



Research Article

Copy Right@ T Raja Rani

# Optimal Machine Learning Methods to Forecast COVID-19 Cases

Dr. T. Raja Rani<sup>\*1</sup>, Dr. T.S.L Radhika<sup>2</sup> and Shalini Pukkella<sup>3</sup>

<sup>1</sup>Research Fellow, School of Engineering, University of Portsmouth, UK

<sup>2</sup>Department of Mathematics, BITS Pilani, Hyderabad Campus, Hyderabad, India

<sup>3</sup>Mechanical & Industrial Department, NUST (affiliated to Glasgow Caledonian University), Muscat, Oman

\*Corresponding author: T Raja Rani, General Studies Department, Military Technological College, Muscat, Oman.

**To Cite This Article:** T Raja Rani, TSL Radhika, Shalini Pukkella, Optimal Machine Learning Methods to Forecast COVID-19 Cases. 2020 - 11(1). AJBSR.MS.ID.001583. DOI: [10.34297/AJBSR.2020.11.001583](https://doi.org/10.34297/AJBSR.2020.11.001583).

**Received:** 📅 November 17, 2020; **Published:** 📅 November 25, 2020

## Abstract

The entire world is undergoing a hard-hitting scenario and trying to combat the COVID-19 by recent technological advancements, which involves machine learning chiefly. The forecasting demand has become a prerequisite as it helps the government officials and other organizations to make well-versed verdicts and impose pertinent measures to benefit the living conditions of the individuals around the globe. Consequently, the current paper focuses on four significant machine learning methods (Facebook Prophet, Auto Regression, Vector Auto Regression, and Holt-Winters), which help forecast the total confirmed and daily confirmed cases. Moreover, the paper reveals the ideal method for the future forecast based on the attained results and efficacy rate. The results of the study reveal the best methods for the considered countries based on the calculation of Error percentage. Out of the four Machine Learning models, AR and FB models stood out as the best methods when compared to the other two.

**Keywords:** Total Confirmed Cases, Daily Confirmed, Total Deceased, Total Recovery, Akaike Information Criterion, Bayesian Information Criterion

## Introduction

In the current striking scenario, forecasting of COVID-19 cases has become quite essential. It helps to create awareness for both government and the public due to which scenario-planning tools take place in every other sector. Moreover, the public will take the utmost care to prevent themselves from nerve-racking disease. The forecasting techniques aid hospitals and organizations' management to share the importance of social distancing, meet essential etiquettes, and better hospitalization and food facilities. It is important to use the accurate forecasting technique as the wrong adopted technique devastates billions' lives and upshot the well-being, economy, and societal chaos on an at-greater scale.

Machine Learning is one of the Artificial Intelligence branches and a strategy for information examination that robotizes the logical model structure. The framework gains from information, recognize examples and settle on choices with less human interruption. Time Series Forecasting is a prominent area of machine learning

and is used for many crucial predictions where time plays a major role. It is also known as extrapolation in the classical statistical handling of Time Series data. The current paper focuses on four main machine learning methods. Facebook Prophet, Auto Regression, Holt-Winters, and Vector Auto Regression are the time series forecasting methods, part of Machine Learning.

### Facebook Prophet (FB)

The prophet method is an additive time series forecasting model where non-linear trends support weekly, yearly, daily seasonality, and holiday effects. The current method uses two leading programming languages called Python and R. However, the current paper got progressed by the Python programming language, where the method primarily builds the model. After this, predicts the actual, lower and upper limits of total or daily confirmed cases. Generally, the method is swift and irregular holidays are well-taken care of [1].

Moreover, an unknown forecast extension of the target limit automatically sets up the cap by logarithmic model instead of linear. However, the pre-defined data format gets preferred and requires tuning specific parameters manually in case of high irregular data. Further, at the time of modeling, no facility is provided to incorporate exogenous variables. The interval width of FB in our present work is 99%. Facebook developed it [2].

### Auto Regression (AR)

The Auto Regression method is a linear regression general forecasting model and is quite flexible as it showcases various time-series patterns. One can easily visualize the stimulated AR progressions. It uses a linear combination of the given values for future forecasts. A proper lag estimation can be done using the AR method by using the Mean Square Error. As the value of lag increases, the rate of error decreases, and it depends on the value of Mean Square Error. Lower the Mean Square Error, accurate will be the predictions [3].

### Holt Winters (HW)

Holt-Winters is one of the populous time series forecasting techniques. In 1957, the economist Holt introduced the "Double Exponential Smoothing" method, and the work got extended further in the year 1960 by Winter to capture seasonality. Thus, the seasonal additive method is popularly called Holt-Winters. The current paper involves double and triple exponential smoothing to attain the best forecasting outcomes. It helps in determining the central value to overcome complex seasonal patterns. Lower the Mean Square Error, accurate will be the forecasting. It is a powerful forecasting method, and its trend is of additive type.

### Vector Auto Regression (VAR)

Vector Auto Regression method is one of the general forecasting models. The augmented Dickey-Fuller test (ADF test) will check whether the given data is stationary or non-stationary. Further, the optimal value of order P (lagged value) is attained. Based on the value of the order, the forecast of the data takes place. If the p-value exceeds 0.05, at a reasonable level of implication, the series turns out to be non-stationary [4].

### Literature Review

Loannidis J et al. discussed the main reasons for the failure of forecasting and suggested some ways to fix a few problems for better forecasting results. Incorrect assumptions related to modeling and epidemiological features stood as fundamental causes for the poor forecasting results apart from errors, sensitivity factors, and absence of proficiency. They suggest gathering appropriate evidence to analyze and provide appropriate solutions, leading to better forecasting outcomes. Consideration of modeling using predictive distributions along with multiple dimensions further helps

besides the continual reappraisal of models. One needs to calibrate analytical intuitions and verdicts when considering extreme values as inputs for future forecasts. Further, the author recommended a holistic approach to avoid the irregularity of risks [5].

Ünlü R & Namlı E applied various machine learning techniques such as Support Vector Machines, Prophet, Long-Short Term Memory, and Holt-Winters to study the confirmed and mortality cases for the top seven countries. The clustering of seven countries done by elbow and silhouette scores methods. The prophet model resulted in the lowest Root Mean Square value amongst all the models. Further, the paper estimates 150 days of confirmed and death cases in Turkey using the Prophet method [6].

Ardabili et al. suggests an alternative towards SIR and SEIR models. They are soft computing and machine learning models, including MLP and ANFIS, which helped generate promising outcomes. The paper suggests machine learning as an optimal tool to study the complex nature of the COVID-19 outbreak and confirms the latent of machine learning for future research. The paper discusses essential modelling tools such as Evolutionary algorithms that solve optimization problems through intelligent methods and the Genetic Algorithm, which solves specific data structure problems. Apart from RMSE, the correlation coefficient was found for using PSO and GA for five selected countries. GWO derived functions stood as the best predictor compared to others [7].

## Preparation and Methodology

### Preparation

Preparation is one of the essential steps for any research. Before the discussion of the Methodology of various Forecasting techniques, planning includes the following necessary steps:

- Collection of COVID-19 data from various online resources [8] and to conduct well-research on the previous literature review to understand the necessary resources and software required.
- To Perform the Forecasting study on Google Colab Platform (open source) using Python programming language where Stats Library included to ensure analysis of total confirmed and daily confirmed cases along with the efficacy.
- Draw the obtained results, which include graphs, and further analyze the data to compute the critical parameters such as to meet the primary purpose of the research.

### Methodology

The methodology is designed through problem definition and by meeting the essential requirements. The research successfully tracked different methodologies for four diverse types of Machine Learning methods, as discussed below.

**Methodology of Facebook Prophet:** Facebook Prophet is one of the easiest forecasting methods which generates high-quality forecasts. Firstly, imported necessary libraries and data set and ensured the data frame is segregated into two columns known as Date and Data. The date is denoted by 'ds,' and Data is denoted by 'y' respectively. There are 295 rows and two columns in the data frame. Now, the model is built with an interval width of about 99% confidence level. The built model is further trained and later fits the model to the best possible future data frame. Lastly, from the attained future data frame, forecasted the total confirmed and daily confirmed cases for India, the US, and Oman, respectively. The model forecasts the total number of cases in three different formats, typically called yhat, yhat\_lower, and yhat\_upper. Here, yhat\_lower and yhat\_upper denotes the lower limit and upper limit of the forecasted value of yhat.

**Methodology of Holt Winters:** Firstly, imported necessary libraries and data set. When computed, the data's length is 295, where 80% of the data is considered for training and the rest, 20% for testing. Next, considered the test dataset to fit the model best using Exponential Weighted Moving Average, Stats model, and Double exponential Stats model. There found a huge difference between EWMA and SES with tested values. However, DES gives the best approximation with the actual value. Further, the model was fit exponential smoothing, and required predictions were generated for the test value. Lastly, mean square error and mean absolute error got computed to best fit using exponential smoothing, and future forecasts of the total confirmed and daily confirmed cases were determined.

**Methodology of Auto Regression:** Auto Regression is a linear combination of seasonal predictors. It runs against a set of lagged values of order p. Firstly, imported necessary libraries and data set. When computed, the length of the data is 295, where 80% of the data is considered for training and the rest, 20% for testing. Further, appropriate test predictions got calculated with minimal Mean Square Error (MSE) of AR. Finally, the total confirmed and daily confirmed cases were forecasted.

**Methodology of Vector Auto Regression:** Firstly, two different data sets imported and joined. Later, the necessary libraries were imported. Applying the Augmented Dickey-Fuller test (ADF test) on

the given data checked whether the data is stationary or non-stationary. If the data set was found non-stationary, it was changed to stationary with the help of the syntax transformed.diff(). Moreover, the grid search for the order p was done using AR of VAR model considering the trained data, which obtained the least AIC value, which gets confirmed as the best value of order p. Thus, the VAR model's close-fitting is achieved for the order p, also known as the VAR model's lagged value. Lastly, with the attained value of lag, the future forecasted value gets calculated.

## Results and Analysis

The current section focuses on the forecasting of the COVID-19 cases using classical machine learning models for the three countries (i.e., India, US and Oman). Over the past few months, much insightful work was done on COVID-19, and the optimal data got congregated for the present study from the World Health Organization (WHO) portal. The data got collected starting from 22nd January 2020 to 11th November 2020 to drive the future forecasts forward [8].

### Study of COVID Cases for Three Countries

As of 16 November 2020, the total number of confirmed cases throughout the world reached 53,766,728 million. Table 1 in the Analysis Section presents the total confirmed, recovered and death statistics. However, we confined our attention to forecast Total confirmed and Daily cases (positive cases).

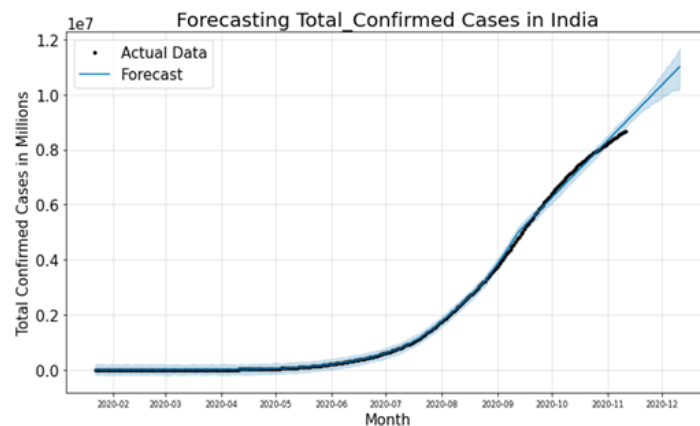
### Classical Machine Learning Methods

The study bounds to forecast the total and daily confirmed cases using important ML methods such as FB, AR, VAR and Holt-Winters. Further, a few important parameters got considered for the declaration of the powerful ML method.

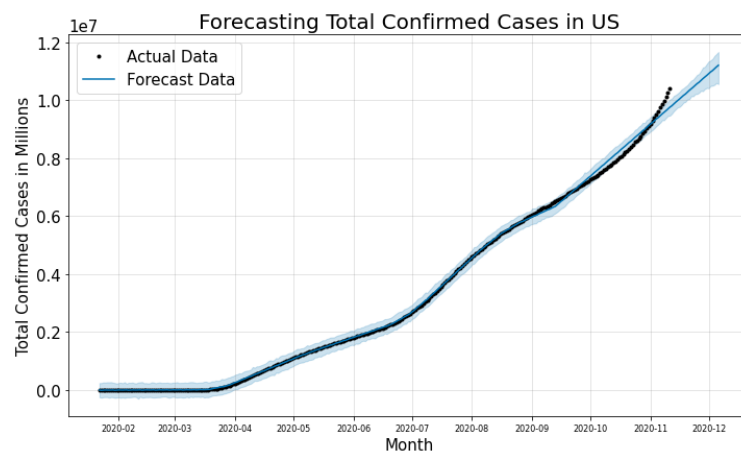
**Results of Facebook Prophet:** Forecasting was done using the FB model on 30 November 2020 as viewed from. According to the total and daily confirmed cases, it is confirmed that the US is hitting its peak when compared to India and Oman. Oman has the least number of confirmed cases when compared to other countries. It implies that all precautionary measures need to US. depict the total confirmed cases in India, US and Oman, respectively.

**Table 2:** Time Series Analysis with Facebook Prophet Library.

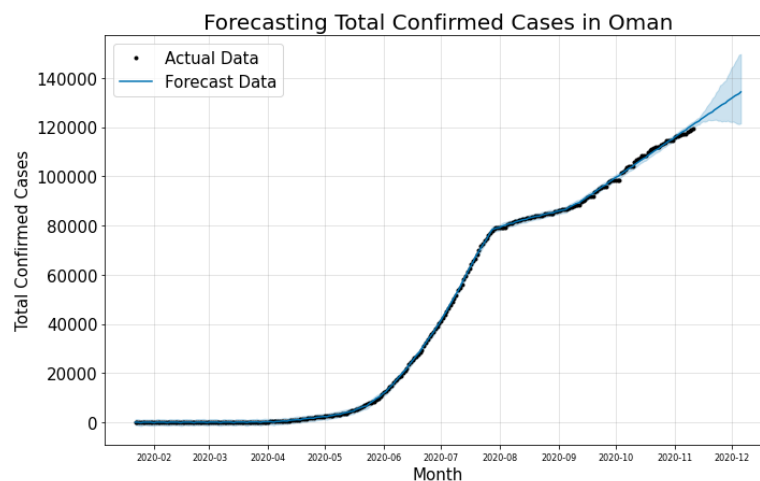
FB Model Forecasting on 30 November 2020				
Country	Forecasting Cases	yhat	yhat lower	yhat upper
India	Total Confirmed	10.268.220	9.871.472	10.678,310
	Dally Continued	17465	6.616	27.559
US	Total Confirmed	10862360	10462530	11258740
	Daily Confirmed	93.542	61.31	126.286
Oman	Total Confirmed	131155	123.685	140417
	Daily Confirmed	335	-664	1285



**Figure 1:** Forecasting total confirmed cases in India.



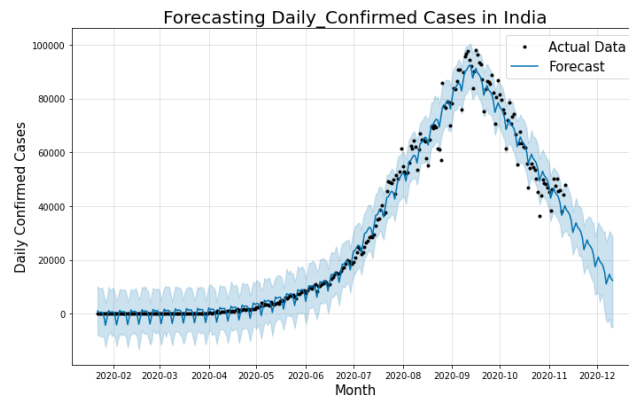
**Figure 2:** Forecasting total confirmed cases in US.



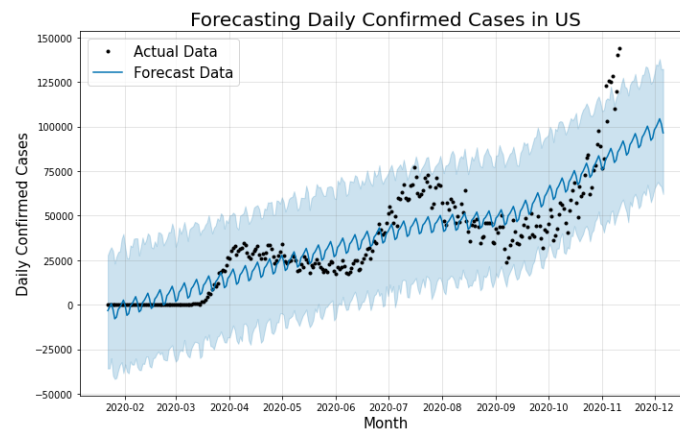
**Figure 3:** Forecasting total confirmed cases in Oman.

From the below figures, it is evident that data of daily confirmed cases in Oman got scattered broadly, and it enflamed the difficulty level of prediction to attain future forecasting results. Moreover, the figures depict the daily confirmed cases in considered countries.

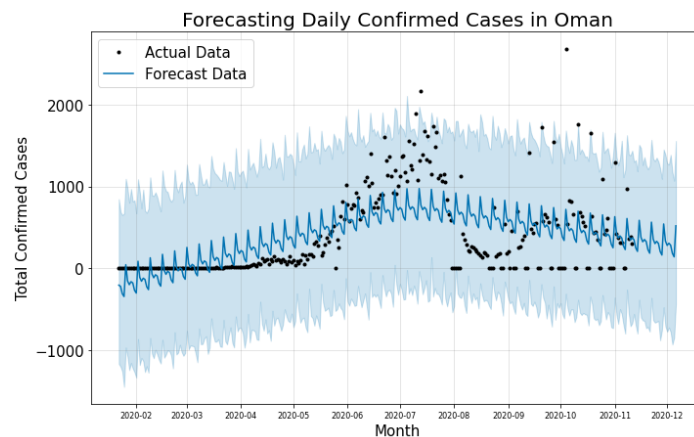
The number of confirmed cases in India and Oman seems to decline by 30 November 2020. Figure 5 gives a clear sign of the start of the second wave in the US because of which several cases are expected to reach its peak by 30 November 2020.



**Figure 4:** Forecasting Daily confirmed cases in India.



**Figure 5:** Forecasting Daily confirmed cases in US.



**Figure 6:** Forecasting Daily confirmed cases in Oman.

**Table 3:** Time Series Analysis with Stats Library.

Holt-Winters Model Forecasting on 30 November 2020		
Country	Forecasting Cases	Number of cases
India	Total Confirmed	9,565,380
	Daily Confirmed	44,489
US	Total Confirmed	131,18,651
	Daily Confirmed	1,52,185
Oman	Total Confirmed	1,26,233
	Daily Confirmed	400

**Results of Holt Winters:** Forecasting was done using the Holt-Winters model on 30 November 2020 as viewed from Table 3. According to the numbers of total and daily confirmed cases, it is deep-rooted that the US is hitting its peak when compared to India and Oman. Oman has the least number of confirmed cases when

compared to other countries. The residents of US need to abide by the informed precautions from the government of the US. Figures 1-3 depict the total confirmed cases in India, US and Oman, respectively.

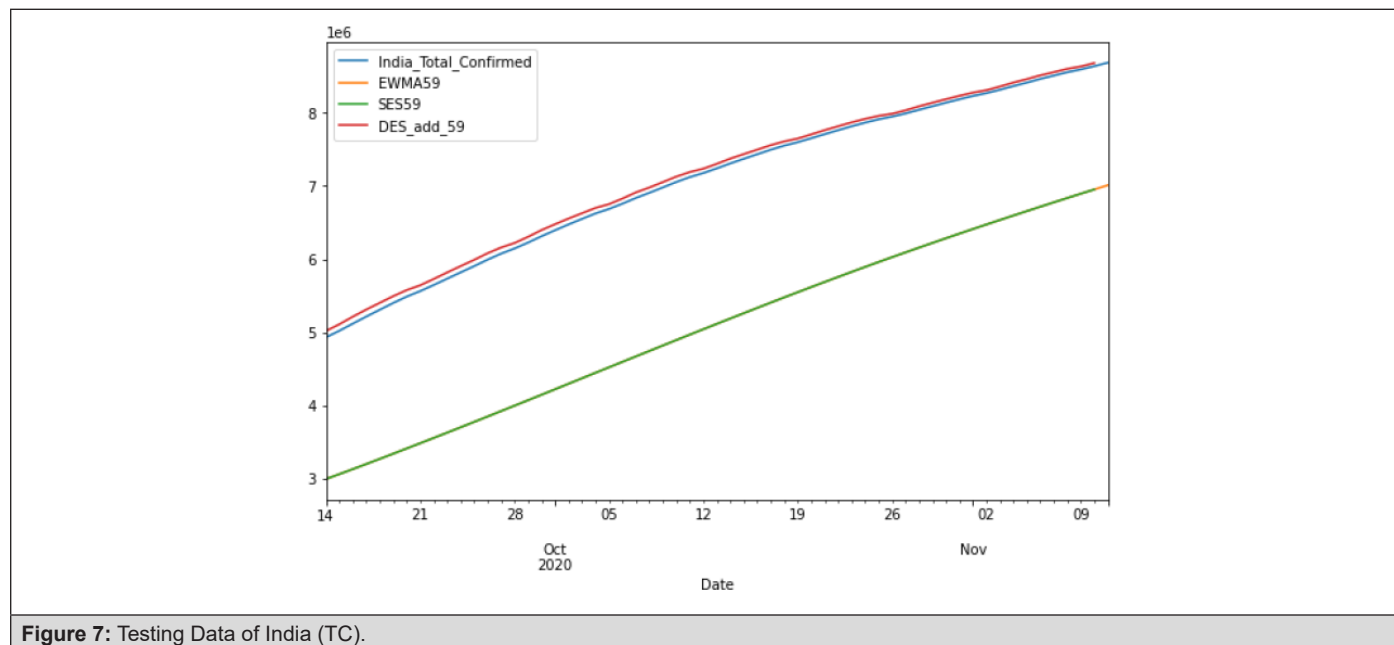
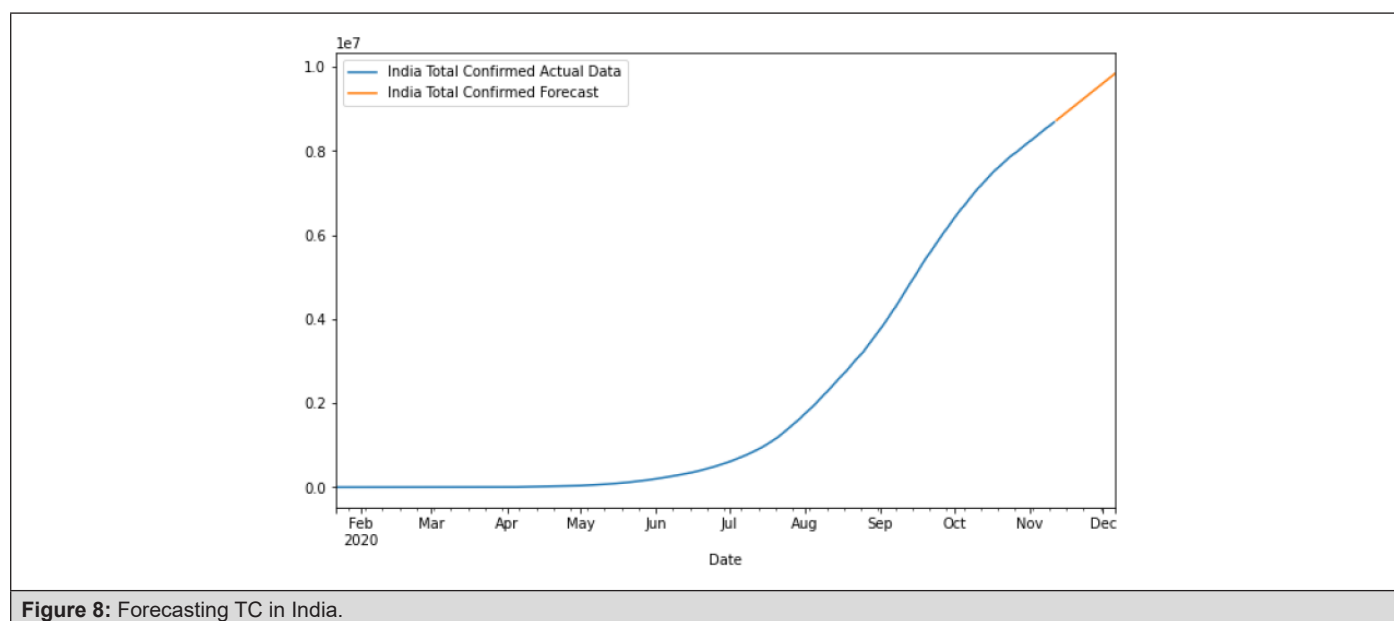
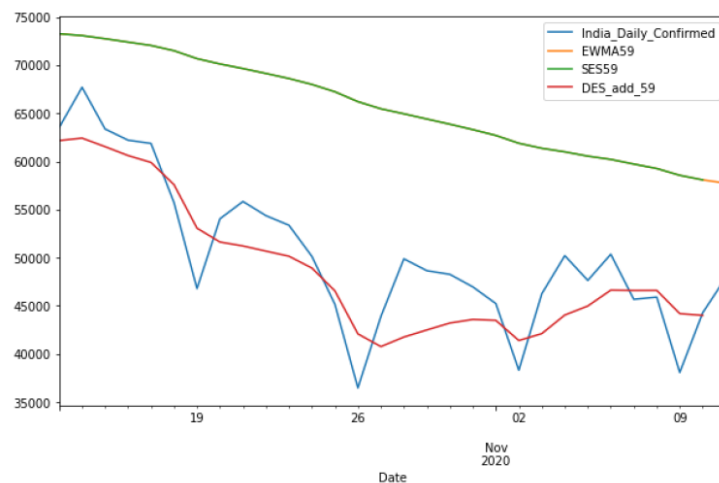


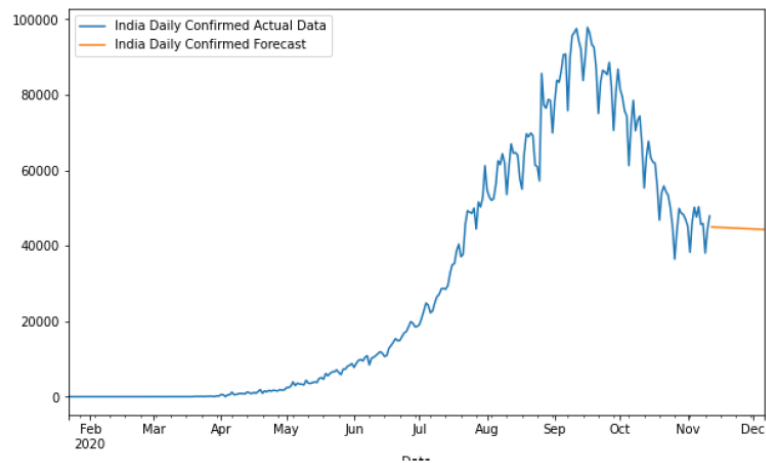
Figure 7 depicts 20% of the testing data of India from 14 September to 11 November 2020. It is visible that EWMA (Exponential Weighted Moving Average) and SES (Single Exponential Smoothing) are in a linear trend and overlapping each other. However, found a massive error concerning the total confirmed testing data. Consequently, Double Exponential Smoothing method (DES)

applied because of which, DES and total confirmed trend lines got very near to each other. The forecasting model is simply in the form of a straight sloped line extending from the recent data point. As Holt-Winters method is a "Triple Exponential" method, introduced smoothing factor and forecasted the total confirmed cases of India from 12-30 November 2020 as seen in the figure below.





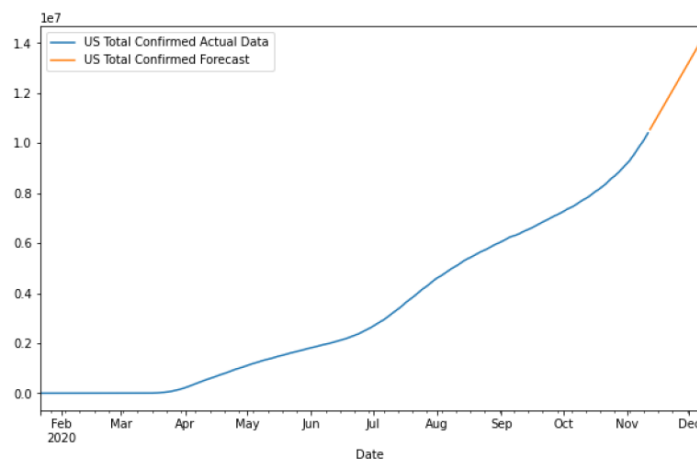
**Figure 9:** Testing Data of India (DC).



**Figure 10:** Forecasting DC in India.

Figures 9 & 10 depict the testing data and forecasting data of daily confirmed cases which got procured in the same manner as that of the total confirmed data. The same applies to forecasting

TC and DC cases of US and Oman. The Figures related to the US and Oman can be seen below.



**Figure 11:** Forecasting TC in US.



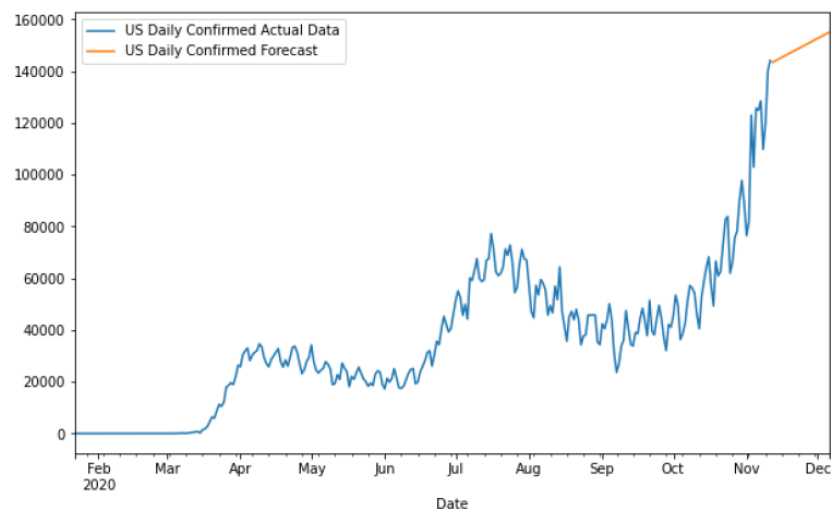


Figure 12: Forecasting DC in US.

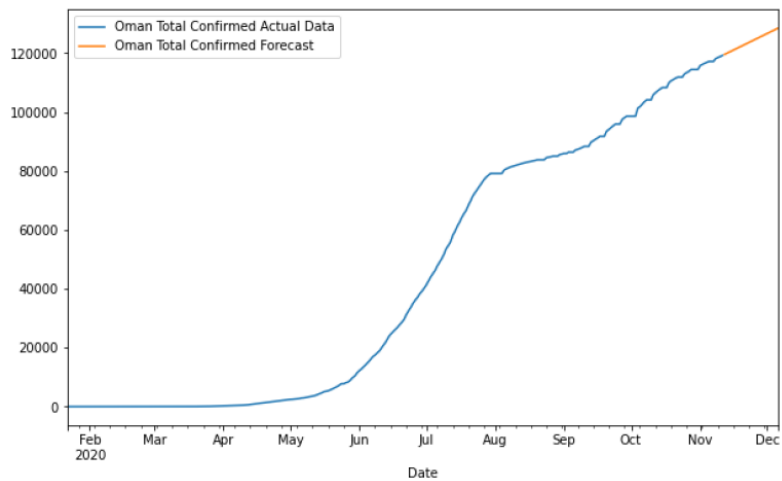


Figure 13: Forecasting TC in Oman.

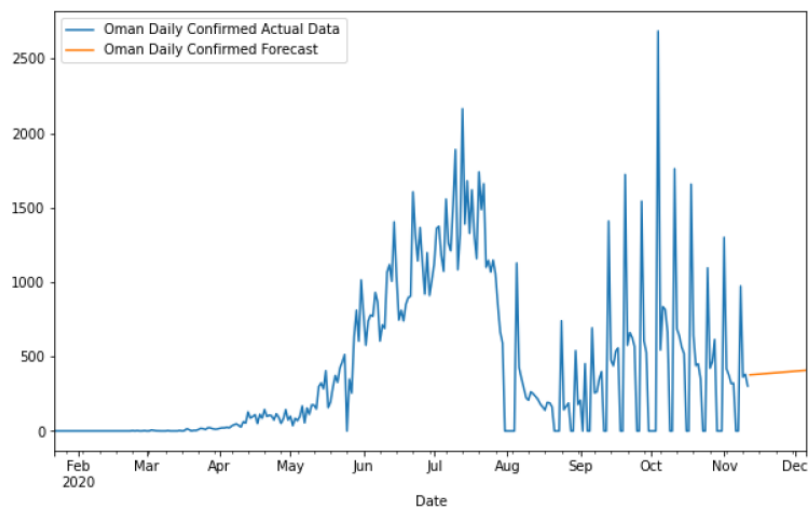


Figure 14: Forecasting DC in Oman.

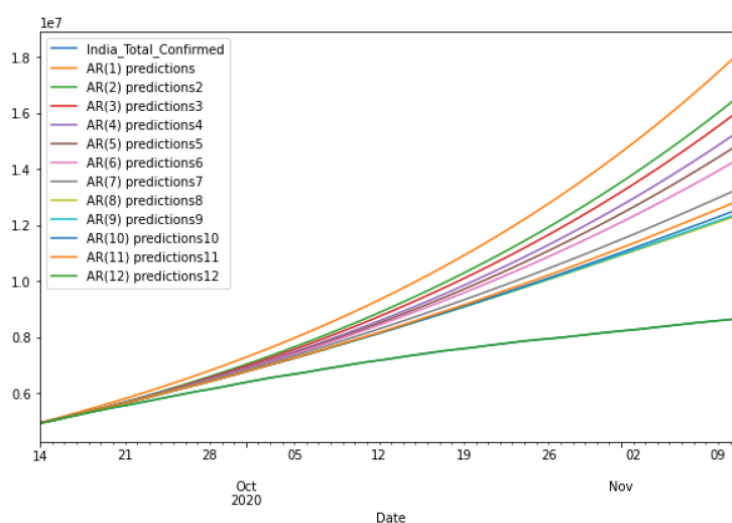


**Results of Auto Regression:** From Table 4, forecasting was done using the Holt-Winters model on 30 November 2020. According to the numbers of total and daily confirmed cases, it is

deep-rooted that the US is hitting its peak when compared to India and Oman. Oman has the least number of confirmed cases when compared to other countries.

**Table 4:** General Forecasting Model (AR).

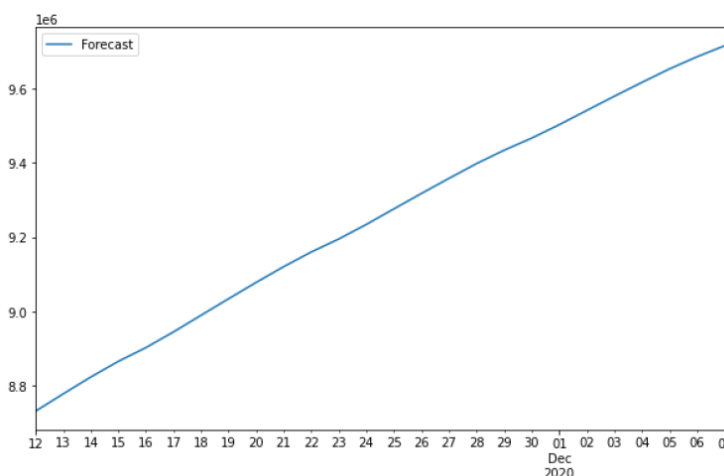
AR Model Forecasting on 30 November 2020		
Country	Forecasting Cases	Number of cases
India	Total Confirmed	94,67,084
	Daily Confirmed	36,950
US	Total Confirmed	134,98,400
	Daily Confirmed	1,95,656
Oman	Total Confirmed	1,24,224
	Daily Confirmed	354



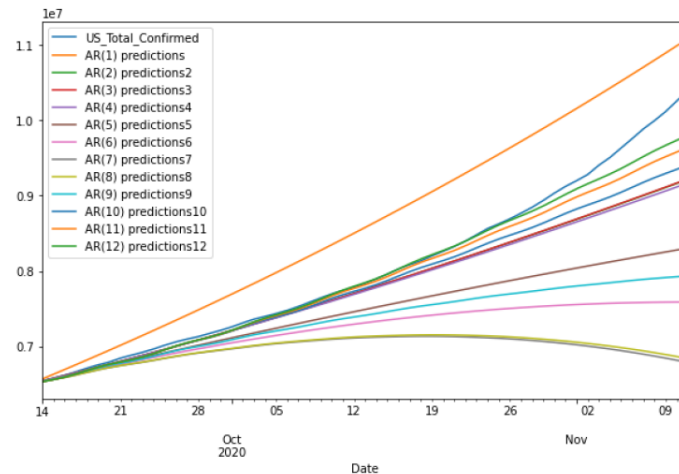
**Figure 15:** Testing TC in India.

Figures 15-16 depict the testing data and forecasting data of total confirmed cases using Auto Regression for India. Initially, the test data set for India got fit as AR1 model with a maximum lag of 1 and because of which obtained high Mean Square Error value (i.e., 684618). Finally, the model got best-fitting as AR12, which rep-

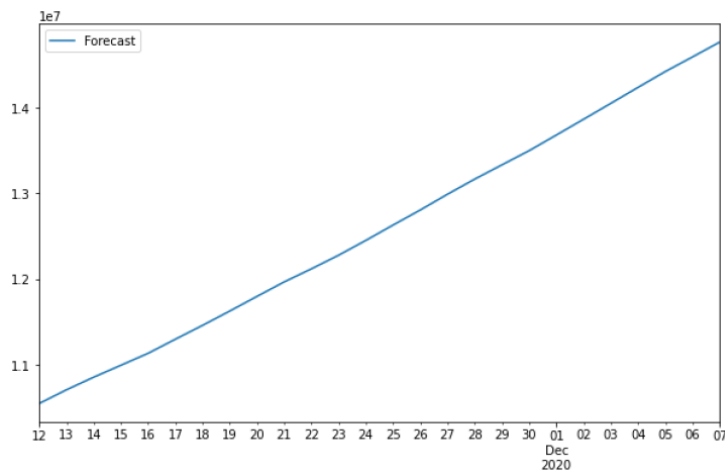
resents the total number of lags (i.e., lags = 12). At this point, the Mean Square Error is minimal as compared to previous and following lags. Lower the Mean Square Error value, appropriate predictions through AR gets generated.



**Figure 16:** Forecasting TC in India.



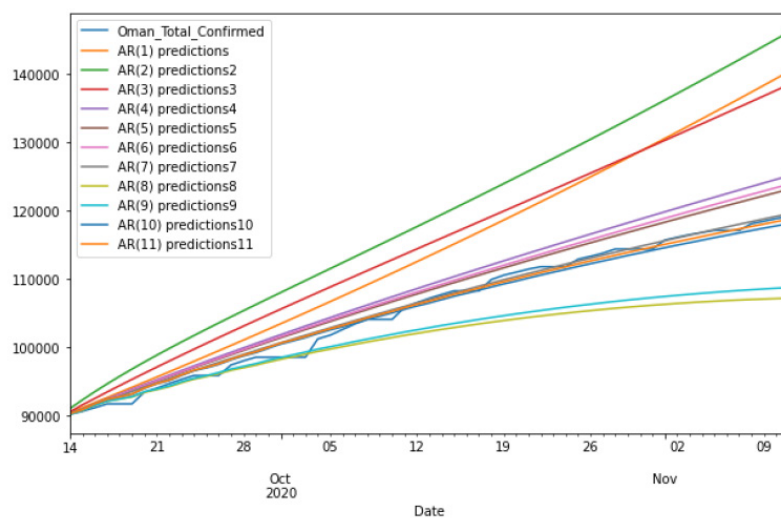
**Figure 17:** Testing TC in US.



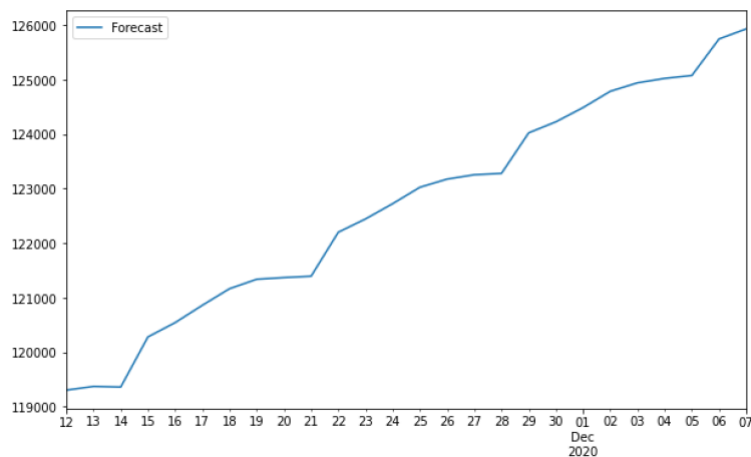
**Figure 18:** Forecasting TC in US.

Figures 17-18 depict the testing data and forecasting data of total confirmed cases using Auto Regression for the US. The same method as that for India followed to forecast the US data. The model

got best-fitting as AR12, which represents the total number of lags (i.e., lags = 12).

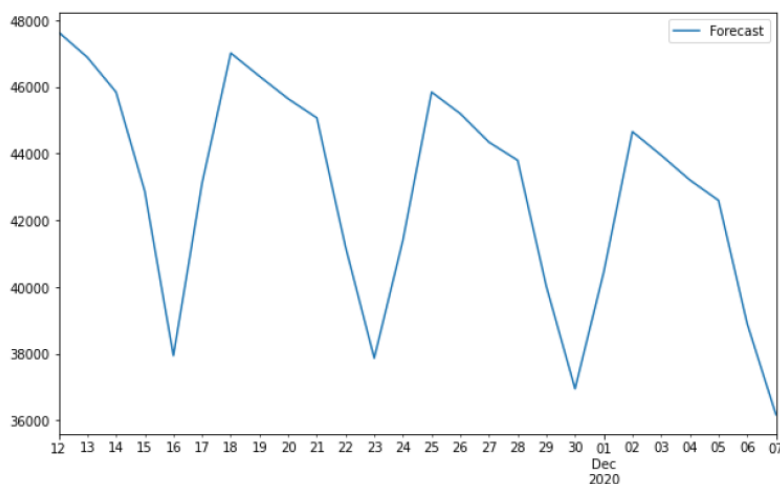
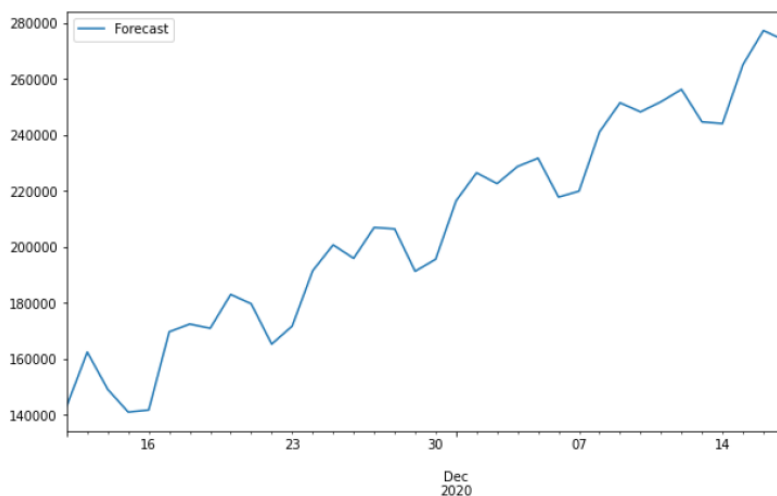


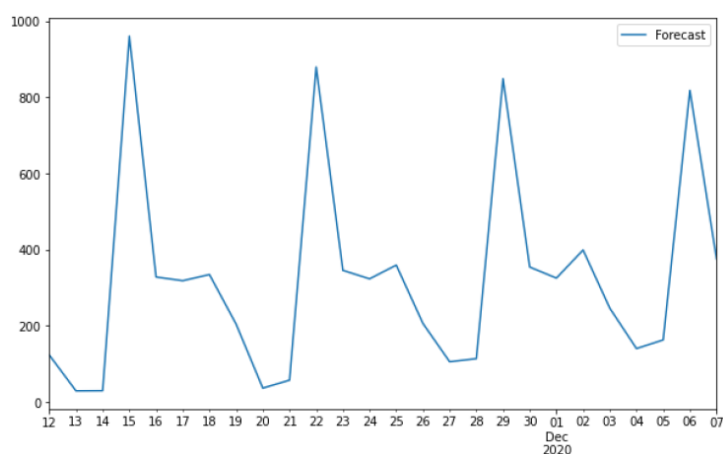
**Figure 19:** Testing TC in Oman.

**Figure 20:** Forecasting TC in Oman.

Figures 19-20 depict the testing data and forecasting data of total confirmed cases using Auto Regression for Oman. The same method as that for India followed to forecast the Oman data. The

model got best-fitting as AR11, which represents the total number of lags (i.e., lags = 11).

**Figure 21:** Forecasting DC in India.**Figure 22:** Forecasting DC in US.



**Figure 23:** Forecasting DC in Oman.

Figures 21-23 depict the forecasting data of daily confirmed cases for three countries which got procured similarly that of the total confirmed data.

**Table 5:** General Forecasting Model (VAR).

VAR Method Forecasting on 30 November 2020		
Country	Forecasting Cases	Number of cases
India	Total Confirmed	11,418,470
	Daily Confirmed	118,732
US	Total Confirmed	11,845,540
	Daily Confirmed	167,844
Oman	Total Confirmed	177,779
	Daily Confirmed	4283

**Results of Vector Auto Regression:** Forecasting was done using the Holt-Winters model on 30 November 2020 as viewed from Table 5. According to the numbers of total and daily confirmed cas-

es, it is deep-rooted that the US is hitting its peak when compared to India and Oman. Oman has the least number of confirmed cases when compared to other countries.

**Table 6:** ADF Test for TC.

Augmented Dickey-Fuller Test			
Test Details	India	US	Oman
ADF test statistic	-5.2996	-7.2007	-11.2392
p-value	0	0	0
Lags used	15	15	16
Observations	276	276	274
Critical Value (1%)	-3.4543	-3.4543	-3.4544
Critical Value (5%)	-2.8721	-2.8721	-2.8721
Critical Value (10%)	-2.5724	-2.5724	-2.5724
Strong evidence against the null hypothesis. Reject the null hypothesis.			
Data has no unit root and is stationary			

Augmented Dickey-Fuller Test performed on given data to check whether the data is stationary or non-stationary. Data was non-stationary till it reached the fourth difference (No. of Differenc-

es=4). At this point, data has a unit root as it was non-stationary. Further, we acquired the above-shown results as data has no unit root and turned out as stationary.

**Table 7:** Grid Search.

Grid Search for order p AR of VAR Model		
Test Details	INDIA & OMAN	INDIA & US
Order	9	9
AIC	32.346	27.036
BIC	32.919	27.609

**Table 8.**

Summary of Regression Results			
Model:	VAR		
Method:	OLS		
Date:	Sun, 15, Nov, 2020		
Time:	11:18:38		
No. of Equations:	2	BIC:	32.9192
Nobs:	227	HQIC:	32.5772
Log likelihood:	-4277.45	FPE:	1.12E+14
AIC:	32.3458	Det (Omega_mle):	9.51E+13

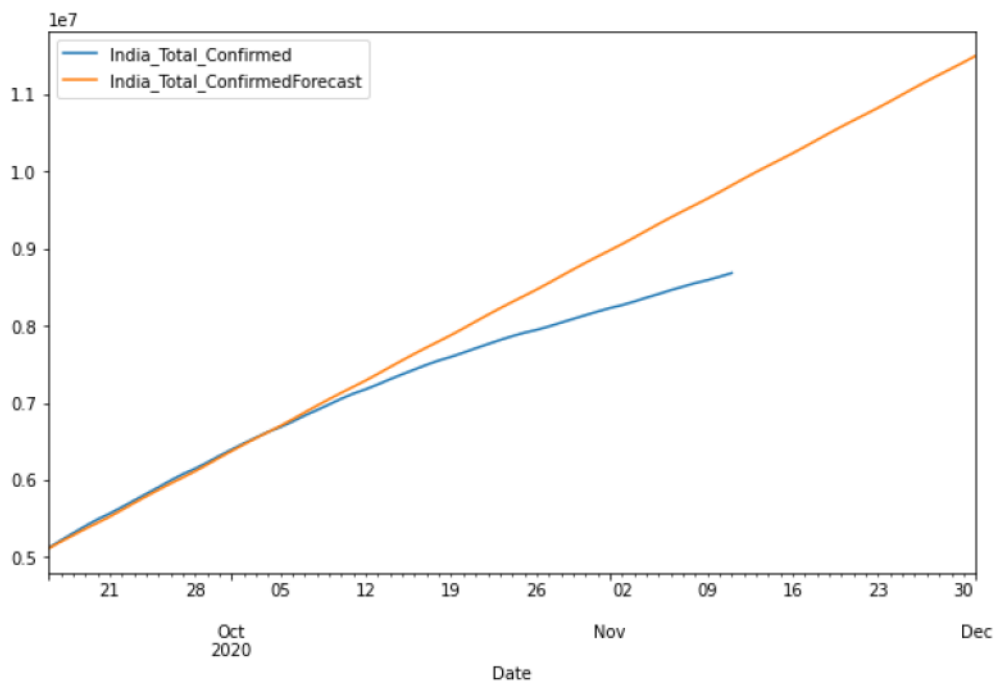
Results for equation India_Total_Confirmed				
	coefficient	std. error	t-stat	prob
const	-10.930831	183.687254	-0.06	0.953
L1. India_Total_Confirmed	-2.168493	0.065869	-32.921	0
1.1.US_Total_Confirmed	-0.054229	0.051342	-1.056	0.291
12.India_Total_Confirmed	-3.057165	0.144675	-21.131	0
L2.05_Total_Confirmed	-0.156156	0.10956	-1.425	0.154
L3. India_Total_Confirmed	-3.411088	0.210965	-16.169	0
L3.1.15_Total_Confirmed	-0.344316	0.149085	-2.31	0.021
14.India_Total_Confirmed	-3.401817	0.247796	-13.728	0
14.U5_Total_Confirmed	-0.432415	0.16573	-2.609	0.009
1.5. India_Total_Confirmed	-3.317106	0.257613	-12.876	0
15.16_Total_Confirmed	-0.413704	0.168985	-2.448	0.014
L6. India_Total_Confirmed	-2.916735	0.25354	-11.504	0
L6.05_Total_Confirmed	-0.540583	0.169759	-3.184	0.001
1.7. India_Total_Confirmed	-2.035179	0.221305	-9.196	0
1.7.U5_Total_Confirmed	-0.62476	0.1544	-4.046	0
1.8. India_Total_Confirmed	-1.023496	0.154577	-6.621	0
18.16_Total_Confirmed	-0.453235	0.114627	-3.954	0
19.India_Total_Confirmed	-0.249095	0.07203	-3.458	0.001
L9.1.15_Total_Confirmed	-0.204636	0.053697	-3.811	0

Results for equation U5_Total_Confirmed				
	coefficient	std. error	t-stat	prob
const	-4.973448	234.702909	-0.021	0.983
LLIndia_Total_Confirmed	-0.00099	0.084163	-0.012	0.991
LLUS_Total_Confirmed	-2.036661	0.065601	-31.046	0
L2. India_Total_Confirmed	-0.022588	0.184855	-0.122	0.903
12.U5_Total_Confirmed	-2.686761	0.139988	-19.193	0
L3. India_Total_Confirmed	0.087505	0.269557	0.325	0.745

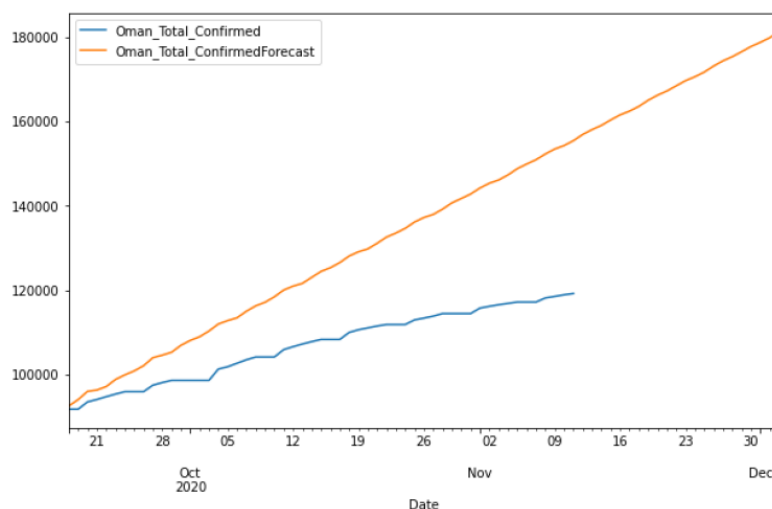
L3.U5_Total_Confirmed	-3.043259	0.190491	-15.976	0
L4. India_Total_Confirmed	0.005196	0.316616	0.016	0.987
L4.U5_Total_Confirmed	-3.143691	0.211759	-14.846	0
L5. India_Total_Confirmed	-0.11945	0.32916	-0.363	0.717
15.115_Total_Confirmed	-3.154236	0.215917	-14.609	0
L6. India_Total_Confirmed	-0.068808	0.323955	-0.212	0.832
L6.U5_Total_Confirmed	-2.868438	0.216906	-13.224	0
L7. India_Total_Confirmed	-0.025138	0.282768	-0.089	0.929
17.115_Total_Confirmed	-2.121906	0.197281	-10.756	0
L8. India_Total_Confirmed	-0.096319	0.197508	-0.488	0.626
L8.U5_Total_Confirmed	-1.11573	0.146462	-7.618	0
L9. India_Total_Confirmed	-0.106125	0.092035	-1.153	0.249
19.115_Total_Confirmed	-0.365085	0.068611	-5.321	0
Correlation matrix of residuals				
	India_Total_Confirmed	U5_Total_Confirmed		
India_Total_Confirmed	1	0.073464		
US_Total_Confirmed	0.073464	1		

We have found order  $p$  for the combination of two countries, India & US as well as India & Oman as the combination is a prerequisite in the VAR model. The order of  $p$  is dependent on the minimal AIC (Akaike Information Criteria) value. Once, the value of the

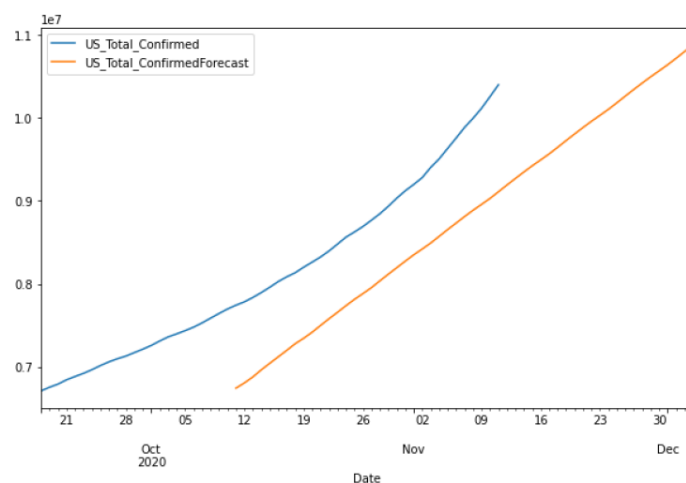
order is set, best-fit of the model takes place for the same order,  $p$  and further forecasting was done as shown in figures below. Finally, the result summary got acquired and can be seen below in Table 8.



**Figure 24:** Forecasting TC in India.



**Figure 25:** Forecasting TC in Oman.



**Figure 26:** Forecasting TC in US.

### Analysis of Best Forecasting Models for Three Countries

The current section treasures the best forecasting model considered for individual countries. Thus, meeting the primary aim of the research paper.

**Forecasting Analysis for India:** shows the actual and forecasted total confirmed cases for four models, and shows the percentage error to determine the efficiency of each model. The best forecasting model to predict the TC cases for India is "AR".

**Table 9:** Comparing the results of the Forecasting models of Total Confirmed cases for India.

Date	Actual Values	Prophet	Holt-Winter	AR	VAR
08-11-2020	85,53,657	86,10,141	84,57,352	85,51,145	8,276.38
09-11-2020	85,91,730	86,60,897	85,97,235	85,94,977	83,82,294
10-11-2020	86,36,011	87,22,012	86,21,393	86,36,279	84,85,454
11-11-2020	86,83,916	88,08,510	86,78,568	86,84,245	85,90,859
12-11-2020		88,84,456	87,30,309	87,31,080	87,00,497
13-11-2020		91,32,819	87,76,702	87,77,554	88,08,253
14-11-2020		91,99,846	88,23,095	88,22,953	89,11,468
15-11-2020		92,65,508	88,69,487	88,65,176	90,15,602



**Table 10:** Error Calculation for Forecasting methods in India.

Date	Prophet (Error %)	Holt-Winter (Error%)	AR (Error%)	VAR (Error %)
2020-11.08	0.7	1.1	0	3.2
202011-09	0.8	0.1	0	2.4
2020.11-10	1	0.2	0	1.7
2020.11-11	1.4	0.1	0	1.1
Best Method to Forecast India Total Confirmed cases is 'AR'				

**Forecasting Analysis for US:** Table 11 shows the actual and forecasted total confirmed cases for four models, and Table 12 shows the percentage error to determine the efficiency of each model. The best forecasting model to predict the TC cases for the US is "FB".

**Table 11:** Comparing the results of the Forecasting models of Total Confirmed cases for US.

Date	Actual values	Prophet	Holt-Winter	AR	VAR
08-11-2020	99,95,860	98,19,250	93,53,054	95,96,655	88,87,345
09-11-2020	10,115.80	98,79,938	94,28,122	96,67,757	89,58,032
10-11-2020	102,56,094	99,64,770	94,88,392	97,34,834	90,29,909
11-11-2020	104,00,227	99,92,183	95,72,541	98,03,279	91,09,990
12-11-2020		100,66,000	105,43,300	105,44,240	91,91,385
13-11-2020		101,36,198	106,86,380	107,03,670	92,70,350
14-11-2020		102,21,980	108,29,450	108,51,280	93,48,437
15-11-2020		102,51,750	109,72,530	109,90,530	94,24,173

**Table 12:** Error Calculation for Forecasting methods in US.

Date	Prophet (Error %)	Holt-Winter (Error%)	AR (Error%)	VAR (Error %)
2020-11.08	1.8	6.4	4.0	11.1
202011-09	2.3	6.8	4.4	11.4
2020.11-10	2.8	7.5	5.1	12.0
2020.11-11	3.9	8.0	5.7	12.4
Best Method to Forecast US Total Confirmed cases is 'Prophet'				

**Table 13:** Comparing the results of the Forecasting models of Total Confirmed cases for Oman.

	Actual Values	Prophet	Holt-Winter	AR	VAR
2020-11.08	118140	118.316	125.342	1,17,732	125.319
2020.11-09	118.503	118.811	125.978	1,18,090	126.522
2020.11-10	1,18,884	119.402	126.613	1,18,439	128.083
11-11-2020	1,19,186	1,19,890	1,27,249	1,18,788	1,29,088
2020.11-12		1,20,409	119.557	119.306	129.773
13-11-2020		1,20,699	1,19,928	1,19,374	1,31,111
2020-11.14		121.158	120.299	1,19,363	132577
2020-11.45		121.724	120.67	128281	133.529

**Table 14:** Error Calculation for Forecasting methods in US.

Date	Prophet (Error %)	Holt-Winter (Error%)	AR (Error%)	VAR (Error %)
2020-11.08	0.1	6.1	0.3	6.1
202011-09	0.3	6.3	0.3	6.8
2020.11-10	0.4	6.5	0.4	7.7
2020.11-11	0.6	6.8	0.3	8.3
Best Method to Forecast Oman Total Confirmed cases is 'AR'				

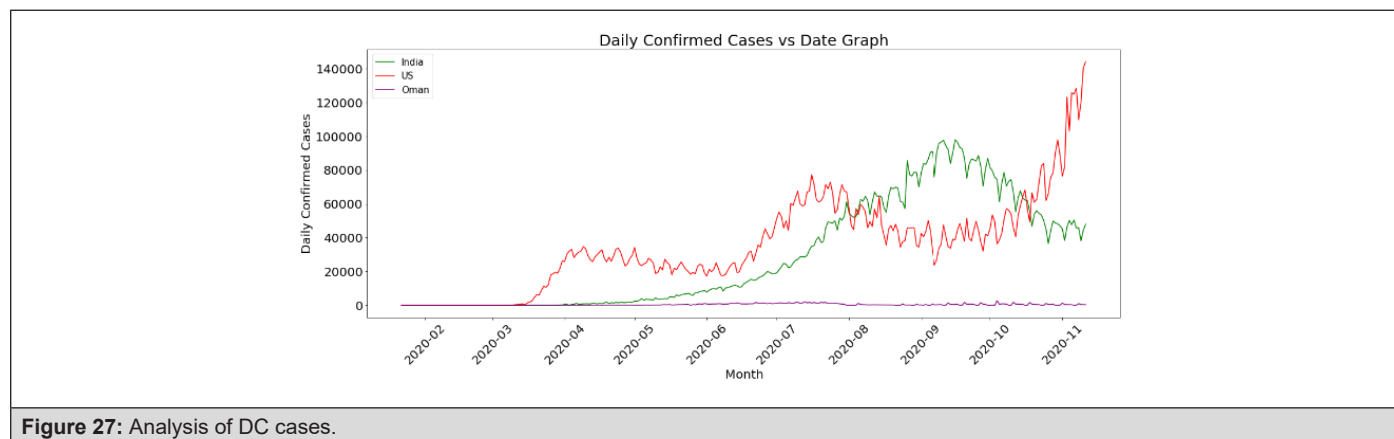
Table 13 shows the actual and forecasted total confirmed cases for four models, and Table 14 shows the percentage error to deter-

mine the efficiency of each model. The best forecasting model to predict the TC cases for Oman is "AR".

**Table 15:** Overall Forecasting Analysis.

Forecasting On 30 November 2020					
Country	Forecasting cases	Prophet	Holt-winters	AR	VAR
India	Total Confirmed	98,71,472	9,565,380	9,467,084	114,18,470
	Daily Confirmed	6,906	44,489	36.95	1,18,732
US	Total Confirmed	104,62,530	13,118,651	13,498,4	118,45,840
	Daily Confirmed	61,310	1,52,185	195,656	1,67,844
Oman	Total Confirmed	123,685	126,233	124,224	177,779
	Daily Confirmed	335	400	354	4283

From Table 15, highlighted total confirmed cases as per the efficacy of the forecasting models, are to be expected on 30 November 2020.

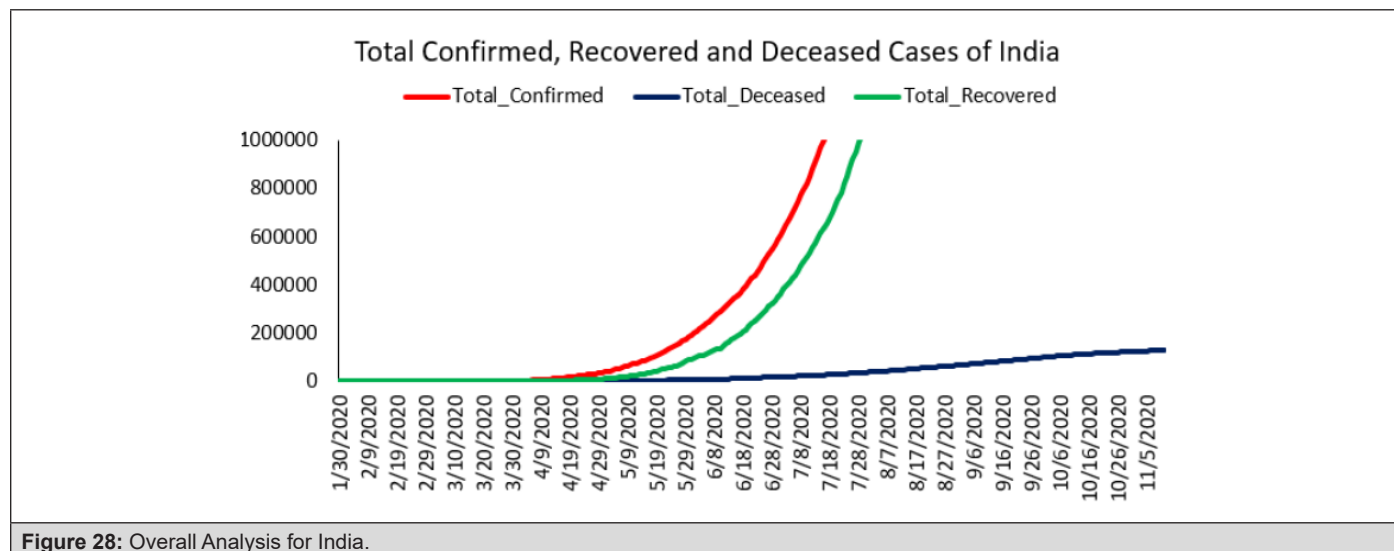


**Figure 27:** Analysis of DC cases.

**Analysis of Daily Confirmed Cases:** Figure 27 depicts the daily confirmed cases for the three countries, where the US overtook the other countries by a step ahead in terms of numbers. Further, it seems to appear that Oman has the least number of confirmed cas-

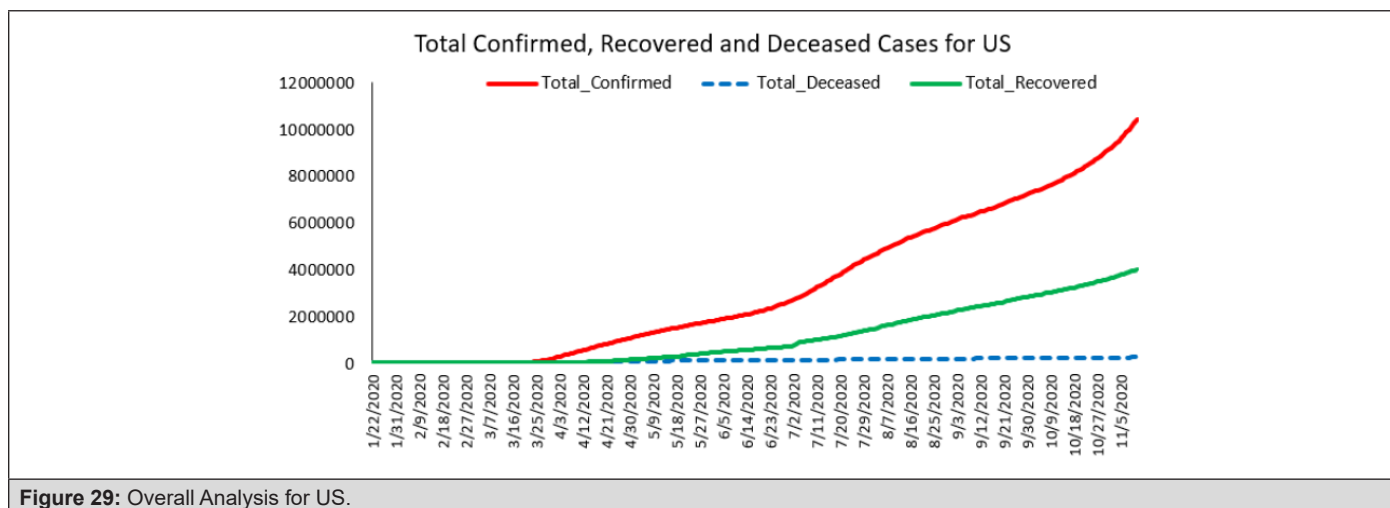
es from the above figure. However, the most affected country will get decided on by comparing the confirmed, recovery and death cases with a respective total population of each country.

#### Overall Analysis for India:



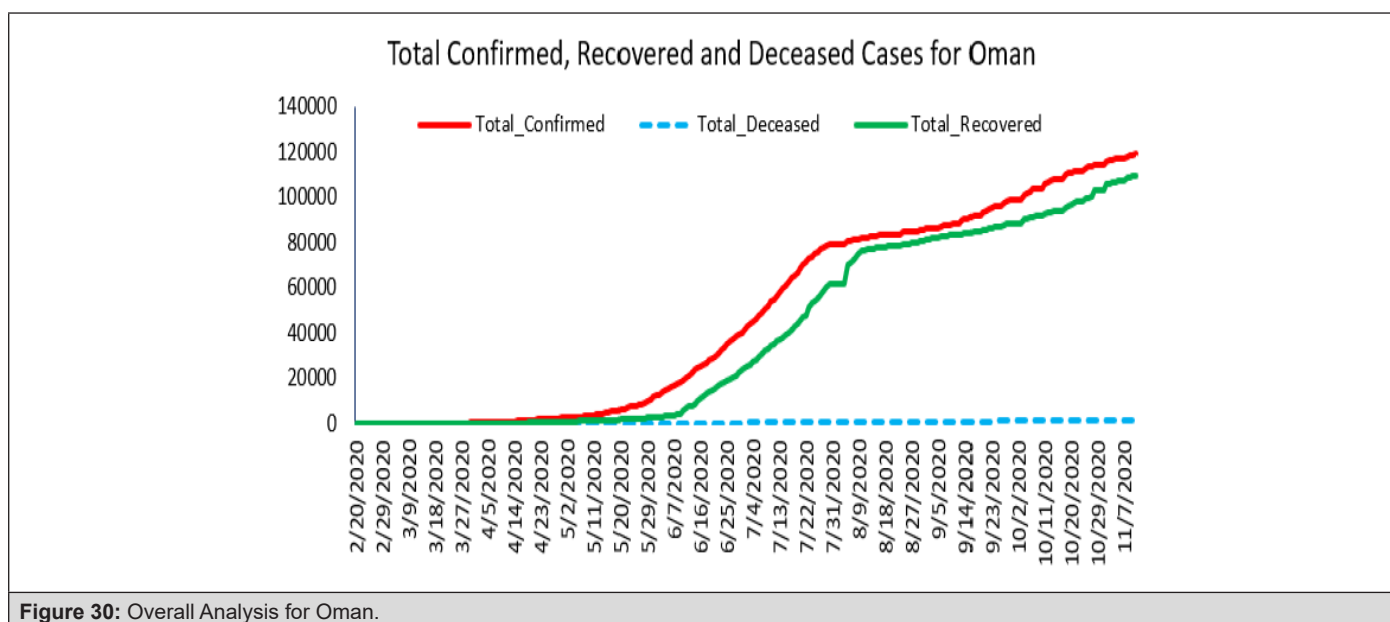
**Figure 28:** Overall Analysis for India.

## Overall Analysis for US:



**Overall Analysis for Oman:** Figures 28-30 reveal the Overall statistics of India, US and Oman individually based on the essential parameters such as confirmed cases, rate of recovery and deceased

cases. India is the least affected country compared to other countries (Figure 29).



## Overall Analysis for Three Countries:

Table 1.

Country	Total Population	TC Cases	TC %	TR Cases	TR %	TD Cases	TD%
India	1,380,004,385	86,83,916	0.63	80,66,501	92.89	1,28,121	1.48
US	3310,02,651	104,00,227	3.1	39,97,175	38.43	2,46,214	2.37
Oman	51,06,626	1,19,186	2.3	1,09,633	91.98	1,321	1.11
		Affected Country: US		Affected Country: US		Affected Country: US	

The following critical points were noticed from the table above:

- Considering the TC, TR and TD (i.e., Total Confirmed, Total Recovered & Total Death) cases, the US stood as the most affected country.
- Even though, Oman is not marked as the affected country

according to the per cent of TC, TR and TD, there is still an alarming call for residents of Oman though population is minimal compared to that of India and the US. The residents of Oman need to abide by the informed precautions from the government of Sultanate of Oman to avoid a hike in cases.

## Discussion

For the current study, relevant data collection was done from [8]. AR and FB models seem to be the best methods out of the four. It is due to the nature of data that resulted in the VAR method could not reach the potentials, even though it is a powerful method. The percentage error calculation helped to decide the best method. It is calculated, as shown below:

$$\% \text{ Error} = \frac{|\text{Actual Data} - \text{Forecast Data}|}{\text{Actual Data}} \times 100$$

The following are the precautionary measures to be implemented by the people globally.

- 1) Wear a mask
- 2) Maintain 6 feet of social distance
- 3) The area needs to be cleaned prior to touching the surface and after
- 4) Implement the habit of frequent hand wash
- 5) Stay home in case of feeling sick
- 6) Avoid gatherings
- 7) Consume Vitamin C & D

The future work involves the Deep learning time series methods to forecast COVID 19 data, Trend analysis and Future price predictions.

## Conclusion

The current research paper has successfully forecasted the total confirmed and daily confirmed cases using four essential Machine Learning techniques which comprises Facebook Prophet, Auto Regression, Vector Auto Regression and Holt-Winters respectively for three different countries (India, US and Oman). The results of the study reveal the best methods for the considered countries based on the calculation of Error percentage. The US surpassed the other two countries and stood as the most affected country. The overall comparison reveals that India is combating against COVID-19 effectively. Considering total population as a significant factor, Oman seems to be in the menace zone, and necessary stringent measures provided by the government of Oman must be followed by the residents of Oman to subdue the effects of the pandemic. Out of the four Machine Learning models, AR and FB models stood out as the best methods when compared to the other two.

## Acknowledgement

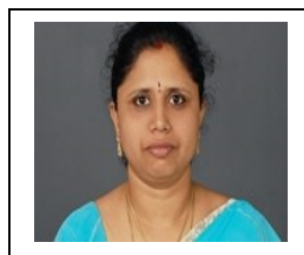
The authors want to acknowledge the management of esteemed institutes for support. Firstly, Military Technological College, Oman,

for the constant support and encouragement on the current research area. Further, we would like to thank the management of BITS Pilani, Hyderabad, India, and the National University of Science and Technology, Oman, for the backing and encouragement to drive the significant research work forward through extended collaboration across different countries.

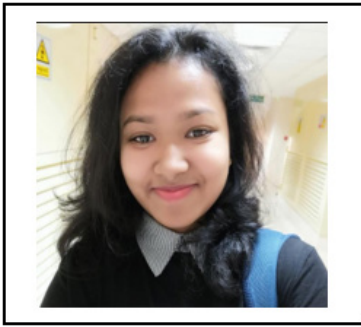
## Biographies



Dr Raja Rani is currently working as a Deputy Head of the General Studies Dept at Military Technological College, Oman and part of a Visiting Fellow Research at Portsmouth University, UK. Further, published several research papers in reputed National and International journals along with the conferences held at UK, Hong Kong, Oman and India. She worked as a peer reviewer of research papers in two international journals of Fluid Dynamics and worked as a co-author for a book published in 2014, entitled: Approximate Analytical Methods for Solving Ordinary Differential Equations, CRC Press (Taylor & Francis Group). She holds 26 years of experience in teaching Engineering Mathematics for under-graduates and post-graduates. Moreover, worked as an Assistant and Associate Professor in various Engineering colleges in India and current research interests are in mathematical modelling with applications to real-world problems focusing on heat and mass transfer (for porous and non-porous medium), bio-fluids, control systems engineering, fractional differential equations, machine and deep learning.



Dr T S L Radhika is working as an Assistant professor in BITS Pilani, Hyderabad Campus, India. An applied mathematician by education and statistician by training she works on mathematical and statistical models for understanding some physical and natural phenomenon. Currently, she is working on models for human arterial blood flow and fluid transport in porous bodies in nonconventional coordinate systems, the spheroidal and bipolar.



Shalini Pukkella currently pursued BEng Computer-Aided Mechanical Engineering (Hons) at the National University of Science and Technology, Oman (affiliated to Glasgow Caledonian University, UK). She Secured a CGPA of 3.79 out of 4. Partakes keen interest in vital research and participated in many student activities and college events on the campus. She worked as a treasurer of ASME on behalf of NUST University. She interned at Glasgow Caledonian University, Scotland, during her final year of study. The first research publication was made in the International Research Journal of En-

gineering and Technology for the paper named "Design and Development of Pick and Place Arm Robot."

### References

1. J Taylor S, Ben L (2017) Prophet: forecasting at scale. PeerJPreprints.
2. (2020) Facebook'S Open Source Forecasting Tool, Prophet. Cartesian Consulting.
3. P Tutorials (2020) What Is an Autoregressive Model?. 365 Data Science.
4. Maitra S (2020) Vector Autoregressive For Forecasting Time Series. Towards data science.
5. Ioannidis J, Cripps S, Tanner M (2020) Forecasting for COVID-19 has failed. International Journal of Forecasting.
6. Unlu R, Namlı E (2020) Machine Learning and Classical Forecasting Methods Based Decision Support Systems for COVID-19. Computers, Materials & Continua 64(3): 1383-1399.
7. Ardabili S, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi Koczy A. et al. (2020) COVID-19 Outbreak Prediction with Machine Learning. medRxiv 13(10): 249.
8. (2020) Novel Coronavirus (COVID-19) Cases Data. HDX.