Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021)
IEEE Xplore Part Number: CFP21OAB-ART; ISBN: 978-1-7281-9537-7

COVID19 Prediction using Time Series Analysis

2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) | 978-1-7281-9537-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICAIS50930.2021.9395877

Aayush Jain
Department of Information Technology
National Institute of Technology
Raipur, India
aajain1301@gmail.com

Tanay Sukhdeve
Department of Information Technology
National Institute of Technology
Raipur, India
tanaysukhdeve007@gmail.com

Himanshu Gadia
Department of Information Technology
National Institute of Technology
Raipur, India
chinugadia@gmail.com

Dr. Satya Prakash Sahu
Department of Information Technology
National Institute of Technology
Raipur, India
spsahu.it@nitrr.ac.in

Satya Verma
Department of Information Technology
National Institute of Technology
Raipur, India
sverma.phd2019.it@nitrr.ac.in

Abstract**-- The ongoing COVID19 pandemic has created havoc all over the world. Millions of lives have been gone and thousands are vulnerable. It has also affected the world economy due to lockdown. So, there is a need to develop a time-series forecasting model for predicting future cases so that necessary precautions can be taken. The aim is to help in coping up with the situation without affecting lifestyle any further. For forecasting, this paper has considered four models: exponential smoothing, ARIMA and SARIMA models, and identifies the suitability of the model for prediction purposes. Also, to incorporate the impact of festive seasons in India to measure the fluctuations in the new cases, SARIMAX model is used. The spread case data of the pandemic collected is of 47 weeks, from 30th January 2020 to 23rd December 2020 for India. Data of the first 45 weeks (90%) is taken for training the model and that of the last 2 weeks (10%) is used for validation purposes. For evaluation purposes RMSE (root mean square error) and MAE (mean absolute error) are taken as parameters for model evaluation. The ARIMA (8,2,1) model performed best among all models based on the RMSE. Considering multiplicative trend and seasonality Triple Exponential Smoothing model gave the best result with respect to MAE.**

Keywords: ARIMA, SARIMA, SARIMAX, time-series, Exponential smoothing, India, COVID19.

I. Introduction

COVID19, which is the significant wellspring of sicknesses going from gentle colds to more intense infections, for example, MERS-CoV and SARS-CoV, as indicated by the World Health Organization (WHO 2020). Another Covid (nCoV) variety has not been found in people before. Contaminations are typically observed as indications of the skin, fever, hack, windedness, and inconvenience relaxing. Many genuine cases, flu, extreme intense respiratory turmoil, organ disappointment, or even demise might be brought about by contamination. Reconnaissance and early notification were significant for the counteraction of irresistible illness flare-ups. In this way, creating community health models and making figures are valuable for the counteraction and the board of COVID-19[1]. Because of their effect on the general wellbeing framework, the expectation of sicknesses is significant as exact as could reasonably be expected. Simulated intelligent models are normally used to estimate epidemiological time patterns throughout the years to guarantee this exactness.

ARIMA model is a statistical analysis model which works on time-series to get a grasp of data and to predict future trends. They are broadly used for overwhelming disease assessment [1]. To evaluate, decipher and react to any illness scourge, particularly in situations like Covid 2019 (COVID19), topographical information is significant [2].

Besides, great contact with other helping associations and individuals is needed to guarantee a strong reaction. Time empowered guides exhibit how microorganisms spread after some time and where wellbeing organizers or heads might need to go for the activity. Coronavirus impacts affect other segments too, for example, the old and the hidden medical conditions [3].

The objective of this study is to understand the pandemic curve in India and to build a model which is susceptible to the dynamic nature of Covid outbreak cases through quantitative study and by the means of exploring data of the epidemic spreading using Exponential Smoothing Methods (Double Exponential or Holt's linear trend method and Triple Exponential or Holt-Winters' additive method) and ARIMA. ARIMA and Exponential Smoothing are preferred because these predictions and forecasts on time series data for COVID-19 cases are by these techniques are upto the mark , with high accuracy. In this critical situation, the motivation behind this article is to investigate and think about the prescient potential in the feeling of the aggregate week by week determining COVID-19 cases in India utilizing AI relapse and factual models[4]. What's more, the proposed work examined the spatial-fleeting example of COVID-19 dissemination at the territorial level.

II. Literature Review

Last couple of decades have witnessed a lot of research aiming on statistical issues that may give a clue to an effective detection of the outbreak of contagious diseases. The huge challenges that researchers face are in detection of outbreak in early stages and feasible evolution of disease so that one can take appropriate preventive and precautionary measures.

From various past studies, it is evident that time series models like exponential smoothing, ARIMA, and SARIMA worked efficiently and gave quite satisfactory results for COVID-19 prediction. Many authors have done research work on forecasting of the COVID19 virus infection. All the previous research concluded the ARIMA model as the best one for forecasting purposes [1,2,5-9].Alok Kumar Sahai et al. [5] forecasted COVID19 cases for five countries Russia, Brazil, USA, Spain and India. Dataset for the study is taken from the worldometers. Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE) were considered as performance measures.

Vasilis Z. Marmarelis et al. [6] worked on a dataset made available by John Hopkins University. The author forecasted the total number of infection cases using the Riccati Modules (RM) model.

Andi Sulasikin et al.[7] predicted the COVID19 cases using the data set available at Jakarta website (https://corona.jakarta.go.id/id). The author considered three methods viz. Holt’s method, Holt-Winters method and ARIMA for the prediction. Author reported the ARIMA model as the best one among others. Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) were used as the evaluation parameters.

Hadeel I. Mustafa et al. [8] did forecasting based on a dataset published by the Iran health ministry regarding the COVID19 breakout. ARIMA (2,1,5) model gives the best prediction results with the performance evaluation being done by MSE and MAE.

W. Regis Anne et al. [9] forecasted with ARIMA model and ARIMA (1,2,2) stood best among others considering the Akaike Information Criteria (AIC) value. The performance of models was measured by RMSE and Mean Absolute Error (MAE).

Saleh I. Alzahrani et al. [10] applied the ARIMA model on the national data of Saudi Ministry of Health to predict the daily confirmed cases of COVID19 in Saudi Arabia. RMSE and MAE was considered as the performance evaluator of the prediction model. Author found ARIMA (2,1,1) as the best model for the prediction task.

Ranjan Gupta et al. [11] identified ARIMA (1,1,2) as the best model for the prediction of COVID19 cases considering the dataset obtained from John Hopkins University.

Sarbhan Singh et al. [12] found ARIMA (0,1,0) as the best one for the prediction of COVID19 cases in Malaysia. Author considered Bayes Information Criteria (BIC) along with MAPE for the model evaluation.

Santanu Roy et al. [13] did disease risk analysis carried out using weighted overlay analysis in GIS platform. The ARIMA (2,2,2) model was used for the prediction purpose. RMSE and MAE were considered as model evaluators. Andres Hernandez et al. [14] forecasted region wise COVID19 cases all over the world. Author utilized the ARIMA model and polynomial function for the prediction with RMSE as a performance measure. Table 1 summarizes the literature of the COVID19 prediction model.

Table I. Literature Review Table

Author and Year	**Model Utilized**	**Performance Metric**	**Dataset**
Alok Kumar Sahai *et al.*, 2020 [5]	ARIMA	MAD, MAPE	(IND, RUS, USA, SPN) https://www.worldometers.info/coronavirus/
Vasilis Z. Marmarelis *et al.*, 2020 [6]	RM	-	John Hopkins University. Github https://github.com/CSSEGISandData/COVID-19
Andi Sulasikin *et al.*, 2020 [7]	ARIMA	R^2, MSE, RMSE	DKI Jakarta Province www.corona.jakarta.go.id/id
Hadeel I. Mustafa *et al.*, 2020 [8]	ARIMA(2,1,5)	MSE, MAE	Iraqi Ministry of Health (Country wise data)
W.Regis Anne *et al.*, 2020 [9]	ARIMA(1,2,2)	RMSE, MAE	John Hopkins university
Saleh I. Alzahrani *et al.*,	ARIMA(2,1,1)	RMSE, MAE	Saudi Ministry of Health (National

2020 [10]			Data)
Rajan Gupta *et al*., 2020 [11]	ARIMA(1,1,2)	R_o	John Hopkins university
Sarbhan Singh *et al*., 2020 [12]	ARIMA(0,1,0)	MAPE, BIC	The official Ministry of Health(MOH), Malaysia and John Hopkins University.
Santanu Roy *et al*., 2020 [13]	ARIMA(2,2,2)	RMSE, MAE	(INDIA) https://www.covid19india.org/
Andres Hernandez et al. (2020) [14]	ARIMA	RMSE	World data from https://ourworldindata.org/coronavirus

III.Dataset and Methodology

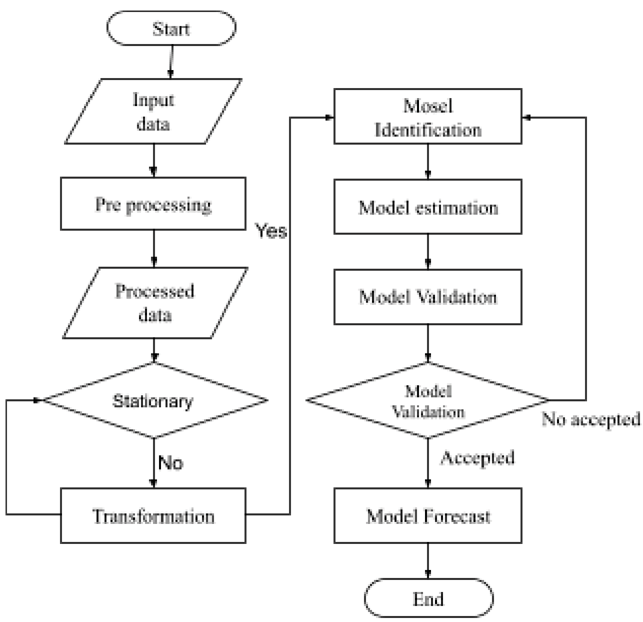

Fig. 1 Proposed Methodology for CoViD forecasting

A. DATA COLLECTION

The information gathered identifies with the true detailed instances of COVID19 in India from 30th January 2020 to 23rd December 2020. The dataset of COVID19 cases was gathered from the API (https://api.covid19india.org/documentation/csv/) [15] which assembles, extricates and distributes every day data from each of the 28 Indian State Health Offices about spread of CoViD-19 and list of festivals is obtained from the government website https://dopt.gov.in/sites/default/files/Holiday%20list%202020.pdf.[16] LibreOffice Calc is utilized to construct a period arrangement information base. The obtained dataset was then divided into two parts: train and test set. The training set comprises 90% of total data (from 30th January 2020 to 10th December 2020), and testing set consists of 10% data (from 11th December 2020 to 23rd December 2020).

B. Methods

1. Exponential smoothing

It is a strategy to constantly change a statistic, generally average, in the light of later experience. It relegates dramatically diminishing loads as the perceptions get more seasoned. As such, ongoing perceptions are given moderately more weight in determining than the more seasoned perceptions [17].

a) Single exponential smoothing

It is otherwise called simple exponential smoothing. It is utilized for small-range anticipating, typically only a few weeks or months into what's to come. The model accepts that the information varies around a sensibly steady mean (no pattern or reliable example of development). The particular equation for single exponential smoothing is:

$$S_t = \alpha * X_t + (1 - \alpha) * S_{t-1} \qquad (1)$$

It is applied recursively to all observations one by one in the time series to obtain a new smoothen value as the weighted mean of the current window and the past smoothen value. The past smoothen value was processed thus from the past noticed value and the smoothen an incentive before the past window, and so on.

So, effectively, each smoothen value can be found as the weighted mean of previous data points, such that the weights of previous windows diminish exponentially depending upon the value of a parameter denoted by α. If α is 1, it means the previous observations have to be neglected totally; else if it happens to be 0, then the present data point is neglected totally, and the smoothen value comprises entirely of the prior smoothen value (which in turn is calculated using the smoothen window prior to it, and so on; implying all smoothen values will be equal to the initial smoothen value S_0).In the middle of qualities will create halfway outcomes. Beginning Value [17] The underlying estimation of S_t assumes a significant job in processing all the resulting esteems. Setting it to y_1 is one strategy for instatement. Another chance is normally the initial four or five perceptions. The more modest the estimation of α, the more significant is the choice of the underlying estimation of S_t .

b) Double Exponential Smoothing

It is the technique that is utilized when there is some sort of trend in the information. Double smoothing with a trend works a lot like Simple smoothing aside from that here, instead of one, the two segments i.e. level and trend, should be refreshed every period. The level is defined as the smoothen value of the approximation of the information toward the finish of every period. The trend can be defined as the smoothened estimate of mean development at the end of each window. Double Exponential Smoothing is given by [17].

$$S_t = \alpha * y_t + (1 - \alpha) * (S_{t-1} + b_{t-1}) \quad 0 < \alpha < 1 \qquad (2)$$

$$b_t = \gamma * (S_t - S_{t-1}) + (1 - \gamma) * b_{t-1} \quad 0 < \gamma < 1 \qquad (3)$$

Initial Values

The initial values for S_t and b_t can be chosen using several methods. S1 is in general set to y1.

Three suggestions for b_1 are:

$b_1 = y_2 - y_1$

$b_1 = [(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)] / 3$

$b_1 = (y_n - y_1)/(n - 1)$

c) Triple Exponential Smoothing

This technique is utilized when the time series has trend and seasonality. Seasonality can be handled using a third boundary. We currently acquaint a third condition to deal with seasonality. The subsequent arrangement of conditions is known as the "Holt-Winters" (HW) strategy, which is named after its creators. There are two primary HW models, based upon the seasonality present in series.

- Multiplicative Seasonal Model
- Additive Seasonal Model

2. Autoregressive Integrated Moving Average (ARIMA) Model

It can be considered as the generalized version of Autoregressive Moving Average (ARMA) that is built by combining the Autoregressive (AR) process and Moving Average (MA) process and builds a compound model of the time series. As indicated by the acronym, ARIMA(p, d, q) has the following as key elements of the model:

- AR (Autoregression) : It works on the principle that variable of interest is regressed on its own lagged observations(p).
- I (Integrated) : To convert the time-series into stationary form by removing the trend. It is done by calculating the difference of data at different points (d).
- MA (Moving Average) : A method based on the relationship between the actual data and the error values when a model based on moving average is applied to the number of lagged values (q).

The equation of an Auto Regressive model having order equal to p, that is, AR(p), is expressed as a linear equation given by:

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t \qquad (4)$$

Here, x_t is the variable of concern that has been stationarized, c is denotes a constant, autocorrelation coefficients is given by Φ_i at lags 1, 2, 3,..., p and ε_t , the error residuals, are the GWN (Gaussian White Noise) series having variance as σ_ε^2 and mean as zero.

MA(q), i.e. an MA model with order as q, can be expressed as:

$$x_t = \mu + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \qquad (5)$$

Here, μ denotes $E(x_t)$, the expectation of x_t (generally taken as zero), θ_i is weight given to the present and previous values of the terms in the time-series, having θ_0 as 1.

An assumption is that ε_t is Gaussian white noise series having variance σ_ε^2 and average zero.

The ARIMA model having order as (p, q) can be formed by combining the above models by adding them as shown in equation (6).

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \qquad (6)$$

Here, $\sigma_\varepsilon^2 > 0$, $\theta_i \neq 0$, $\varphi_i \neq 0$. The parameters 'p' and 'q' are called the AR and MA orders, respectively.

ARIMA forecasting, also known by the name "Box and Jenkins forecasting", can deal with time-series which is non-stationary because it has an additional component called "integrate" that difference the time-series to change it into a stationary one. The ARIMA model is generally denoted by ARIMA(p, d, q). In this paper, auto ARIMA has been used to forecast COVID-19 outbreak. To check if the time-series is stationary or non-stationary, the "Augmented Dickey–Fuller" test was conducted. The time-series contains very less seasonal effects; thus, it can be considered to be nonseasonal and stationary data. In this paper, auto ARIMA is used to find the p, d and q values to define the order of model, and ARIMA (8,2,1) has been found as the best fit for the given time-series, hence this model is used to forecast. Akaike's Information Criterion (AIC) is widely used to find the parameters of an ARIMA model. It is given by [18]:

$$AIC\ (p) = n \ln (RS / n) + 2K \qquad (7)$$

Table II. Models Specifications

METHOD	PARAMETERS
Double Exponential Smoothing (Multiplicative)	$\alpha = 0.782, \beta = 0.0001$
Double Exponential Smoothing (Additive)	$\alpha = 0.782, \beta = 0.0001$
Triple Exponential Smoothing(trend = add, seasonality = add)	$\alpha = 0.606, \beta = 0.083, \gamma = 0.350$
Triple Exponential Smoothing(trend = mul, seasonality = add)	$\alpha = 0.606, \beta = 0.082, \gamma = 0.350$
Triple Exponential Smoothing(trend = add, seasonality = mul)	$\alpha = 0.606, \beta = 0.082, \gamma = 0.285$
Triple Exponential Smoothing(trend = mul, seasonality = mul)	$\alpha = 0.606, \beta = 0.110, \gamma = 0.285$
SARIMA	ar.L1 -1.003787e+00 ar.L2 -1.026829e+00 ar.L3 -9.745787e-01 ar.L4 -9.657847e-01 ar.L5 -9.381011e-01 ar.L6 -8.945952e-01 ar.S.L7 9.012076e-01 ma.S.L7 -1.368124e+0 ma.S.L14 3.642204e-01 ma.S.L21 2.580331e-01 ma.S.L28 -1.383605e-01 sigma2 5.424714e+06
SARIMAX	const -12.687284 Festival 93.276648 ar.L1 -1.693499 ar.L2 -1.894284 ar.L3 -1.851363 ar.L4 -1.812705 ar.L5 -1.807257 ar.L6 -1.722801 ar.L7 -0.841860 ar.L8 -0.190542 ma.L1 0.46
ARIMA	const -1.989183 ar.L1 -1.696156 ar.L2 -1.894129 ar.L3 -1.850823 ar.L4 -1.811462 ar.L5 -1.805883 ar.L6 -1.721284 ar.L7 -0.840485 ar.L8 -0.187829 ma.L1 0.472074

Here n represents the number of observations and RS gives the residual sums of squares. The model having minimum AIC value is chosen as the best predicting model.

3. Seasonal ARIMA (SARIMA)

SARIMA is an improvised version of ARIMA which is formed by incorporating seasonal factors in the ARIMA model. A SARIMA model is denoted as[19]:

$$\text{ARIMA } (p, d, q) \times (P, D, Q)S \quad (8)$$

Here p,d,q are used for non-seasonal computations and denote the AR order, differencing and MA order respectively, whereas P,D,Q represent the seasonal terms and individually they denote AR order, differencing and MA order respectively. The duration of the seasonal pattern is denoted by S. The seasonal and non-seasonal parts of the SARIMA model are similar to each other, but the former is associated with backshifts of the seasonal periods.

4. SARIMAX

ARIMA/SARIMA is applicable only on univariate time-series. The interdependence among different variables can't be studied using ARIMA. But, real life problems may contain relationships of some sort between different variables. The ever changing rules of the multivariate time series can't be expressed using the ARIMA model. The reason being the inefficiency of the measuring model. Therefore, a need to build a model that can work on a multivariate time-series was observed. So, a steady time-series model working on multiple variables, using the pre-existing ARIMA model, was created by Box and Jenkins. Technically, stability between input time-series and exogeneous time-series is a must. The limiting constraint prohibiting the development in the field of the analysis to multivariate time-series is needed to be satisfied as well. In this field, Granger and Engle proposed the concept of cointegration. This theory says that the residual of input time-series and output time series, after regression, are required to stabilize, and so, stabilizing the series itself is not required. With the help of the concept of cointegration, the development of multivariate time-series analysis was made possible and introduction of multivariate time-series analysis led to the improvement of precision of forecasting[20].

C. Architecture

The model having the below mentioned structure was defined as the "Dynamic Regression Model", which was further simplified and called the SARIMAX model.

$$y_t = \mu + \sum_{i=1}^{k} \frac{\Theta_i(B)}{\phi_i(B)} B^{l_i} x^{it} + \varepsilon_t \quad (9)$$

$$\varepsilon_t = \frac{\Theta(B)}{\phi(B)} a_t \quad (10)$$

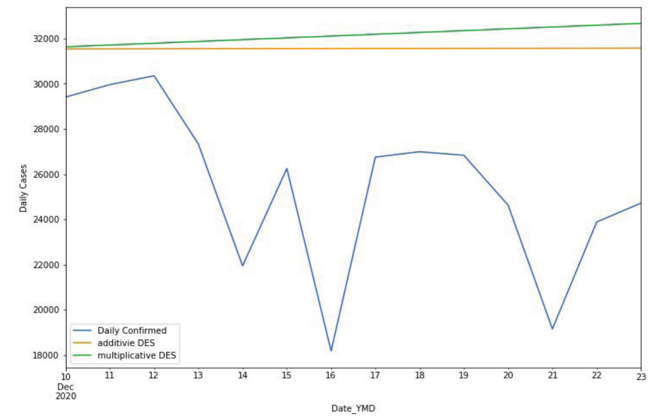

Fig. 2 : Prediction using Double Exponential Smoothing

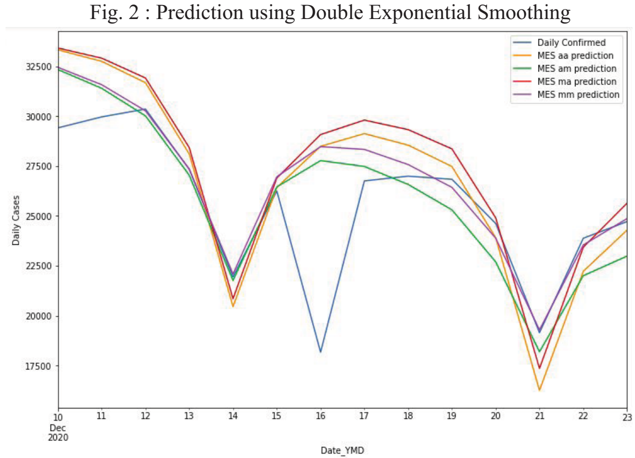

Fig. 3 : Comparison between Prediction using Triple Exponential Smoothing with (a) trend = add, seasonality = add; (b) trend = add, seasonality = mul (c) trend = mul, seasonality = add (d) trend = mul, seasonality = mul

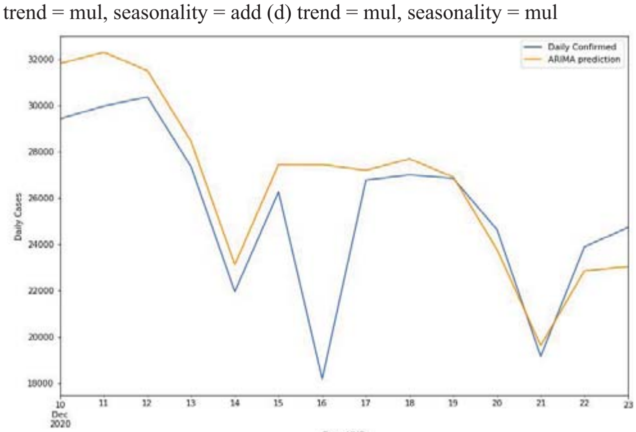

Fig. 4 : Prediction using ARIMA

Here, $\Phi_i(B)$ represents the multinomial of auto-regression coefficients of the *I-th* input, $\Theta_i(B)$ represents the multinomial of average coefficients of the i-th input, l_i denotes the degree of the lag of the i-th input variable, $\{\varepsilon_t\}$ is the residual of regression of time-series, $\Phi(B)$ represents the coefficients of auto-regression of residual series multinomial, and the white noise time-series with zero mean is denoted by $\{a_t\}$.[20]

D. Working

The basic concept of the SARIMAX model is as follows:; Firstly, the regression model for output time-series and input time-series is created, assuming that both $\{y^t\}$ i.e. output time series, and $\{x^{it}\}$ i.e. input time series ,where (i = 1, 2,…, k), are stationary:

$$y_t = \mu + \sum_{i=1}^{l} \frac{\Theta_i(B)}{\varphi_i(B)} B^{l_i} x^{it} + \varepsilon_t \qquad (11)$$

Now, as known, the linear combination of a stationary time-series is also stationary, $\{\varepsilon_t\}$ i.e. residual time-series, is also stationary because, as mentioned above, input and output time-series, both, are stationary.

IV. Experimental Results

The comparison of original data from 11th December 2020 to 23rd December 2020 was done and the forecast efficiency was checked using rooted mean square error (RMSE) and Mean Absolute Error(MSE).

The RMSE for Double Exponential Smoothing(Holt method) with parameters $\alpha = 0.78, \beta = 0.0001$ is found to be 7677.64 for additive trend and 7077.60 for multiplicative trend as shown in Table II and Table III and the comparison between predicted and actual value is shown in Fig. 2

The result of Triple Exponential Smoothing was minimum for multiplicative trend and multiplicative seasoning with parameters $\alpha = 0.751, \beta = 0.0001$ as depicted in Table I and visualized in Fig. 3.

For using ARIMA, the initial step was to test for unit root in the given time-series. A visual representation of the data by plotting the graph was observed and it was observed that the series was rising exponentially and was non-stationary. ADF test was done on time-series and it was found that the series has a non-stationary trend. After 2^{nd} order differencing, the Dicky Fuller test was again conducted and it was found that the series is now stationary which infers that the series is non-seasonal. The specifications of the model which were obtained by using the "Hannan Rissanen algorithm" [14] was (p,d,q) = (8,2,1). Fig. 4 visualizes the result of ARIMA.

To understand the effect of other factors (exogenous variables) i.e. the effects of festivals on the number of cases in India, SARIMAX model is used. The result can be visualized in Fig. *5* The ARIMA(8,2,1) model was then used to forecast daily confirmed cases of test set for 14 days up to

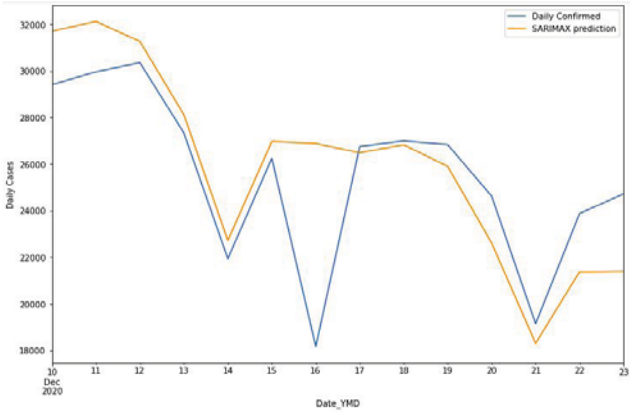

Fig. 5 : Prediction using SARIMAX

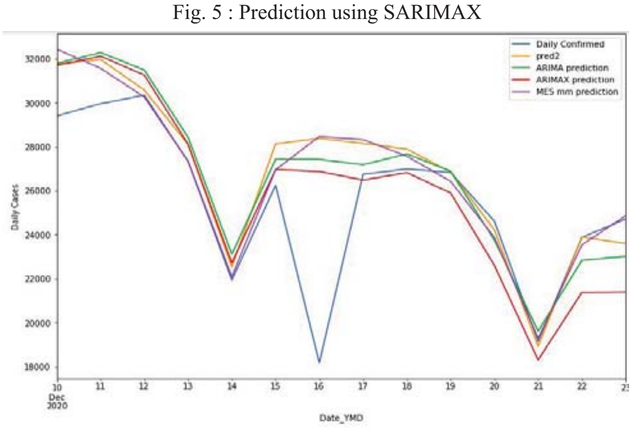

Fig. 6: Comparison between (a) Triple Exponential Smoothing; (b) ARIMA (c) SARIMA (d) SARIMAX

Table III. Comparison of all methods

METHOD	**RMSE**	**MAE**
Double Exponential Smoothing (Multiplicative)	7677.64	6696.91
Double Exponential Smoothing (Additive)	7077.60	6104.24
Triple Exponential Smoothing(trend = add, seasonality = add)	3441.34	2215.03
Triple Exponential Smoothing(trend = mul, seasonality = add)	3401.80	2326.12
Triple Exponential Smoothing(trend = add, seasonality = mul)	3280.59	1723.58
Triple Exponential Smoothing(trend = mul, seasonality = mul)	3221.96	**1416.48**
SARIMA	2961.48	1581.58
SARIMAX	2819.80	1884.81
ARIMA	**2773.27**	1700.78

23rd December 2020. The RMSE was lowest for the ARIMA model, followed by the SARIMA.

E. Evaluation of Performance

To judge the performance of the proposed models used in the paper, the difference obtained between the obtained and forecasted result is compared.The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are chosen as a measurement for doing comparison. [13]

• RMSE- It is the square root of MSE, and is given by:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}[P_t - Z_t]^2} \qquad (12)$$

P_t = Value predicted at time t

Z_t = Value observed at time t

• MAE- It is given by:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|P_t - Z_t| \qquad (13)$$

P_t = Value predicted at time t

Z_t = Value observed at time t

The graphical model visualizes the data driven gist to determine the most important characteristics of the images in various stages. Fig. 6 shows the graphical analysis with respect to Mean Square Error. It compares the input image and output image for each period of the proposed plot]. MSE also portrays the slip-up size against each pixel, which is restricted profitably during de-noising and restoration cycle fairly.

V. Conclusion

In this paper, an experimental study was conducted for forecasting of COVID-19 pandemic spread pattern. Apart from that, a comparison of the difference between forecasted and original value was done. It was found that the ARIMA model with parameters (p,d,q) = (8,2,1) worked best among all with RMSE of **2773.27**. While in the case of MAE, triple exponential smoothing with parameters (α,β,γ) = (0.606, 0.110, 0.285) performed best with a value of **1416.48**. If considered RMSE value of 2773.27 to be in an acceptable range, then ARIMA model for forecasting can be used as it acquires AR for considering past value and MA for current and preceding residual series knowledge. The linearity of COVID patterns can be easily captured using an efficient linear model and ARIMA model fulfils that purpose. The limitation of this model is that it can be handy in calculating linear relationships which are not affected by multiple factors but there are multiple factors that can affect the spreading of COVID. like climatic conditions and social influences like social distancing, wearing masks, using sanitizer and many more. To overcome this limitation, the ARIMAX model is used, with the help of which, multivariate time-series analysis can also be performed.

References

[1] 10101998 Coronavirus Update (Live). Cases and 501644 deaths from COVID19 virus pandemic - worldometer. Available at: https://www.worldometers. info/coronavirus/. [Accessed 28 June 2020].

[2] Murugesan B, Karuppannan S, Mengistie AT, Ranganathan M, Gopalakrishnan G (2020) Distribution and trend analysis of COVID-19 in India: geospatial approach.

[3] Wang Li Y, Peng B, Zhou R, Zhan CY, Liu Z. Mathematical modeling and epidemic prediction of COVID-19 and its significance to epidemic prevention and control measures. Ann Infect Dis Epidemiol 2020;5(1):1052. 2020.

[4] Boulos MNK (2004) Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom. Int J Health Geogr. https://doi.org/10.1186/1476-072X-3-1 Health 2016;8:1845e7. https://doi.org/10.4236/health.2016.815176.

[5] Alok Kumar Sahai , Namita Rath , Vishal Sood , Manvendra Pratap Singh. ARIMA modelling & forecasting of COVID-19 in top five affected countries.

[6] V. Z. Marmarelis, "Predictive Modeling of Covid-19 Data in the US: Adaptive Phase-Space Approach," in IEEE Open Journal of Engineering in Medicine and Biology, vol. 1, pp. 207-213, 2020, doi: 10.1109/OJEMB.2020.3008313.

[7] A. Sulasikin, Y. Nugraha, J. Kanggrawan and A. L. Suherman, "Forecasting for a data-driven policy using time series methods in handling COVID-19 pandemic in Jakarta," 2020 IEEE International Smart Cities Conference (ISC2), Piscataway, NJ, USA, 2020, pp. 1-6, doi: 10.1109/ISC251055.2020.9239066.

[8] H. I. Mustafa and N. Y. Fareed, "COVID-19 Cases in Iraq; Forecasting Incidents Using Box - Jenkins ARIMA Model," 2020 2nd Al-Noor International Conference for Science and Technology (NICST), Baku, 2020, pp. 22-26, doi: 10.1109/NICST50904.2020.9280304.

[9]Anne W, Regis. (2020). ARIMA modelling of predicting COVID-19 infections. 10.1101/2020.04.18.20070631.

[10] Saleh I. Alzahrani , Ibrahim A. Aljamaan, Ebrahim A. Al-Fakih. Forecasting the spread of the COVID-19 pandemic in Saudi Arabia
using ARIMA prediction model under current public health interventions

[11]Rajan Gupta and Saibal K Pal. Trend Analysis and Forecasting of COVID-19 outbreak in India.

[12] Sarbhan Singh, Bala Murali Sundram, Kamesh Rajendran, Kian Boon Law. Forecasting daily confirmed COVID-19 cases in Malaysia using ARIMA models.

[13] Santanu Roy, Gouri Sankar Bhunia, Pravat Kumar Shit. Spatial prediction of COVID-19 epidemic using ARIMA techniques in India

[14] Andres Hernandez-Matamoros, Hamido Fujita, Toshitaka Hayashi. Forecasting of COVID19 per regions using ARIMA models and
polynomial functions

[15]Website for collecting Indian data https://api.covid19india.org

[16] Website for obtaining list of holidays in India https://dopt.gov.in/sites/default/files/Holiday%20list%202020.pdf

[17]Kalekar, Prajakta S.. "Time series Forecasting using Holt-Winters Exponential Smoothing." (2004).

[18]S. Siami-Namini, N. Tavakoli and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, 2018, pp. 1394-1401, doi: 10.1109/ICMLA.2018.00227.

[19]Peng Chen *et al* 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* 394 052024

[20]Jalalkamali, A., Moradi, M. & Moradi, N. Application of several artificial intelligence models and ARIMAX models for forecasting drought using the Standardized Precipitation Index. *Int. J. Environ. Sci. Technol.* 12, 1201–1210 (2015).