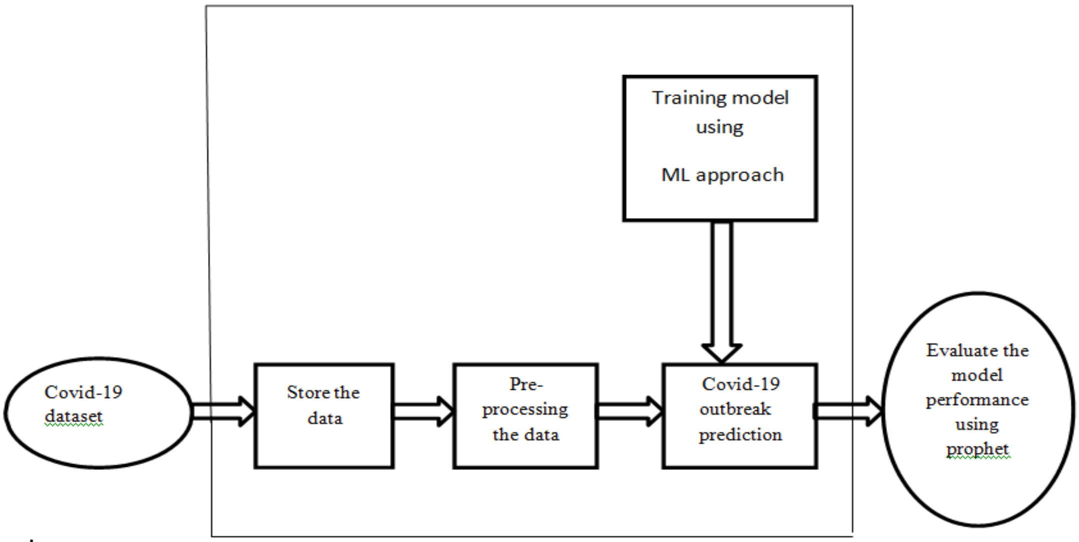

Figure 1 Block diagram of classification.

Methodology

To get insight into the expected spread and consequences of infectious illnesses, effective outbreak prediction models are required. Prediction models are used by governments and other legislative bodies to suggest new policies and analyze the performance of existing ones. In order to anticipate the COVID-19 epidemic, this system compares machine learning and soft computing models. The model is written in Python to examine its impact thus far and to analyze the COVID-19 epidemic across multiple locations, as well as to forecast the number of confirmed, recovered and death cases based on the existing data. Predictions are made with the help of time series forecasting techniques.

Prophet

Time series forecasting tool

The use of Supervised Training. Using historical data training, create a model to predict value from date. If you feed future model dates, it will produce the expected values of those days. Traditional approach with time series forecasting: Create a model that predicts day value from past N days (**Figure 2**).

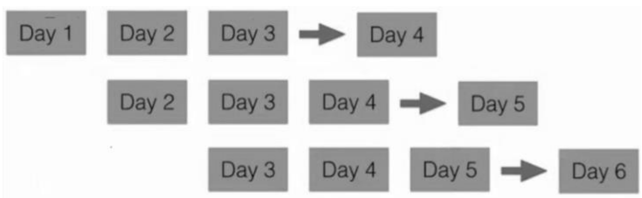

Figure 2 Traditional approach.

Problems with traditional time series model: The time gap between data points should be consistent in the data set. Day with NA in the data set are not allowed. The season of several years (Week and Year) is difficult to manage. Parameter tweaking must be performed by a professional. Expertise is required.

Prophet is an open source forecasting package created by Facebook's research team. It can be used to model time series and forecast trends in the future. This programme enables both professionals and non-experts to generate high-quality forecasts with minimal effort, i.e. it was designed for ease of use without specialist understanding of time series forecasting or statistics.

Benefits of prophet approach

Uneven time intervals between data are not an issue. It's not an issue to spend the day with NA. Seasonality with multiple periods (Week and Year) is automatically handled. By default, it works properly. The parameters are simple to understand.

The Prophet uses a model of a rotten time series with three main elements of the model: Trends, seasons and holidays. They are included in the following Eq. (27):

$$y(t) = g(t) + s(t) + h(t) + \varepsilon t \tag{27}$$

where, $g(t)$: Piecewise linear or logistic growth curve for modelling non-periodic changes in time series

$h(t)$: Implications of holidays (given by the user) on irregular schedules

$s(t)$: Periodic changes (e.g. Weekly/yearly seasonality)

εt: Error term accounts for any odd changes that the model does not account for.

Seasonal effects $s(t)$ are approximated by the function given in the Eq. (28):

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\frac{2\pi nt}{P} + b_n \sin\frac{2\pi nt}{P} \right) \tag{28}$$

Here P is the yearly period data. The seasonal model is predicted with Parameters $[a_1, b_1, ..., a_n, b_n]$ for a given N. Prophet has a built-in feature, which allows inputs of customized recurring events, which is a built-in feature in it.

Modules imported to carry out the work

Pandas: Wes McKinney created Pandas, a high-level data manipulation tool. It is based on the Numpy package, and its primary data structure is known as the DataFrame. DataFrames are used to store and manipulate tabular data in the form of rows of observations and columns of variables.

Matplotlib: It is a Python package that allows you to create static, animated and interactive visualisations. Matplotlib makes simple things simple and difficult things possible.

Seaborn: It is a matplotlib-based Python data visualisation package. It offers a high-level interface for creating visually appealing and informative statistical visuals.

Plotly: The Python graphing package from Plotly creates interactive, publication-quality graphs.

Folium: The Python-manipulated data on interactive leaflet map is very simple to visualize in folium.

Prophet: It's a well-known time series software suite. It works effectively with data that has substantial seasonal impacts as well as data from multiple seasons. It is a method of forecasting time series data using an additive model. It is extremely user-friendly and customizable, requiring no effort to set up.

Results and discussion

Analyses of the propagation and effect of the new coronavirus are performed here. This may assist to explain the storm's quick expansion. The developed model analyses COVID-19 across several regions, visualises it in the form of charts and tables, and predicts the confirmed cases, patients diagnosed, and death cases in number based on the information provided.

Authors have discussed corona virus cases in India, including increased cases, confirmed cases and recovered cases in contrast to Italy and South Korea in the **Figures (3) - (7)**.

As of May 3rd, 2021, **Figure 3** shows a table of varied conditional status of corona virus effects in various Indian states.

	S. No.	Name of State / UT	Total Confirmed cases	Active	Cured/Discharged/Migrated	Deaths	Total cases
0	1	Andhra Pradesh	1583	1062	488	33	1583
1	2	Andaman and Nicobar Islands	33	1	32	0	33
2	3	Arunachal Pradesh	1	0	1	0	1
3	4	Assam	43	9	33	1	43
4	5	Bihar	517	396	117	4	517
5	6	Chandigarh	97	77	19	1	97
6	7	Chhattisgarh	57	21	36	0	57
7	8	Delhi	4549	3123	1362	64	4549
8	9	Goa	7	0	7	0	7
9	10	Gujarat	5428	4096	1042	290	5428
10	11	Haryana	442	192	245	5	442
11	12	Himachal Pradesh	40	1	34	2	40
12	13	Jammu and Kashmir	701	406	287	8	701
13	14	Jharkhand	115	85	27	3	115
14	15	Karnataka	614	295	293	25	614
15	16	Kerala	500	95	401	4	500
16	17	Ladakh	42	25	17	0	42
17	18	Madhya Pradesh	2837	1883	798	156	2837
18	19	Maharashtra	12974	10311	2115	548	12974
19	20	Manipur	2	0	2	0	2
20	21	Meghalaya	12	1	10	1	12
21	22	Mizoram	1	1	0	0	1
22	23	Odisha	162	101	60	1	162
23	24	Pondicherry	12	6	6	0	12
24	25	Punjab	1102	964	117	21	1102
25	26	Rajasthan	2886	1459	1356	71	2886
26	27	Tamil Nadu	3023	1614	1379	30	3023
27	28	Telengana	1082	508	545	29	1082
28	29	Tripura	16	14	2	0	16
29	30	Uttarakhand	60	20	39	1	60
30	31	Uttar Pradesh	2645	1848	754	43	2645
31	32	West Bengal	963	762	151	50	963

Figure 3 Analyzing the present condition in India (As of 3rd May 2021).

As of May 3rd, 2021, **Figure 3** shows a table of varied conditional status of corona virus effects in various Indian states.

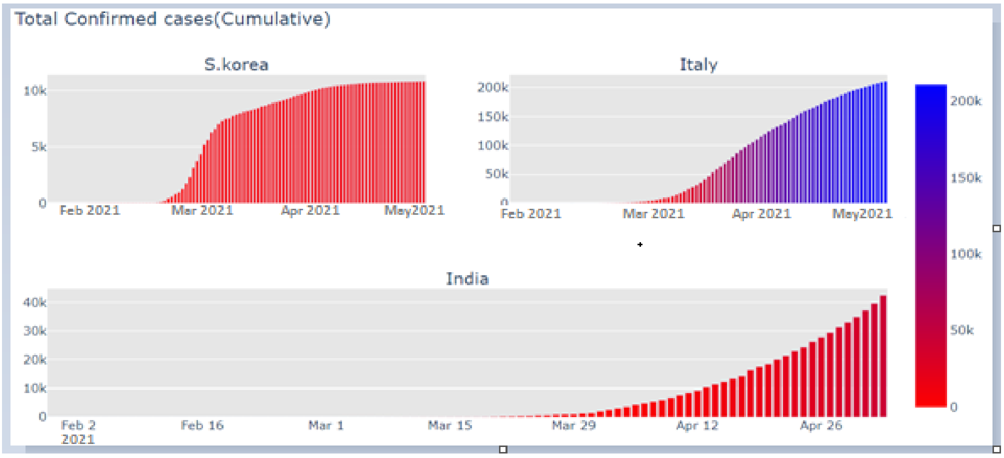

Figure 4 Comparison between the raise of cases in S. Korea, Italy and India.

Figure 4 illustrates a graph of the increase in coronavirus cases in several nations such as South Korea, Italy and India, and a comparative research can be done very quickly using the graphs.

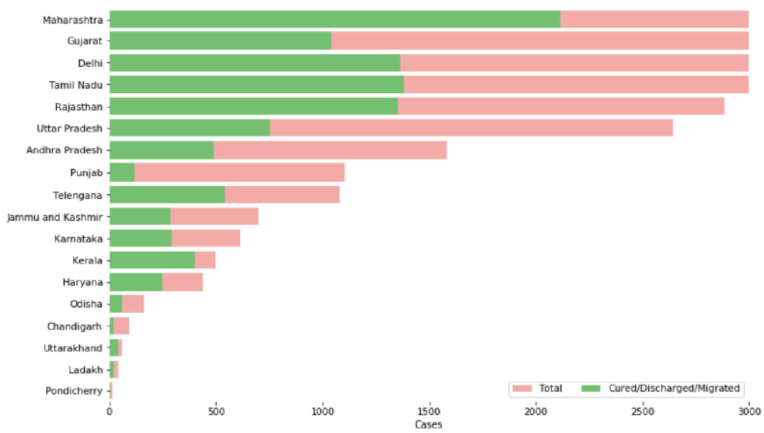

Figure 5 Confirmed vs. recovered figures (India).

The graph of recovered versus confirmed cases in India is shown in **Figure 5**. It allows us to assess the recovery rate for confirmed instances.

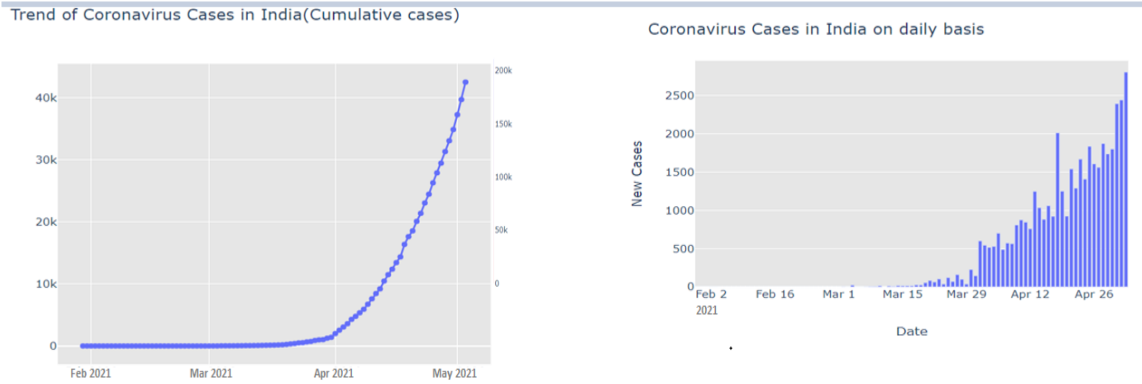

Figure 6 How coronavirus cases are rising in India.

Figure 6 depicts a graph of the increase in corona cases in our country on a fortnightly and monthly basis. This makes it easier to figure out what the next step is in breaking the chain.

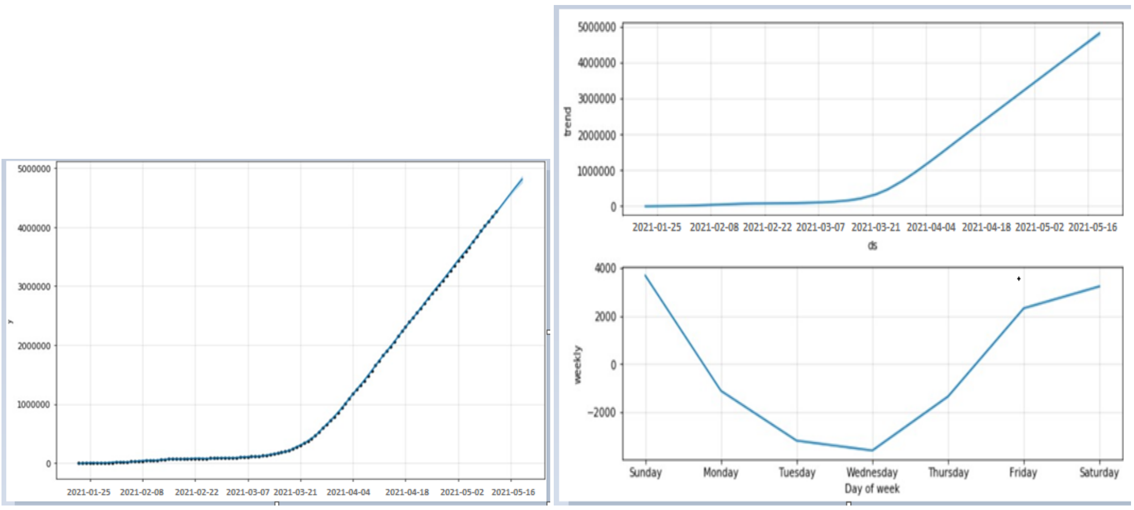

Figure 7 Forecasting confirmed NCOVID-19 cases worldwide with Prophet.

Figure 7 displays a graph that uses prophet, a time series software suite, to anticipate or forecast confirmed cases on a daily or fortnightly basis. Now we move on to Karnataka state and analyze the information we have gathered about the coronavirus. The results of analysis are depicted in **Figures (8) - (12)**.

	S. No.	Name of District	Total Confirmed cases	Active	Cured/Discharged/Migrated	Deaths	Total cases
0	1	Bagalkot	77	10	66	1	77
1	2	Ballari	47	32	14	1	47
2	3	Bangalore Rural	15	8	6	1	15
3	4	Bangalore Urban	336	153	172	10	336
4	5	Belagavi	147	54	92	1	147
5	6	Bidar	130	99	27	4	130
6	7	Chikkaballapura	136	113	20	3	136
7	8	Chikmagalur	18	16	2	0	18
8	9	Chitradurga	39	34	5	0	39
9	10	Dakshina Kannada	111	75	29	6	111
10	11	Davanagere	150	42	104	4	150
11	12	Dharwad	45	34	11	0	45
12	13	Gadag	35	22	12	1	35
13	14	Hassan	157	127	30	0	157
14	15	Haveri	14	11	3	0	14
15	16	Kalaburagi	253	118	128	7	253
16	17	Kodagu	3	2	1	0	3
17	18	Kolar	22	17	5	0	22
18	19	Koppal	4	4	0	0	4
19	20	Mandya	257	224	33	0	257
20	21	Mysuru	94	5	89	0	94
21	22	Raichur	134	132	2	0	134
22	23	Ramanagara	1	1	0	0	1
23	24	Shivamogga	41	34	7	0	41
24	25	Tumakuru	31	23	6	2	31
25	26	Udupi	177	170	6	1	177
26	27	Uttara Kannada	77	34	43	0	77
27	28	Vijayapur	96	37	54	5	96
28	29	Yadgiri	241	231	9	1	241
29	30	Others*	34	12	21	1	34

Figure 8 Analyzing the present condition in Karnataka (As of 30th May 2021).

Figure 8 shows a table of various conditional status of coronavirus effects in various Karnataka locations as of May 30, 2021.

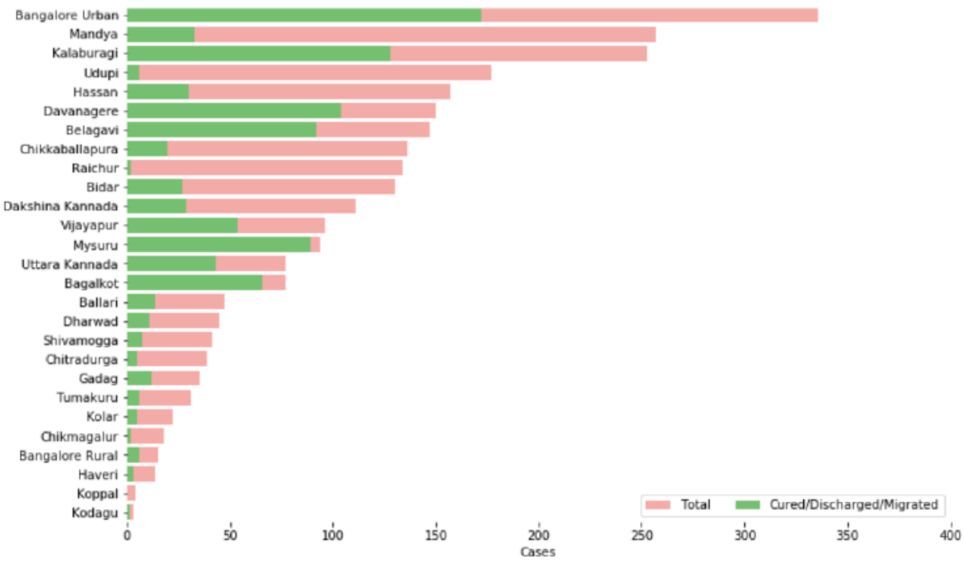

Figure 9 Confirmed vs rcovered cases (Karnataka).

The graph of recovered versus confirmed cases in Karnataka is shown in **Figure 9**. It allows us to assess the recovery rate for confirmed instances.

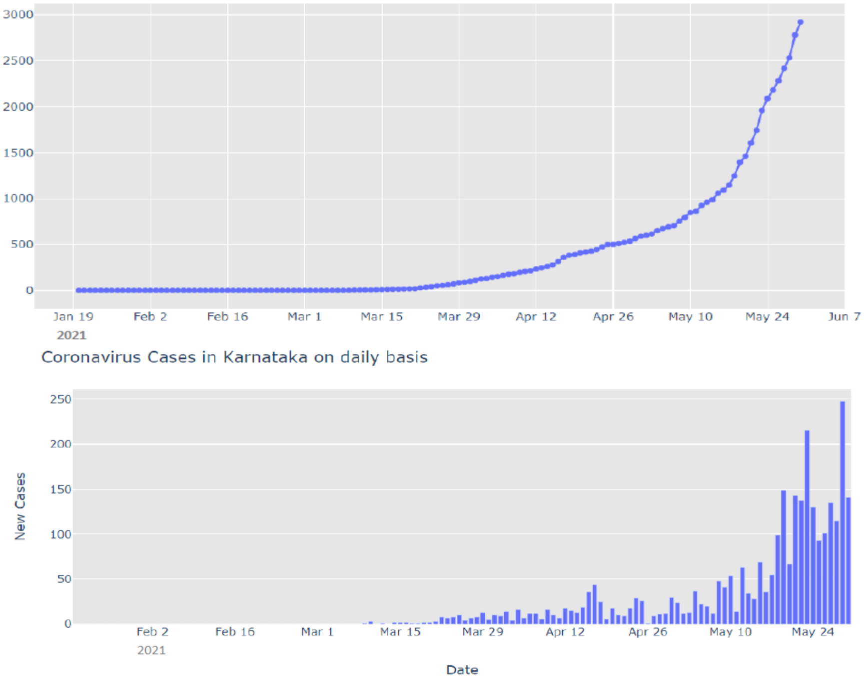

Figure 10 How coronavirus cases are rising in Karnataka.

Figure 10 depicts a graph of the increase in corona cases in our state on a biweekly and weekly basis. This makes it easier to figure out what the next step is in breaking the chain.

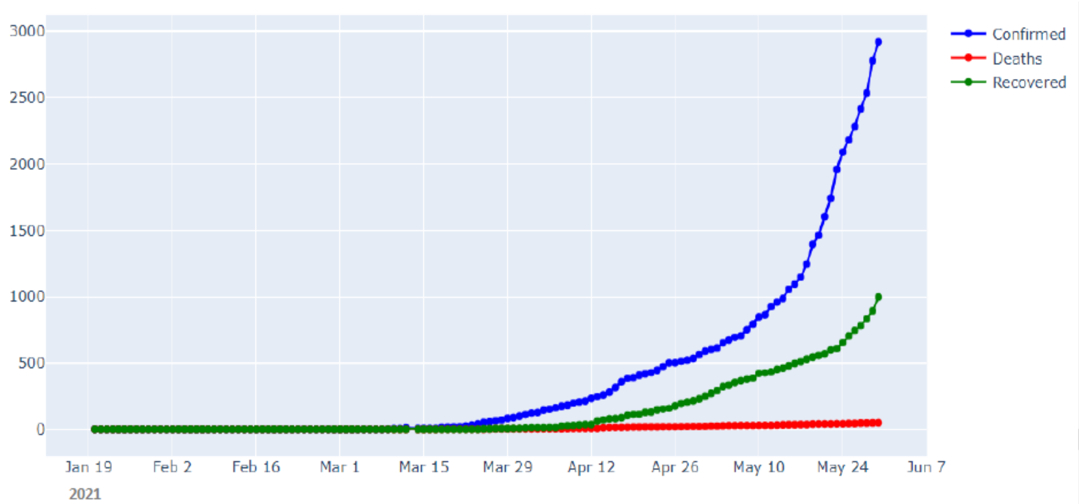

Figure 11 Confirmed vs recovered vs deaths.

Figure 11 depicts a recovery record with confirmed cases and death cases. Located in the state. So that they can find a way to prevent deaths before they happen.

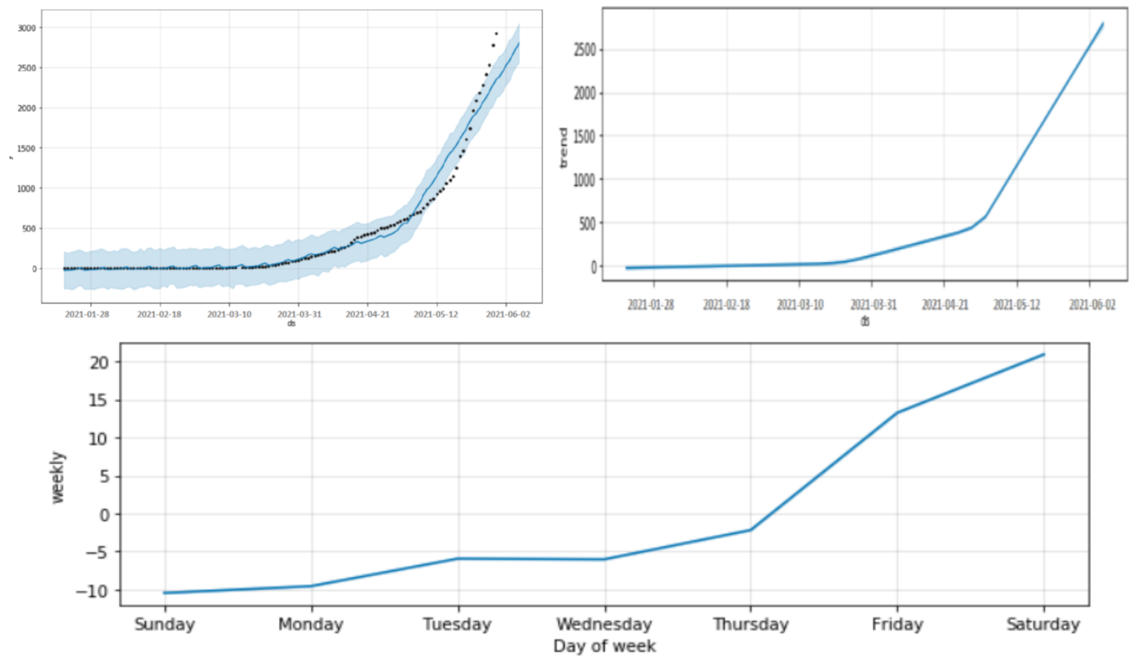

Figure 12 Forecasting confirmed COVID-19 cases in Karnataka with Prophet

Figure 12 shows a graph that uses prophet, a time series software suite, to anticipate or forecast confirmed cases on a daily or fortnightly basis.

Conclusions

The effects of COVID-19 are expected to be high. If this outbreak is not contained India would confront a significant lack of hospitals exacerbating the situation. Work may aid in the development of some novel concepts and ways that can be utilized as guidelines to issue an advice on the role of integrative Medicine in the management of COVID-19. And the pandemic could be contained if the Indian government takes bold forceful actions to establish and extend the country's lockdown. This epidemiological model described here is an attempt to anticipate the future spread of COVID-19 based on the current circumstances so that policy decisions can be made and appropriate actions can be taken.

References

[1] L Zhong, L Mu, J Li, J Wang, Z Yin and D Liu. Early prediction of the 2019 novel coronavirus outbreak in the Mainland China based on simple mathematical model. *IEEE Access* 2020; **8**, 51761-9.
[2] SF Ardabili, A Mosavi, P Ghamisi, F Ferdinand, AR Varkonyi-Koczy, U Reuter, T Rabczuk and PM Atkinson. COVID-19 outbreak prediction with machine learning. *Preprints* 2020, http://dx.doi.org/10.20944/preprints202004.0311.v1
[3] S.G. Kanakaraddi, K.C. Gull, J Bali, A.K. Chikaraddi and S Giraddi. *Disease prediction using data mining and machine learning techniques. In*: S Roy, LM Goyal and M Mittal (Eds.). Advanced prognostic predictive modelling in healthcare data analytics, Springer, Singapore, 2021, p. 71-92.
[4] R Takele. Stochastic modelling for predicting COVID-19 prevalence in, East Africa Countries, *Infect. Dis. Model.* 2020; **5**, 598-607.
[5] World Health Organization, Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov, accessed January 2022.
[6] K Tantrakarnapa and B Bhopdhornangkul. Challenging the spread of COVID-19 in Thailand. *One Health* 2020; **11**, 100173.
[7] RH Shumway. *Applied Statistical Time Series Analysis*. Vol 1. Prentice-Hall, 2010.
[8] GEP Box and GM Jenkins. *Time series analysis: Forecasting and control*. Holden-Day, San Francisco, 1976.

[9] Y Wang. Predict new cases of the coronavirus 19 in Michigan, U.S.A. or other countries using Crow-AMSAA method. *Infect. Dis. Model.* 2020; **5**, 459-77.
[10] RO Ogundokun, AF Lukman, GBM Kibria, JB Awotunde and BB Aladeitan. Predictive modelling of COVID-19 confirmed cases in Nigeria. *Infect. Dis. Model.* 2020; **5**, 543-8.
[11] S Chopra, P Ranjan, V Singh, S Kumar, M Arora, MS Hasan, R Kasiraj, D Kaur, NK Vikram, A Malhotra, A Kumari, KB Klanidhi and U Baitha. Impact of COVID-19 on lifestyle-related behaviours: A cross-sectional audit of responses from nine hundred and ninety-five participants from India. *Diabetes Metab. Syndrome* 2020; **14**, 2021-30.
[12] X Xie, Z Zhong, W Zhao, C Zheng, F Wang and J Liu. Chest CT for typical coronavirus disease 2019 (COVID-19) Pneumonia: Relationship to negative RT-PCR testing. *Radiology* 2020; **296**, E41-E45.
[13] DB Duane. Significance of Skewness in Recent Sediments, Western Pamlico Sound, North Carolina. *J. Sediment. Petrol.* 1964; **34**, 864-74.
[14] RJ Siegert, KD Taylor, M Weatherall and DA Abernethy. Is implicit sequence learning impaired in Parkinson's disease? A meta-analysis. *Neuropsychology* 2006; **20**, 490-5.
[15] Y Wang. Predict new cases of the coronavirus 19 in Michigan, U.S.A. or other countries using Crow-AMSAA method. *Infect. Dis. Model.* 2020; **5**, 459-477.
[16] S Chen, Q Guo, H Leung and É Bossé. A Maximum likelihood approach to joint image registration and fusion. *IEEE Trans. Image Process.* 2011; **20**, 1363-72.
[17] K Ayinde, AF Lukman, SO Olarenwaju and MO Attah. Some new adjusted ridge estimators of linear regression model. *Int. J. Civil Eng. Technol.* 2018; **9**, 2838-52.
[18] AF Lukman, K Ayinde, SS Kun and ET Adewuyi. A modified new two-parameter estimator in a linear regression model. *Model. Simulat. Eng.* 2019; **2019**, 6342702.
[19] SI Alzahrani, IA Aljamaan and EA Al-Fakih. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. *J. Infect. Publ. Health* 2020; **13**, 914-9.
[20] H Singh and J Dhar. *Mathematical population dynamics and epidemiology in temporal and spatio-temporal domains*. CRC Press, 2018.
[21] CM Bender and SA Orszag. *Advanced Mathematical Methods for Scientists and Engineers I*. Springer New York, 1999.