International Journal of Advances in
Soft Computing and Intelligent Systems (IJASCIS)
2022, Vol 01, No 01, 36-41

Science Transactions © 2022

Original Paper

FORECASTING THE SPREAD OF COVID-19 WITH PROPHET MODEL USING BELGIUM DATASET

Rama Adiga

[1] Nitte (Deemed to be University), Nitte University Centre for Science Education and Research (NUCSER), Division of Bioinformatics and Computational Genomics, Deralakatte, Paneer Campus, Mangalore, India 575018

rama_adiga@nitte.edu.in

ABSTRACT

The Coronavirus pandemic emerged as the deadliest in recent times originated from China in 2019 affecting the entire globe. The data used in this work is obtained from publicly available official Belgian website. Model parameters of ARIMA were selected and the optimized parameters were used to forecast the COVID-19 cases. Daily time series data from April 2020 to July 2021 of total infected cases from Belgium were collected. ARIMA forecast for 60 and 100 days was computed using the seasonal ARIMA model and PROPHET model which resulted in the forecast with actual (observed) data and the predicted data. ARIMA and PROPHET model suggested the continuing trend of infection in Belgium. The predicted ARIMA forecast model (3,0,0) (0,1,0) had MAPE 39.8 with Mean Error 0.005 and RMSE 0.54. The obtained model may be used to assess the successive waves of infection in other countries. The Prophet model outperforms ARIMA, in accuracy in forecasts. The RMSE of the model was 259.7536 and MAE was 208.8522

KEYWORDS

Arima modelling, Prophet, accuracy, forecast, Facebook

1. INTRODUCTION

COVID-19 virus has been spreading as a pandemic across the globe. A respiratory infection spreading through small droplets or upon contact has been causing severe acute respiratory syndrome [1]. The pandemic originated from Wuhan of China with lives lost worldwide with the trend still continually increasing [1,2]. Predicting the developing trend of the successive waves of COVID-19 in five European countries have been studied in France, Spain, UK, Germany, and Italy [3]. Globally the risk of second rebound of the pandemic COVID-19 pandemic caused great anxiety. Looking at the history of the deadly virus there are multiple waves of pandemic causing significant spread and deaths.

In the year 1918 the Spanish flu for example first made its appearance in the USA which was transmitted to Europe during World War I through the returning soldiers in early spring of the

same year. It was deemed to be contagious having all the hallmarks of the seasonal flu. The first wave however, was not particularly deadly with symptoms of malaise and fever lasting for 3 days. Amidst speculation, however, the Spanish flu emerged as a mutated strain which spread through the warring troops to France and USA from England, causing greater fatalities and often termed the 'rebound". Similar was the spread of the H7N9 pandemic which caused five epidemics in China.

1.1 The danger of subsequent waves

The countries resorted to lockdown as the only measure to contain the virus from spreading which included banning international flights to restrict the movement of millions, suspending schools and educational institutes and business operations. The psychological impact on the population with long stay-at- home and the outcome of unlocking by the governments of countries brought the administrators and scientific communities on the discussion table for a valid solution.

Multiple models have been developed by the scientists which present the best-case and worst-case scenarios, under varying sets of circumstances. The randomness and the complexity of virus transmission present with uncertainties in key epidemiological parameters. ARIMA model have been used to predict the European Centre for Disease Prevention and Control (ECDC) data to on the number of infected cases and mortality rate of COVID-19 [5]. However, only a few countries participated in the study. Other studies investigated the rising trends of COVID19 and provided a forecasting model including from India [5].

The objective of the study was to find an appropriate ARIMA model and a PROPHET model for forecasting the spread of the virus. If such a model can be used to predict the spread in other countries.

2. FORMAT GUIDE

2.1. Methods

The COVID-19 dataset is taken from the database hosted by the Belgian institute for health, Sciensano and is responsible for the collaborations on epidemiological data on COVID-19 pandemic and provides insight into the dynamics. The dataset is downloaded from https://epistat.wiv-isp.be/covid/ for further analysis.

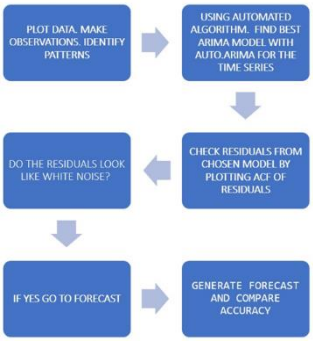


Figure 1: Workflow

As represented in the dataset, the death toll had an exponential phase in 2020 followed by a sharp decline with slight increase until June 21 which coincides with the unlock down seen in the state. In the beginning of March 2020, there was no seasonality in the data, nor was it expected to be relevant. However later on the seasonality was visible. The workflow was indicated in Figure

1.All analysis was carried out in R version 4.1.1 (statistical computing 2021) after installing suitable packages.

3. RESULTS

3.1 ARIMA Model

The graphics function of auto ARIMA (auro_arima) finds appropriate parameters like p, d, q.

The difference tests like Augmented Dickey-Fuller, Phillips–Perron or Kwiatkowski–Phillips–Schmidt–Shin tests are used to decide the parameters like difference, *d* for a test of stationarity.

The models are fitted within the scope of start_*p*, max_*p*, start_*q*, max_*q* ranges . Auto_Arima can also identify the ideal p and q values for tests for seasonality such as the Canova–Hansen test.

3.1.1 Data Preprocessing and transformation

The ADF test was performed to test for stationarity. The P-value was found to be 0.07 which is greater than 0.05, hence the data is accepting the null hypothesis, indicating the data is non-stationary.

adf.test(rnorm(6), k=0)

To obtain optimal performance of ARIMA models the data need to be stationary. Transformation of data takes place when it is converted to stationarity. Differencing is used primarily for data transformation.

The auto.arima gave the initial Arima model as (2,0,1) as the best model.The parameters given by the auto_arima model was iterated and was subjected to differencing. Figure 2 depicted plot of differenced data against time plotted monthly. The seasonality in the data is revealed using auto arima model with the Belgium dataset.

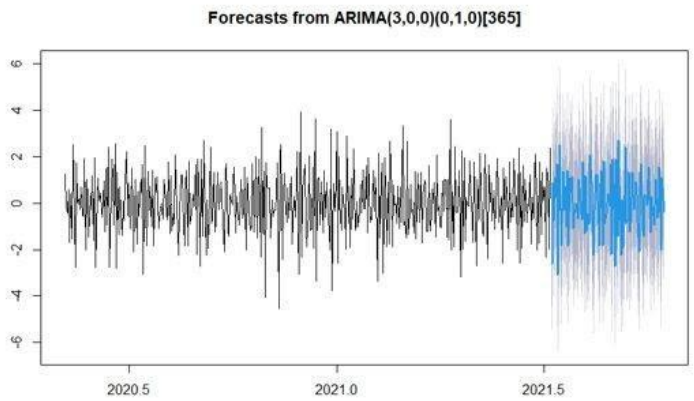


Figure 2. ARIMA forecast model (3,0,0)(0,1,0) for next 100 days

3.1.2Forecasting

A widely popular method for outbreak detection of diseases is ARIMA models. The prerequisite for fitting an ARIMA model is stationarity of time series. To build the ARIMA model the first step was to test for unit root in the time series. Since the time series in the study was non stationary.

This was done by conducting the Augmented Dicky Fuller Test (ADF). The tests suggested that the data had to be transformed to stationarity by differencing. The residuals of the ARIMA series were checked and found to be stationary (Figure 3).

The ACF plot decays slowly indicating the seasonal AR model. To handle the data showing seasonality we fit a seasonal ARIMA model to the study data with auto.arima. Differencing of the first order on the seasonal model set to true and specifying the argument D-1 as one of the parameters in the auto.arima function.

The analysis was carried out using library(forecast) with auto.arima.

fit <- auto.arima(diff(myts), D=1, seasonal = TRUE, approximation = FALSE)

The residual plot and PACF plots from the auto_arima model, was shown in Figure 5 and Figure 6 respectively.

ACF plots show that the residual error is not correlated.Table 1 shows the measures of accuracy

The value 0.54 of RMSE is similar to standard deviation and is a measure of how much the residual distribution is.Around 39.8% MAPE implies the model is about 60.2% accurate in predicting the observation.

The results described in the current study found ARIMA (3,0,0)(0,1,0) to be optimal for forecasting the spread of infection. The forecasting for next 100 days showed ARIMA model exhibits seasonality in prediction. A better measure of estimating accuracy of forecast is MAPE which implies the model accuracy of 60.2%. The forecast indicated the subsequent waves would appear in a seasonal manner for the next 100 days. In spite of restrictions across the border and the vaccinations rate increasing the model forecast indicate a gradual decrease with sudden spurts of infection. The study may be extended to the other European countries with similar control measures such as vaccination by the government. Though vaccine coverage has been directly linked in controlling, its effectiveness has also been debated in pandemic control [6-11].

Future work would aim to correlate other parameters such as vaccination rate and its effectiveness in eradication of coronavirus in populations. Other spatial-temporal models such as Bayesian model or spatial fixed/random effects panel models may be investigated.

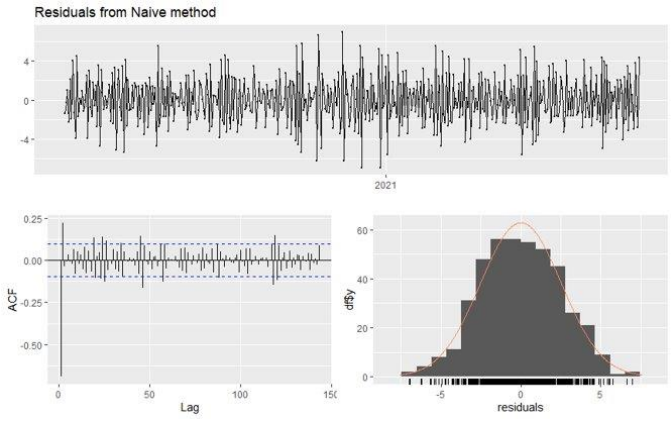


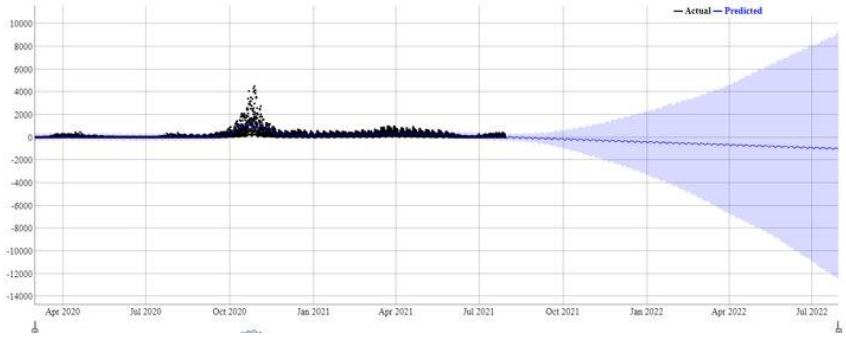
Figure 3. Residual Checks from Naïve method and ACF plot

Table 1. Accuracy check for ARIMA model

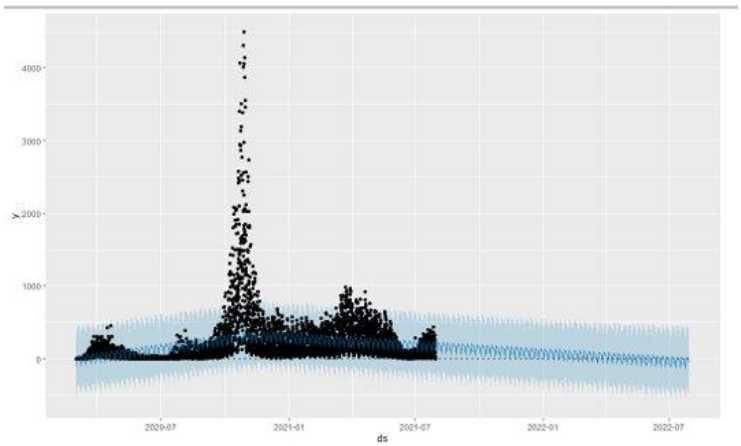
Measures of accuracy	Value
Mean Absolute Error (MAE)	0.172
Mean Error(ME)	0.005
Root Mean Square Error (RMSE)	0.54
Mean Absolute percentage Error (MAPE)	39.8
AIC	230.7
AICc	231.4
Log likelihood	-111.3
ACF	-0.04

3.2 PROPHET MODEL

Prophet is a model developed by open-source tool of Facebook Inc., which was modelled in R. A very flexible tool for forecasting and is very intuitive.



Prophet model



Prophet model: logistic growth model with a saturating minimum

Table 2: Evaluation Metric for the Prophet model

Sl No.	horizon	mse	rmse	mae	smape	coverage
1	6 days	53301.78	230.8718	184.4525	0.8310489	0.9308449
2	7 days	53627.78	231.5767	183.9293	0.8396263	0.9259259
3	8 days	56040.72	236.7292	189.9405	0.8755413	0.9244792
4	9 days	58656.67	242.1914	193.6253	0.888002	0.9198495
5	10 days	63341.08	251.6765	201.7680	0.9225915	0.9105903
6	11 days	67471.92	259.7536	208.8522	0.9212559	0.9024884

3. CONCLUSIONS

Papers In the present paper the dataset from Begium health web site were used to build an Arima forecast model. The residual checks were done for fit of the model. The Prophet model was also constructed and it scored over Arima model in accuracy.

ACKNOWLEDGEMENTS

The author acknowledges the contribution of Gagan Punacha for assisting in manuscript preparation. I also thank Prof. Anirban Chakraborty, Director, NUCSER and the management for the facilities provided in the institute.

REFERENCES

[1] Rambaut, A, Holmes, E.C, O'Toole Á, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus OG et al.(2020) A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. vol 5, pp1403–1407

[2] Zhu, N., D Zhang, W Wang, X Li, B Yang et al.(2019). A novel coronavirus from patients with pneumonia in China. N. Engl. J. Med. vol 382, pp.727–733.

[3] Faranda D, Alberti T(2020). Modeling the second wave of COVID-19 infections in France and Italy via a stochastic SEIR model. Chaos. Vol: 30, pp 111101. doi: 10.1063/5.0015943.

[4] Bayyurt Lutfi, Bayyurt Burcu (2020) Forecasting of COVID-19 Cases and Deaths Using ARIMA Models, medRxiv .04.17.20069237. 10.1101/2020.04.17.20069237.

[5] Tandon,H., Ranjan, P., Chakraborty,T., Suhag, V(2020) Coronavirus (covid-19): Arima based time-series analysis to forecast near future . arXiv:2004.07859 (2020)

[6] Coletti, P., Libin, P., Petrof, O. et al. A data-driven metapopulation model for the Belgian COVID-19 epidemic: assessing the impact of lockdown and exit strategies. BMC Infect Dis 21, 503 (2021). https://doi.org/10.1186/s12879-021-06092-w

[7] McKinsey & Company, 2020. COVID-19: Implications for business. Available at: https://www.mckinsey.com/business-functions/risk/our-insights/covid-19-implications-for-business.

[8] Yuan J., Li M., Lv G. and Lu K., 2020. Monitoring transmissibility and mortality of COVID-19 in Europe. International Journal of Infectious Diseases, 95, pp. 311–315. pmid:32234343

[9] Wolff, S., and S. Ladi. 2020. "European Union Responses to the Pandemic: Adaptability in Times of Permanent Emergency." Journal of European Integration 42 (8).

[10] Manoj K, Madhu A. An application of time series arima forecasting model for predicting sugarcane production in India[J]. Stud ITI Bus Econ. 2014;9(1):81–94.

[11] Sato, R.C. Disease management with ARIMA model in time series. Einstein 2013 11(1):128–31.