International Journal of Management and Humanities (IJMH)
ISSN: 2394-0913 (Online), Volume-4 Issue-11, July 2020

Implication of ARIMA Time Series Model on COVID-19 Outbreaks in India

D. Mallikarjuna Reddy

Abstract*: This research paper focuses on a Time Series Model to predict COVID-19 Outbreaks in India. COVID-19 Corona virus disease has been recognized as a worldwide hazard, and most of the studies are being conducted using diverse mathematical techniques to forecast the probable evolution of this outbreak. These mathematical models based on various factors and analyses are subject to potential bias. Here, we put forward a natural Times Series (TS) model that could be very useful to predict the spread of COVID-19. Here, a popular method Auto Regressive Integrated Moving Average (ARIMA) TS model is performed on the real COVID-19 data set to predict the outbreak trend of the prevalence and incidence of COVID-19 in India. Every day data of fresh COVID-19 confirmed cases act as an exogenous factor in this frame. Our data envelops the time period from 12th March, 2020 to 27th June, 2020. The time series under study is a non-stationary. Results obtained in the study revealed that the ARIMA model has a strong potential for prediction. In ACF and PACF graphs. Lag 1 and Lag 40 was found to be significant. Regressed values imply Lag 1 and Lag 40 was significant in predicting the present trend. The model predicted maximum COVID-19 cases in India at around 14, 22,337 with an interval (12, 80,352 - 15, 69, 817) during 1st July to 30th July period cumulatively. As per the model, the number of new cases shall increases drastically in India only. The results will help governments to make necessary arrangements as per the estimated cases. This kind of investigation, implications of ARIMA models and fitting procedures are useful in forecasting COVID-19 Outbreaks in India.

***Keywords*: ARIMA, Time Series, COVID-19, ACF, PACF, Forecasting**

I. INTRODUCTION

Novel Virus belongs to Corona viruses' family that has been spread from animal to human beings and it was found in Wuhan, China, in December 2019. This can cause serious illness and even death too [1]. It's been identified as a Zoonotic corona virus, similar to severe acute respiratory syndrome corona virus (SARS-CoV) and Middle East Respiratory Syndrome Corona virus (MERS-CoV), and was named 2019-nCoV (2). The incidence scope of this illness is indistinguishable, since at present the occurrences of this virus infection are so dynamic [1]. There is clear variation among countries in epidemiological surveillance and detection capacity for suspected cases. Several cases of COVID-19 infections were also reported outside China [3]. Fore studies reveal that The first serious case of COVID-19 epidemic was reported by Chinese authorities along with WHO on 11th January 2020 whose features are similar to pneumonia of unknown origin that caused maximum spoil to developed countries similar to America, Spain, Russia, UK, Italy, France, Germany, Iron etc., affecting 215 countries and territories around the Globe. This COVID-19 pandemic decease has infected almost 10,097,334 people and caused deaths of 5, 01,585 people across the world till June 27th June, 2020. Around 5,473,348 people recovered also without any special treatment. Also in number 4,122,401 were currently infected patients from them 4,064,578 (99%) in mild condition 57,823 (1%) serious or critical stage 5, 974,933 cases which had an outcome that is about closed cases throughout world 5,473,348 (92%) recovered and discharged and 5, 01, 585 (8%). Deaths (https://www.worldometers.info/coronavirus). Where as in India the first case was reported on 15th February, 2020 [4] and thereafter it has been continued and reached to 5,29,889 cases with deaths 16,112 and Recovered case 3,10,236 with 58.54 percent recovery rate on 28th June, 2020 (https://www.worldometers.info/coronavirus/country/india). In this condition when the illness does not have any specific treatment, the prevention of disease and preparation in healthcare services is very important. Modeling and future forecast of daily number of confirmed cases can help the treatment system in providing services for the new patients. The statistical prediction models [5-6] could be helpful in forecasting and controlling this global epidemic threat. Here in this study, Auto Regressive Integrated Moving Average (ARIMA) model [7] could be constructive to predict confirmed cases of COVID-19. The best ARIMA model is identified, and then 60 future days (2-months) is predicted in India. Since the present trend is not supporting the results discussed in[4]. The daily occurrence data of COVID-19 confirmed cases during 12th March, 2020 to 27th June, 2020 were collected from the official website of Govt. of India (https://www.mygov.in/covid-19) and were used to build these models. The purpose of this study is first to find the best predicting models for confirmed cases in the country. These models can suggest in predicting patients in near future to have better treatment in Indian hospitals. rectification is not possible.

II. DESCRIPTION OF THE DATA

Author (s) can send paper in the given email address of the journal. There are two email address. It is compulsory to send paper in both email address. The daily occurrence data of COVID-19 confirmed cases during 12th March, 2020 to 27th June, 2020 were collected from the official website of Govt. of India

Revised Manuscript Received on July 15, 2020.
* Correspondence Author
D.Mallikarjuna Reddy*, Department of Mathematics, School of Science GITAM Deemed to be University, Hyderabad, India 502329: mallik.reddyd@gmail.com

Retrieval Number: K10570741120/2020©BEIESP
DOI: 10.35940/ijmh.K1057.0741120
Journal Website: www.ijmh.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

(https://www.mygov.in/covid-19), and SPSS 20 software was used to build a time-series model Auto Regressive Integrated Moving Average (ARIMA) [1]. ARIMA model was applied to this dataset consisting of 108 rows day wise determinations cumulatively. Fig. 3 shows that the overall frequency of COVID-19 presented an increasing trend that is reaching the outbreaks in table. The difference between cases of one day and cases of the previous day Δ (Z_t-$Z_{(t-1)}$)) a variable increase in the number of confirmed cases. Descriptive analysis of the data was performed to evaluate the incidence of new confirmed cases of COVID-19 and to prevent eventual bias.

III. Methodology -ARIMA (p,d,q) Model

The ARIMA model introduced by Box and Jenkins (1976) includes Autoregressive as well as moving average parameters and explicitly includes differencing in the formulation of the model. Specifically, the three types of parameters in the model are: the autoregressive parameters (p), the number of differencing passes (d), and moving average parameter (q). In the notation introduced by Box and Jenkins, models are summarized as ARIMA (p, d, q).

The foremost step in the process of modeling is to check for the stationary of the time series data. This is done by observing the graph of data or autocorrelation and the partial autocorrelation functions. Another way of checking stationary is to fit the first order AR model to the raw data and test whether the coefficient Ø is less than one. The next step is to identify an appropriate sub-class of the general ARIMA model.

$$\phi(B)\nabla^d Z_t = \theta(B)a_t \tag{1}$$

Which may be used to represent a given time series.The general procedure is

(i) To difference Z_t as many times as is needed to produce stationary.

$$\phi(B)\omega_t = \theta(B)a_t \tag{2}$$

where $\omega_t = (1-B)^d Z_t = \nabla^d Z_t$ (3)

To identify the resulting ARIMA process. The Autocorrelation and Partial Autocorrelation functions will be used as main tools in attaining (i) and (ii) Stationary in time series means a constant mean, variance and Autocorrelation through time. Generally the hypothetical series will not be stationary. Therefore the series needs to differenced until it is stationary (usually at most two differences will be sufficient to make the series stationary).

$$\omega_t = \nabla^d Z_t \tag{4}$$

After differencing the series to identify the sub class of ARIMA model, which is suitable for the series, the components of Autoregressive (p) and Moving Average (q) should be identified. The number of parameters of Autoregressive (p) can be identified by using correlogram of partial autocorrelations and the number of parameters of moving average (q) can be identified by using correlogram of autocorrelations. Then p parameters of Autoregressive $Ø_1$, $Ø_2$, - - - , $Ø_p$ and then q parameters of Moving Average θ_1, θ_2, - - - θ_q have to be estimated by using Least Squares (LS) or Maximum Likelihood (ML) methods.

IV. Results & Discussions

To obtain the daily occurrences of COVID-19, ARIMA (p, d, q) was chosen as the top ARIMA model,[8]. Thus i.e ARIMA (0, 2, 1) (p=0, d=2, q=1) has been fitted to forecast COVID-19 Confirmed Cases (Cumulative) using SPSS-20 and a fore study was taken as reference for the methodology of the analysis [9]. Logarithmic transformation was performed to evaluate the influence of seasonality on the forecast. The COVID-19 daily conformed new cases have been considered to identify the order of an moving average (MA) autoregressive (AR) model. A plot of the daily new COVID-19 cases against time is presented fig.1.

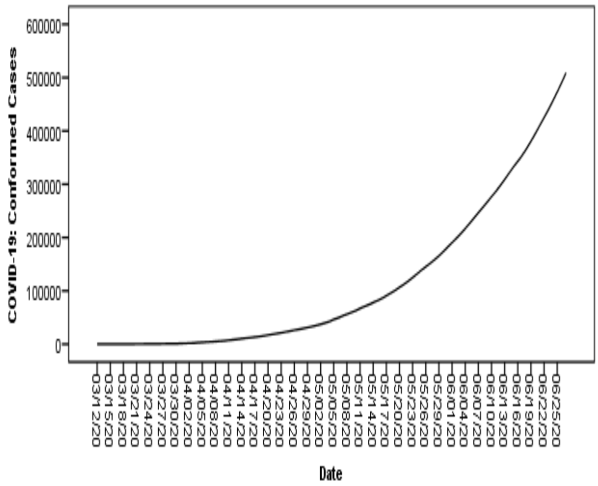

Fig. 1.Time Series plot of daily new Conformed COVID-19 cases

The correlogram(i.e ACFv/sLag) reporting the ACF and PACF [10] showed that significant thus Fig. 2 is the plot of autocorrelation function Graph of COVID-19 Daily new cases at 1st differencing. The Lag 1-lag16 all lags were showed significant and which was reported in Table 1. The underlying process assumed is independence (white noise) and based on the asymptotic chi-square approximation. The statistical significance level was considered at 0.05.

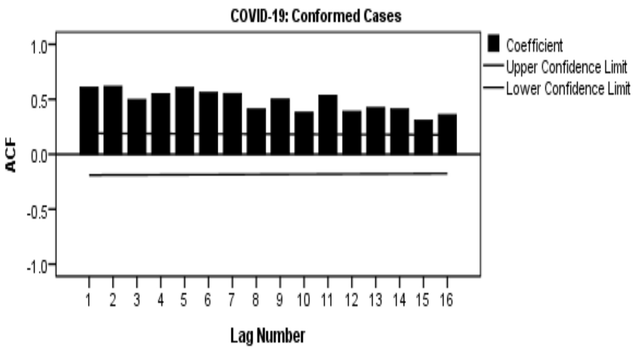

Fig. 2.Correlogram graph for the COVID-19 confirmed cases .

Retrieval Number: K10570741120/2020©BEIESP
DOI: 10.35940/ijmh.K1057.0741120
Journal Website: www.ijmh.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table- I: Autocorrelations of COVID-19: Conformed Cases

Autocorrelations					
Series: COVID-19: Conformed Cases					
Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic		
			Value	df	Sig.[b]
1	.608	.095	41.054	1	.000
2	.617	.094	83.656	2	.000
3	.497	.094	111.555	3	.000
4	.548	.094	145.860	4	.000
5	.606	.093	188.203	5	.000
6	.560	.093	224.708	6	.000
7	.551	.092	260.468	7	.000
8	.410	.092	280.469	8	.000
9	.501	.091	310.561	9	.000
10	.381	.091	328.119	10	.000
11	.533	.090	362.925	11	.000
12	.389	.090	381.662	12	.000
13	.423	.089	404.055	13	.000
14	.410	.089	425.253	14	.000
15	.307	.088	437.329	15	.000
16	.359	.088	453.962	16	.000
a. The underlying process assumed is independence (white noise).					
b. Based on the asymptotic chi-square approximation.					

Fig. 3 is the plot of Partial autocorrelation function (PACF) Graph of COVID-19 daily conformed new cases at 1st Differencing In PACF graph. Lag 1 and Lag 16 is significant. regressed values imply Lag 1 and Lag 16 is significant in predicting the current values. All the Partial Autocorrelations of COVID-19: Confirmed Cases with Lag kept in Table-2.
The models returned the highest stationary R2 and relatively lower BIC values showed in Table-3.
Table -5 shows the different parameters of autoregressive (p) and moving average (q) among the several ARIMA model experimented upon. ARIMA (0, 2, 1) is considered best for COVID-19 prediction. The investigational results kept in Table -6 which are predicted values of ARIMA (0, 2, 1) that has been found the best model for prediction of COVID-19 cases in India. Fig.4 reveals graphical visualization of predicted COVID-19 daily conformed cases in India.

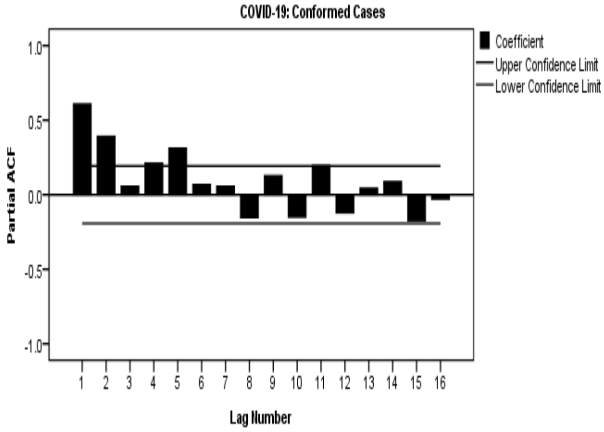

Fig.3 PACF for the COVID-19 confirmed cases

Table-II Partial Autocorrelations of COVID-19: Conformed Cases

Partial Autocorrelations		
Series: COVID-19: Conformed Cases		
Lag	Partial Autocorrelation	Std. Error
1	.608	.096
2	.392	.096
3	.057	.096
4	.211	.096
5	.312	.096
6	.069	.096
7	.057	.096
8	-.154	.096
9	.128	.096
10	-.149	.096
11	.196	.096
12	-.122	.096
13	.044	.096
14	.088	.096
15	-.174	.096
16	-.028	.096

Table- III: Model Fit of COVID-19: Confirmed Cases

Model Fit										
Fit Statistic	Mean	Min.	Max.	Percentile						
				5	10	25	50	75	90	95
Stationary R-squared	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
R-squared	1	1	1	1	1	1	1	1	1	1
RMSE	456.88	456.88	456.88	456.88	456.88	456.88	456.88	456.88	456.88	456.88
MAPE	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03	2.03
MaxAPE	13.87	13.87	13.87	13.87	13.87	13.87	13.87	13.87	13.87	13.87
MAE	310.72	310.72	310.72	310.72	310.72	310.72	310.72	310.72	310.72	310.72
MaxAE	1615.84	1615.84	1615.84	1615.84	1615.84	1615.84	1615.84	1615.84	1615.84	1615.84
Normalized BIC	12.337	12.337	12.337	12.337	12.337	12.337	12.337	12.337	12.337	12.337

Retrieval Number: K10570741120/2020©BEIESP
DOI: 10.35940/ijmh.K1057.0741120
Journal Website: www.ijmh.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table- IV: Model Statistics of COVID-19: Confirmed Cases

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
COVID-19: Conformed Cases -Model_1	0	.349	24.869	17	.098	0

Model Statistics

Table- V: ARIMA Model Parameters

					Estimate	SE	t	Sig.
COVID-19: Conformed Cases -Model_1	COVID-19: Conformed Cases	Square Root	Constant		.115	.024	4.720	**.000**
			Difference		2			
			MA	Lag 1	.720	.070	10.283	**.000**

Table- VI ARIMA (p, d, q) forecast values of COVID-19 (Cumulative)

Model	*COVID-19: Confirmed Cases -Model_1(Cumulative)*			*Model*	*COVID-19: Confirmed Cases -Model_1(Cumulative)*		
	Forecast	UCL	LCL		Forecast	UCL	LCL
6/28/2020	527187	529708	524671	8/1/2020	1502402	1666926	1344338
6/29/2020	545913	550081	541756	8/2/2020	1543689	1717161	1377202
6/30/2020	565139	571022	559278	8/3/2020	1585829	1768537	1410656
7/1/2020	584874	592607	577179	8/4/2020	1628834	1821071	1444709
7/2/2020	605129	614868	595448	8/5/2020	1672713	1874778	1479369
7/3/2020	625911	637822	614084	8/6/2020	1717481	1929677	1514643
7/4/2020	647230	661486	633091	8/7/2020	1763149	1985782	1550540
7/5/2020	669096	685874	652474	8/8/2020	1809728	2043111	1587068
7/6/2020	691518	710998	672241	8/9/2020	1857232	2101680	1624235
7/7/2020	714505	736871	692398	8/10/2020	1905671	2161507	1662050
7/8/2020	738067	763506	712952	8/11/2020	1955059	2222609	1700520
7/9/2020	762214	790916	733911	8/12/2020	2005408	2285001	1739654
7/10/2020	786956	819113	755281	8/13/2020	2056730	2348702	1779461
7/11/2020	812301	848111	777071	8/14/2020	2109038	2413730	1819949
7/12/2020	838261	877924	799286	8/15/2020	2162345	2480100	1861126
7/13/2020	864845	908564	821935	8/16/2020	2216662	2547831	1903002
7/14/2020	892064	940046	845025	8/17/2020	2272004	2616941	1945584
7/15/2020	919926	972382	868563	8/18/2020	2328382	2687447	1988881
7/16/2020	948443	1005587	892555	8/19/2020	2385810	2759367	2032903
7/17/2020	977626	1039676	917011	8/20/2020	2444301	2832718	2077658
7/18/2020	1007483	1074661	941935	8/21/2020	2503868	2907520	2123155
7/19/2020	1038026	1110558	967337	8/22/2020	2564523	2983789	2169403
7/20/2020	1069266	1147381	993223	8/23/2020	2626281	3061545	2216410
7/21/2020	1101212	1185145	1019601	8/24/2020	2689155	3140805	2264187
7/22/2020	1133877	1223864	1046477	8/25/2020	2753157	3221588	2312741
7/23/2020	1167269	1263553	1073860	8/26/2020	2818302	3303913	2362083
7/24/2020	1201401	1304227	1101757	8/27/2020	2884603	3387797	2412221
7/25/2020	1236283	1345902	1130175	8/28/2020	2952074	3473260	2463164
7/26/2020	1271927	1388592	1159122	8/29/2020	3020728	3560321	2514923
7/27/2020	1308343	1432313	1188606	8/30/2020	3090580	3648999	2567507
7/28/2020	1345542	1477081	1218634	8/31/2020	3161643	3739312	2620924
7/29/2020	1383536	1522910	1249213	9/1/2020	3233930	3831279	2675184
7/30/2020	1422337	1569817	1280352	9/2/2020	3307457	3924921	2730298
7/31/2020	1461955	1617817	1312058	9/3/2020	3382238	4020255	2786274

For each model, forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

Retrieval Number: K10570741120/2020©BEIESP
DOI: 10.35940/ijmh.K1057.0741120
Journal Website: www.ijmh.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

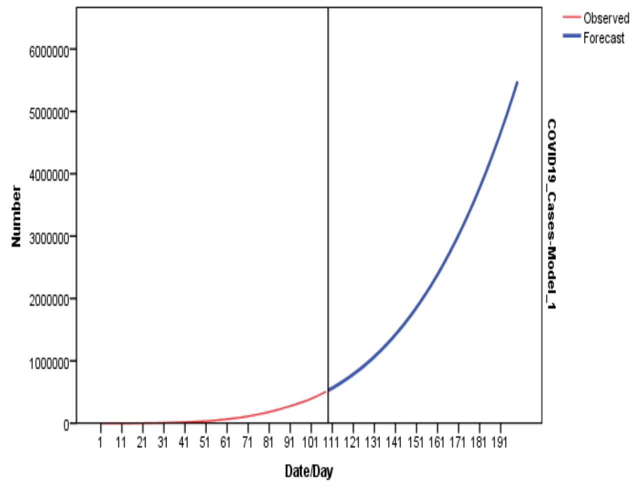

Fig.4 Forecasting COVID-19 confirmed cases (Cumulative) using ARIMA (0, 2, 1)

V. CONCLUSION

The results obtained with the model have been confirmed with the prospective of ARIMA models. The model implicated to predict COVID-19 cases on daily basis satisfactorily on short-term basis. The model predicted maximum COVID-19 cases in India at around 14, 61, 955 end of July 2020 within the range of (13, 12,058-16, 17,817) with 95 percent confidence level. As per the model, the number of new cases shall increases drastically in India may be in the month of July and August 2020. The results will help governments to make necessary arrangements as per the estimated cases. This kind of analysis and implications of ARIMA models and fitting procedures are useful in forecasting COVID-19 Outbreaks in India..

REFERENCES

1. Paules CI, Marston HD, Fauci AS. Corona virus infections—more than just the common cold. JAMA. 2020;323(8):707-8.
2. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS corona virus. Journal of Travel Medicine. 2020.
3. Niehus R, De Salazar PM, Taylor A, Lipsitch M. Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. medRxiv. 2020.
4. Sitaram pandey, Amitava Samanta. Real-time forecasting of COVID-19 prevalence in India using ARIMA model International Journal of Management and Humanities (IJMH), Volume-4 Issue-10, pp 78-83, June 2020.
5. Manohar Dingari, D.Mallikarjuna Reddy, V. Sumalatha Air Traffic Forecasting using Time Series Models International Journal of Recent Technology and Engineering Volume-8 Issue-4, November 2019.
6. Manohar Dingari, D. Mallikarjuna Reddy, V. Sumalatha Air Traffic Forecasting Using Artificial Neural Networks International Journal of Scientific & Technology Research Volume 8, Issue 10, October 2019.
7. Manohar Dingari, D. Mallikarjuna Reddy, V. Sumalatha Time Series Analysis for Long Memory Process of Air Traffic Using Arfima International Journal of Scientific & Technology Research Volume 8, Issue 10, October 2019
8. G. Baiocchi, W. Distaso, GRETL: econometric software for the GNU generation, J. Appl. Econom. 18 (2003) 105e110.
9. Y.W. Wang, Z.Z. Shen, Y. Jiang, Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China, PloS One 13.
10. D. Mallikarjuna Reddy "Second Order Statistics of Time Series of Various Real Time Problems in Conjunction with Periodogram Technique" International Journal of Latest Trends in Engineering and Technology Vol. 3 Issue 1 2013
11. J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, A. Lachhab, Forecasting of demand using ARIMA model, Int. J. Eng. Bus. Manag. 10 (2018).
12. S. Cao, F. Wang, W. Tam, L.A. Tse, J.H. Kim, J. Liu, Z. Lu, A hybrid seasonal prediction model for tuberculosis incidence in China, BMC Med. Inf. Decis. Making 13 (2013) 56.
13. Y.-W. Cheung, K.S. Lai, Lag order and critical values of the augmented Dickeye Fuller test, J. Bus. Econ. Stat. 13 (1995) 277e280.
14. C. COVID & R.Team., ―Severe outcomes among patients with corona virus disease 2019 (COVID-19)—United States‖, February 12–March 16, 2020. MMWR Morb Mortal Wkly Rep, 69(12), pp. 343-346.
15. J. Gao, Z. Tian,, & X. Yang , ―Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies‖, 2020, Bioscience trends.
16. D. Fanelli,, & F. Piazza, ―Analysis and forecast of COVID-19 spreading in China, Italy and France‖. Chaos, Solitons & Fractals, 2020, 134, 109761.
17. T. Chakraborty & I. Ghosh, (2020). ―Real-time forecasts and risk assessment of novel corona virus (COVID-19) cases: A data-driven analysis‖. Chaos, Solitons & Fractals, 2020, 109850.
18. Z. Yang,, Z. Zeng,, K. Wang,, S.S. Wong, W. Liang, M. Zanin, ... & J. Liang , ―Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease‖, 2020, 12(3), pp 165.
19. Z. Ceylan , ―Estimation of COVID-19 prevalence in Italy, Spain, and France. Science of the Total Environment‖, 2020, 138817.
20. A.A. Ariyo, A.O. Adewumi, & C.K. Ayo, ―Stock price prediction using the ARIMA model‖. In 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation (pp. 106-112). IEEE.

AUTHORS PROFILE

Dr D.Mallikarjuna Reddy, completed his BSc (MPC) from Kakatiya University in the year 1995, MSc (Statistics) from Kakatiya University in the year 1998 and PhD from Kakatiya University, Warangal, Telangana State India in the year 2012. He has Twenty years of teaching experience including NIT Warangal, and ten years of research experience He has worked as research fellow for DST sponsored project Govt.of India under the guidance of Prof. Mallareddy Perati, Department of Mathematics Kakatiya University, Warangal. His area of research is Queuing Systems, Performance Analysis of Routers, Time Series, Organizational Culture-IT industries India and Bio-Statistics. Dr Reddy has published 35 publications in reputed international journals with one book entitled "Engineering Mathematics for Gate students, S. Chand Publishers in 2017 and chapters in other two books. Dr .Reddy received a best teacher award state level, Telangana State for the year 2016-17 from SRF foundation, Telangana . He presented his research work in 15 International conferences and participated more than 40 workshops and FDP's. He is a Reviewer and EB member for 15 international reputed journals like IEEE, IJCNDS, AIP, IANG etc. Also He acted as Treasurer for ICMSA-2019 international conference conducted by Department of mathematics GITAM Hyderabad. Under his able guidance seven research scholars perusing their Ph Ds. Two of them submitted and waiting for defense. Also, He has done data analysis using MINI TAB, SPSS, for 200 plus dissertations/project works, research articles and acknowledged. He is currently working as Assistant Professor in the Department of Mathematics, School of Science, For GITAM Deemed to be University, Hyderabad, Telangana India-502329.