INTERNATIONAL JOURNAL OF STATISTICS AND MEDICAL INFORMATICS

Vol. 13 No. 1/2020, pp: 1-7

Overview of different models for predicting COVID-19 Cases

Editor, IJSMI

1. Introduction

During the month of Dec 2019 Chinese people from Wuhan the capital of Hubei Province were affected by the Virus called 'Covid-19' (initially it was called as Novel Coronavirus (2019-nCoV)) which was originated from the Sea food market as per the Chinese Government reports. During the first week of January 2020 Chinese Authorities identified this new virus [1]. It slowly started spreading to other parts of China. During the month January it started spreading to Thailand, South Korea and Japan through travellers from the Wuhan City [1]. Number of papers have already published in predicting the COVID-19 cases using mathematical and artificial intelligence modelling [2, 3, 4, 5]. This paper provides evaluation of different predictive models such as statistical models, machine learning and deep learning models for predicting the COVID cases which was declared as 'Pandemic' by World Health Organization (WHO). This paper used situation reports of WHO [1] as the basis for the predicting the number of cases which was reported daily and included country wise statistics. The focus of the paper is on academic part of the problem. Open Source statistical software R along with GUI tool R Studio [6, 7] is used to build the forecasting model.

2. Methodology

The paper uses different predictive models using Time series models [8, 9, 10, 11] and Machine Learning [12,13] models.

2.1. Time series models for forecasting

Time series models helps to predict the future values based on the past observation which are associated with time values. The time values can be day, month or years. Some of the models used in the time series models are given below

2.1.1. Auto Regressive Moving Average (ARIMA)

Auto Regressive Moving Average models[14] includes three components namely Autoregressive (AR) , Integrated(I) and Moving Averages which is used to predict the future values of the series from its past values wherein the regression function happens on its own past values.

2.1.2. General Autoregressive Conditional Heteroscedasticity (GARCH) models to model volatility

General Autoregressive Conditional Heteroscedasticity [15] Models are non-stationary models wherein changes in the variance is calculated from the series past values of variance. It includes both variance of the error terms and the variance itself.

2.1.3. Simple Exponential Smoothing model

Simple Exponential smoothing model[12] provides more weight to recent observation and less weight to older observations.

Vol. 13 No. 1/2020, pp: 1-7

Readers are encouraged to refer author's Book on for more details about Forecasting [17]. (www.ijsmi.com/book.php)

2.2. Machine learning models

Machine learning models are built using computer algorithms which learn on its own in the due course to perform specific task such prediction and classification under supervised or unsupervised context.

Different Machine Learning models which are used in forecasting are given below:

2.2.1. Artificial Neural Network – Multi Layer Perceptron(MLP)

Multi-Layer Perceptron (MLP) [18,19] model includes input layer, hidden layer and output layer in the prediction process wherein the input layer receives the raw data input and passes to the hidden layer with weights assigned to the input data using transfer function. Hidden Layer process the sum of the products of weights using activation function and pass on to the output layer.

Readers are encouraged to refer author's Book on Machine Learning [20]. (www.ijsmi.com/book.php)

Data Analysis and results

The data used in the analysis is date wise cases reported by World Health Organization. R software packages such as Forecast, tseries and fGarch for time series models.

Auto Regressive Moving Average (ARIMA)

The following table-1 and chart-1 provides worldwide number of COVID 19 cases forecast is obtained by using Auto ARIMA function (The model assumes the interventions by Governments remain same – Quarantine, Social distancing etc.)

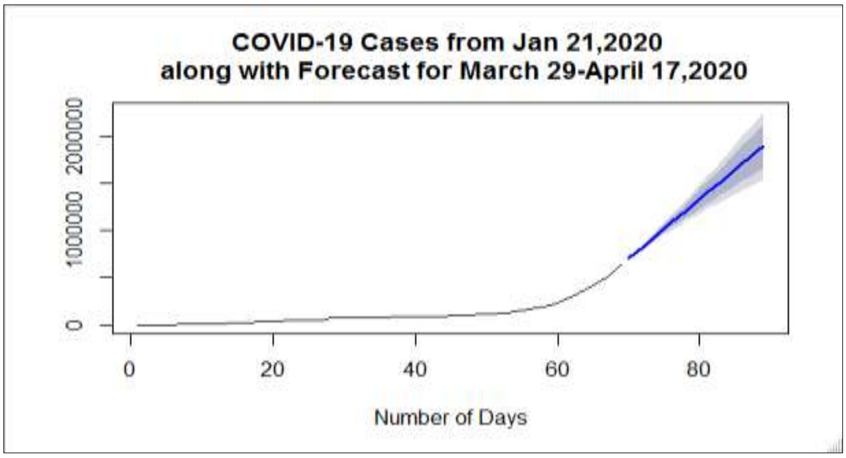
Table1 – Worldwide Forecast values of COVID-19 cases from March 29 to April 17, 2020- ARIMA

Date	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
29-03-2020	697794	692165	703422	689186	706402
30-03-2020	760814	749754	771874	743899	777728
31-03-2020	823815	806013	841616	796590	851039
01-04-2020	886821	861355	912288	847874	925769
02-04-2020	949826	915828	983825	897830	1001823
03-04-2020	1012832	969526	1056138	946602	1079062
04-04-2020	1075837	1022506	1129168	994275	1157400
05-04-2020	1138843	1074817	1202868	1040924	1236761
06-04-2020	1201848	1126498	1277198	1086610	1317086
07-04-2020	1264854	1177582	1352125	1131384	1398323
08-04-2020	1327859	1228098	1427620	1175288	1480430
09-04-2020	1390864	1278070	1503658	1218361	1563368

Vol. 13 No. 1/2020, pp: 1-7

10-04-2020	1453870	1327521	1580219	1260636	1647104
11-04-2020	1516875	1376469	1657282	1302142	1731609
12-04-2020	1579881	1424931	1734830	1342906	1816855
13-04-2020	1642886	1472925	1812848	1382952	1902820
14-04-2020	1705892	1520463	1891320	1422303	1989480
15-04-2020	1768897	1567558	1970235	1460976	2076818
16-04-2020	1831902	1614224	2049581	1498992	2164813
17-04-2020	1894908	1660470	2129345	1536367	2253449

Chart 1 - Worldwide Forecast values of COVID-19 cases from March 29 to April 17, 2020- ARIMA



General Autoregressive Conditional Heteroscedasticity (GARCH) models

The following table-2 provides volatility aspects in the number of cases predicted (sigma values)

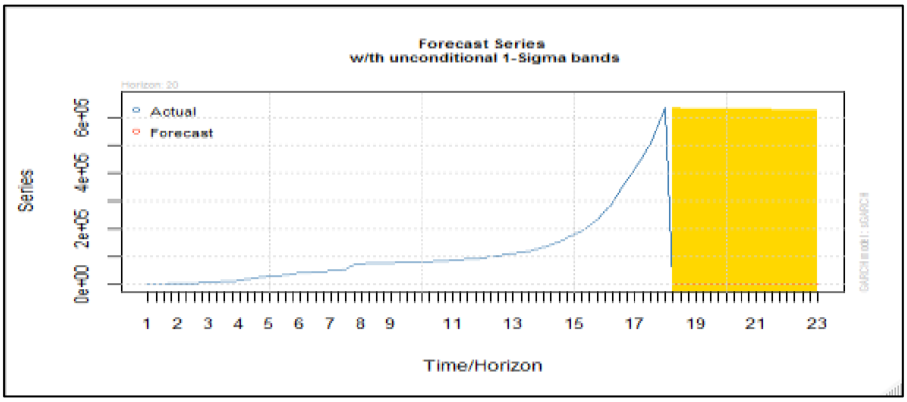
Table2 - Volatility aspects in the number of worldwide cases from March 29 to April 17, 2020- GARCH

Date	Series	Sigma
29-03-2020	79529	555045
30-03-2020	79529	554785
31-03-2020	79529	554525
01-04-2020	79529	554265
02-04-2020	79529	554005
03-04-2020	79529	553746
04-04-2020	79529	553486
05-04-2020	79529	553227
06-04-2020	79529	552968
07-04-2020	79529	552709
08-04-2020	79529	552450
09-04-2020	79529	552191
10-04-2020	79529	551932
11-04-2020	79529	551674
12-04-2020	79529	551416
13-04-2020	79529	551157
14-04-2020	79529	550899
15-04-2020	79529	550641
16-04-2020	79529	550384
17-04-2020	79529	550126

INTERNATIONAL JOURNAL OF STATISTICS AND MEDICAL INFORMATICS

Vol. 13 No. 1/2020, pp: 1-7

Chart2 - Volatility aspects in the number of worldwide cases from March 29 to April 17, 2020- GARCH



Simple Exponential Smoothing Model

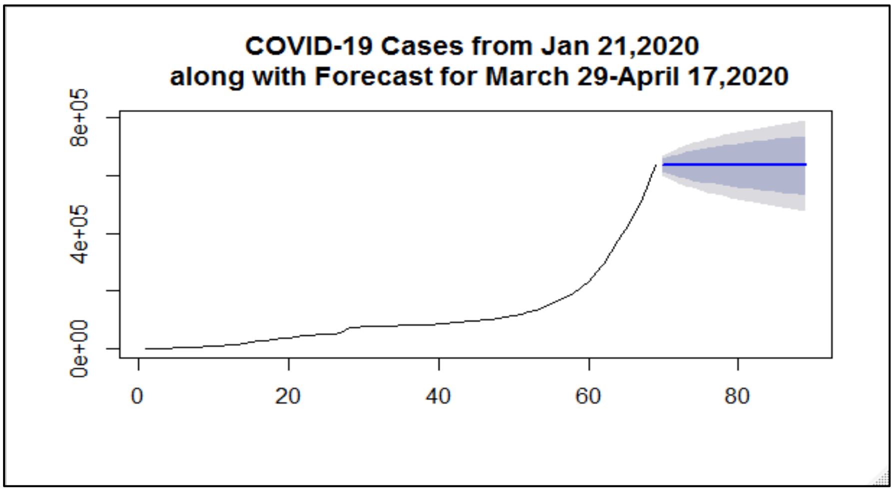
The following table-3 and chart-3 provides worldwide forecast values of COVID-19 cases from March 29 to April 17, 2020 – Simple Exponential Smoothing model. In the below results, simple exponential model with values Hi80 and Hi 80 looks valid.

Table3 – Worldwide forecast values of COVID-19 cases from March 29 to April 17, 2020 – Simple Exponential Smoothing model

Date	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
29-03-2020	634829	611882	657775	599735	669922
30-03-2020	634829	602379	667278	585202	684456
31-03-2020	634829	595087	674570	574049	695608
01-04-2020	634829	588940	680718	564647	705010
02-04-2020	634829	583523	686134	556364	713294
03-04-2020	634829	578627	691031	548875	720782
04-04-2020	634829	574124	695534	541988	727669
05-04-2020	634829	569932	699725	535578	734079
06-04-2020	634829	565996	703662	529558	740099
07-04-2020	634829	562273	707385	523864	745794
08-04-2020	634829	558731	710926	518448	751210
09-04-2020	634829	555348	714310	513273	756385
10-04-2020	634829	552102	717555	508309	761348
11-04-2020	634829	548979	720678	503534	766124
12-04-2020	634829	545966	723691	498925	770732
13-04-2020	634829	543052	726605	494468	775189
14-04-2020	634829	540228	729430	490149	779509
15-04-2020	634829	537485	732173	485954	783703
16-04-2020	634829	534818	734840	481875	787783
17-04-2020	634829	532219	737438	477901	791756

Chart3 – Worldwide forecast values of COVID-19 cases from March 29 to April 17, 2020 – Simple Exponential Smoothing model

Vol. 13 No. 1/2020, pp: 1-7



Artificial Neural Network – Multi Layer Perceptron (MLP)

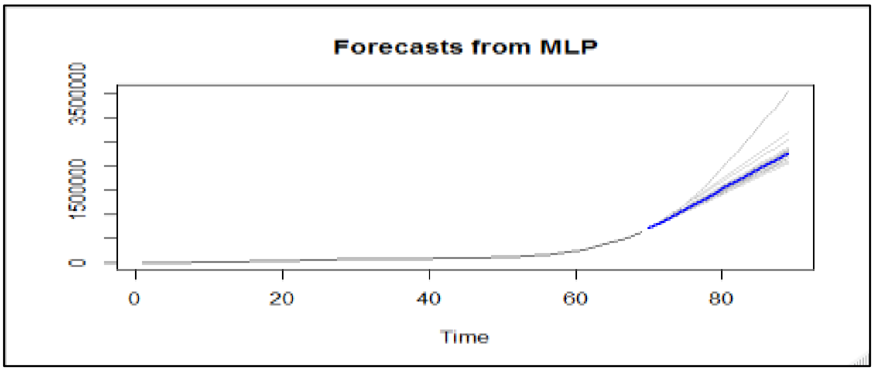
The following table -4 and Chart -4 provides the forecasting values of COVID-19 cases from March 29 to April 17, 2020 using Multi-Layer Perceptron model. nnfor package is used to build MLP model.

Table4 - Worldwide forecast values of COVID-19 cases from March 29 to April 17, 2020 –Multi Layer Perceptron

Point	Forecast
29-03-2020	706751.2
30-03-2020	781259.7
31-03-2020	857511.1
01-04-2020	935885.2
02-04-2020	1015818.6
03-04-2020	1097668.7
04-04-2020	1180314.1
05-04-2020	1263445.8
06-04-2020	1348129.9
07-04-2020	1432466.3
08-04-2020	1516955.3
09-04-2020	1602919.5
10-04-2020	1686632
11-04-2020	1769301.7
12-04-2020	1852112.6
13-04-2020	1935104.5
14-04-2020	2018110.9
15-04-2020	2101034.2
16-04-2020	2184284.8
17-04-2020	2268239.8

Vol. 13 No. 1/2020, pp: 1-7

Chart4 - Worldwide forecast values of COVID-19 cases from March 29 to April 17, 2020 –Multi Layer Perceptron



Conclusion

The paper provided an overview of forecasting COVID-19 cases using time series and machine learning models.

Reference

1. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (last accessed (29th March 2020)
2. Liu, Z., Magal, P., Seydi, O., & Webb, G. (2020). Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data. arXiv preprint arXiv:2002.12298.
3. Ziff, A. L., & Ziff, R. M. (2020). Fractal kinetics of Covid-19 pandemics (with update 3/1/20). MedRxiv preprint.
4. Li, Q., & Feng, W. (2020). Trend and forecasting of the COVID-19 outbreak in China. arXiv preprint arXiv:2002.05866.
5. Hu, Z., Ge, Q., Jin, L., & Xiong, M. (2020). Artificial intelligence forecasting of covid-19 in china. arXiv preprint arXiv:2002.07112.
6. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
7. RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.
8. Harvey, A. C. (1990). Forecasting, structural time series models and the Kalman filter. Cambridge university press.
9. Deb, S., & Majumdar, M. (2020). A time series method to analyze incidence pattern and estimate reproduction number of COVID-19. *arXiv preprint arXiv:2003.10655*.
10. Elmousalami, H. H., & Hassanien, A. E. (2020). Day Level Forecasting for Coronavirus Disease (COVID-19) Spread: Analysis, Modeling and Recommendations. arXiv preprint arXiv:2003.07778.

Vol. 13 No. 1/2020, pp: 1-7

11. Castle, J. L., Doornik, J. A., & Hendry, D. F. (2020). Short-term forecasts of COVID-19-preliminary version 2020-03-19.
12. Bontempi, G., Taieb, S. B., & Le Borgne, Y. A. (2012, July). Machine learning strategies for time series forecasting. In European business intelligence summer school (pp. 62-77). Springer, Berlin, Heidelberg.
13. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. PloS one, 13(3).
14. Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika, 71(3), 599-607.
15. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 987-1007.
16. Ostertagova, E., & Ostertag, O. (2012). Forecasting using simple exponential smoothing method. Acta Electrotechnica et Informatica, 12(3), 62.
17. IJSMI, Editor (2019), Forecasting models - an overview with the help of R software: Time Series Prediction - Past, Present and Future (Machine Learning), ISBN- 9781081552800
18. Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. Econometric Reviews, 29(5-6), 594-621.
19. Kamara, A. F., Chen, E., Liu, Q., & Pan, Z. (2020). Combining contextual neural networks for time series classification. Neurocomputing, 384, 57-66.
20. IJSMI, Editor (2018), Machine Learning: An overview with the help of R software , ISBN-9781790122622.