Prophet Forecasting and Temporal Modeling of Covid-19 Cases in the Philippines

Mary Colleen Bautista[1], Adrian Joshua Nunez [2]and Shirlee Ocampo[3]
Mathematics and Statistics Department, De La Salle University
[1]*email: mary_colleen_bautista@dlsu.edu.ph*
[2]*email: adrian_joshua_nunez@dlsu.edu.ph*
[3]*email: shirlee.ocampo@dlsu.edu.ph*

Abstract: This paper aims to make use of machine learning models to forecast the trend of COVID-19 in the Philippines to approximate when surges might occur and for the country to be better prepared for the next wave. This study uses three forecasting techniques, Simple Exponential Smoothing (SES), Auto Regressive Integrated Moving Average (ARIMA) and Prophet Forecasting Model. The Prophet Method is a forecasting method developed in 2017, it is new compared to the other forecasting method used. The model performances were compared with the use of Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and with the addition of Akaike's Information Criterion (AIC) for ARIMA model selection. For the study, the dataset used was obtained from Our World In Data (OWID) website and contained the number of daily confirmed cases in the Philippines. The training set was used to forecast future values for the Two methods, the forecasted result was then compared to their test data to measure model performance against each other. Results showed that Prophet outperformed ARIMA with it having the lowest RMSE and MAPE.

Key Words: ARIMA, Prophet, COVID-19, Forecasting Techniques, SES

1. INTRODUCTION

The COVID-19 pandemic has turned the world upside down. According to Maron (2020), it is believed to originate from a wet seafood market in Wuhan, China. However, it has now spread and infected around 215 countries all over the world. The WHO (n.d.) stated that it is caused by a virus known as SARS-CoV-2 which was spread rapidly by human-to-human interaction. More than 167 million people were affected and more than 3.46 million have died because of the virus as reported by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (n.d). COVID-19 transmission has been declared by the World Health Organization (WHO) as a public health emergency as cases continue to rise globally. As of current, the Philippines is also facing this disease and has been struggling with the prevention of the sudden surges in cases. To counteract the surge in cases, the Government continues to implement quarantine guidelines. But according to Amnesty International (2021) these guidelines have not effectively prevented the transmission of the disease as we have seen a second surge in cases. From seeing 2000 to 3000 daily reported cases from the last quarter of 2020 growing to form a second wave in mid-March that reached over 15,000 new cases in April. According to the Department of Health (DOH) COVID-19 Tracker (n.d.) it showed that during the 2nd wave, daily positivity rate reached 25.2 percent dated April 2,

2021. This was the highest daily positivity rate recorded since the pandemic started back on January 30, 2020. More than a year has passed and yet still the daily positivity rate showed much higher results in the 2nd wave as compared to the 1st wave. Everyone is working persistently to find a solution to pacify the virus. In agreement with the WHO (n.d.), strategic planning with predictive modeling and forecasting to analyze COVID-19 data may allow us to fight this virus.

COVID-19 is a highly infectious disease caused by the Coronavirus, cases continue to increase as days pass and severely affects a significant portion of the world's population. During these sudden surges in cases, hospitals in densely populated areas have been finding it difficult to keep up and have been reaching "critical" levels in COVID-19 bed occupancy (Gulla, 2021). As a response to the problems, the research used three forecasting techniques: Simple Exponential Smoothing (SES), Auto Regressive Integrated Moving Average (ARIMA) and Prophet Method. Compared to the other techniques, described as a model with interpretable parameters that can easily be modified intuitively, the Prophet Method is a new forecasting method developed in 2017 by Sean J. Taylor and Ben Letham, both of which are a part of Facebook's research team. The study aims to forecast COVID-19 cases in the Philippines using the Prophet, ARIMA and SES. The performances of these forecasting methods were compared using Root Mean Square Error (RSME) and Mean Absolute Percentage Error (MAPE). Through this research, the most robust model generated is useful for prediction of future cases to assist the government and hospitals in preparing for the possible next surge in COVID-19 cases.

2. METHODOLOGY

2.1 Splitting Data into Training and Test Sets

The COVID-19 data set retrieved from Our World in Data which had a total of 488 time points. Since the division of the data is 80% for the training set and 20% for the test set, we had a total of 390 time points for the training set and 98 time points for the test set for daily cases the time series plot for training set and test set (red plot) can be seen in Figure 1.

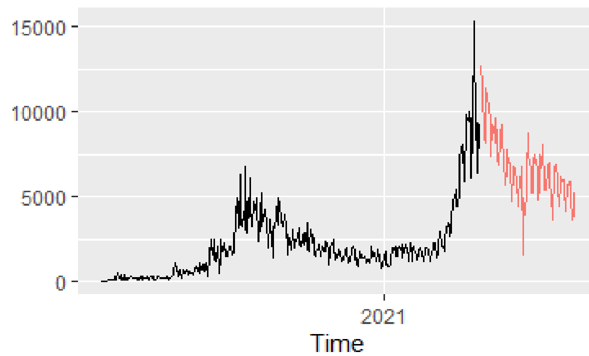

Fig. 1. Time Series Plot of Daily Cases for the Training Set (black plot) and Test Set (red plot)

2.2 Simple Exponential Smoothing (SES)

Simple Exponential Smoothing (SES) according to Ismail et al. (2020) is a forecasting method for data that has no trend nor seasonality. The method requires one parameter, and this parameter is alpha (α) which is also known as the smoothing factor and ranges from 0 to 1 (Singh, 2021). From the study of Chaurasia and Pal (2020), the formula for this method is

$$Y_t = Y_{t-1} + \alpha(A_{t-1} - Y_{t-1}) \quad \text{(Eq.1)}$$

where

Y_t= predicted value for next period at time t

α = smoothing constant (0< α<1)

A_{t-1} = actual value at time t-1

2.3 ARIMA

According to Maurya and Signh (2020), the combination of two models makes up ARIMA. These two models are Autoregressive (AR) and Moving Average (MA) and they are combined through Integration "I". A paper by Satrio et al. (2021) mentioned that ARIMA models have three parameters p, d and q where p represents "AR", d represents "I" and q represents "MA". Malki et al. (2020) and Satrio et al. (2021) illustrates the models AR, MA and ARIMA model (d = 1) respectively as,

$$AR(p): y_t = \mu + \sum_{i=1}^{p} (\phi_i y_{t-i}) + \omega_t \quad \text{(Eq. 2)}$$

where

p = number of autoregressive lags
y_t = predicted value at time t

μ = mean value of the time series data
ϕ_i = AR(p) coefficients (i = 1,2, …, p)

ω_t = white noise

$$MA(q): y_t = \mu + \sum_{i=1}^{q} (\theta_j \omega_{t-j}) + \omega_t$$
(Eq. 3)
where:

q = number of moving average model lags
y_t = predicted value at time t

μ = mean value of the time series data
θ_j = MA(q) coefficients (j = 0, 1, 2, …, q)

ω_t = white noise

ARIMA in eq. 4 is the combination of the AR and MA model above with the number of differences equal to 1 (d = 1).

$$ARIMA(p, d, q): y'_t = \mu + \sum_{i=1}^{p} (\phi_i y'_{t-i}) + \sum_{i=1}^{q} (\theta_j \omega_{t-j}) + \omega_t \quad \text{(Eq. 4)}$$

2.4 Prophet Forecasting

In 2017, Sean J. Taylor and Ben Letham wanted to address the problem in producing reliable and high-quality forecasts due to the low count of expert analysts in time series modeling. Therefore, to address these problems they proposed a regression model with easy to interpret parameters that can be modified intuitively, a method that can help analysts forecast at scale and effectively use their expertise in time series forecasting, they called this model the Prophet Forecasting Model (Taylor & Letham, 2017). The model handles outliers well and is robust when it comes to shifts in the trend and missing data (Žunić et al., 2020). It implements an additive time series forecasting model which supports the following model components: holidays, seasonality (yearly, weekly, and daily), and non-linear trends. Using a decomposable time series model with the aforementioned components, is computed by,

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad \text{(Eq. 5)}$$
where:

y_t = predicted value at time t

$g(t)$ = logistic growth curve for modelling non-periodic changes in time series (trend)
$s(t)$ = periodic changes (seasonality)

$h(t)$ = effects if holidays to the forecast

ε_t = error term

2.5 Statistical Measures for Comparison of Model Performance

To choose the most robust model, the data was split into two to make a training set (from March 15, 2020, to April 8, 2021) and a test set (from April 9, 2021 to July 15, 2021) which was used to validate each of the model performances. Forecasted values obtained from the 5 methods from the training set were used to compare to the values of the test set. To evaluate each of the model performances, the researchers used the metric Root Mean Square Error (RMSE) since its useful whenever large errors are particularly undesirable, Mean Absolute Percentage Error (MAPE) since its a widely used measure when checking for forecast accuracy, and Akaike's Information Criterion (AIC) in order for different possible models to be determined which is the best fit. The method with the lowest value for RMSE, MAPE and AIC is the most robust model of them all.

3. RESULTS AND DISCUSSION

3.1 Simple Exponential Smoothing (SES)

The basic idea of the model is to assume that the future forecast will be more or less similar to the past data. After using the SES function in R, the researchers obtained the results for daily cases RMSE = 766.8478 MAPE = 30.10847. They then applied the same model for the test set, summary results showed for daily cases RMSE = 2817.379 MAPE = 46.84422. The time series plot for the training set of daily cases along with the forecasted values can be seen in Figure 2.

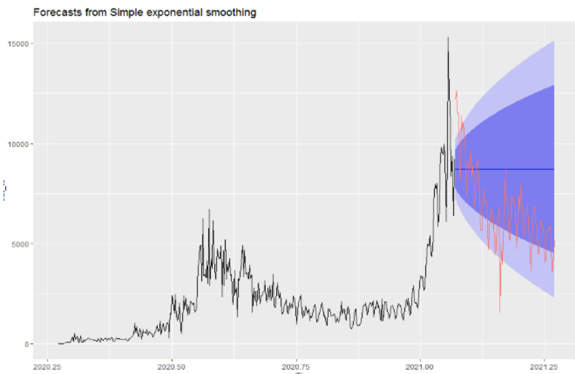

Fig. 2. Time Series Plot for The Training Set and Test Set with The Forecasted Values for Daily Cases from SES

3.2 ARIMA

3.2.1 Exploratory Data Analysis

As can be observed in the time series plot for the daily cases in Figure 1, there is an increasing trend. An increasing trend suggests that there is a non-stationary time series. Therefore, it is necessary to apply logarithmic transformation or differencing to make the time series stationary. Applying logarithmic transformation, the data was still non-stationary from the result of the Augmented Dickey-Fuller (ADF) test for stationarity. Upon first differencing the data, ADF test showed that the data was now stationary. The first differenced time series can be seen in Figure 3.

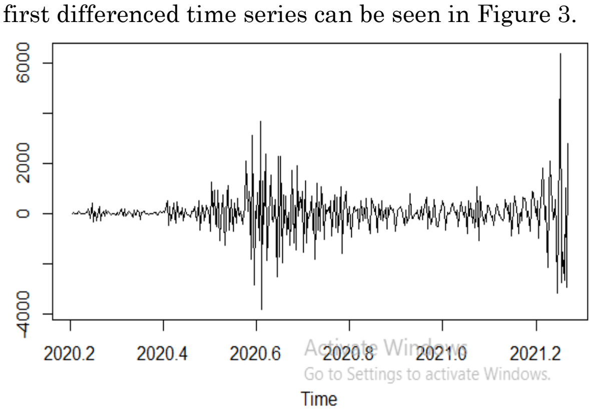

Fig. 3. First Differenced Time Series Plot for Daily Cases

3.2.2 Arima Model

Building the ARIMA model in R, upon mix and matching the significant lags it was found that the best model for daily cases is ARIMA(3,1,3). Summarizing ARIMA(3,1,3), results showed RMSE = 715.6495, MAPE = 30.21328, and AIC = 6234.11. Applying the same model for the test set, summary of results showed RMSE = 1044.918 and MAPE = 14.64506. The time series plot for the training set and test set along with the forecasted values can be seen in Figure 4.

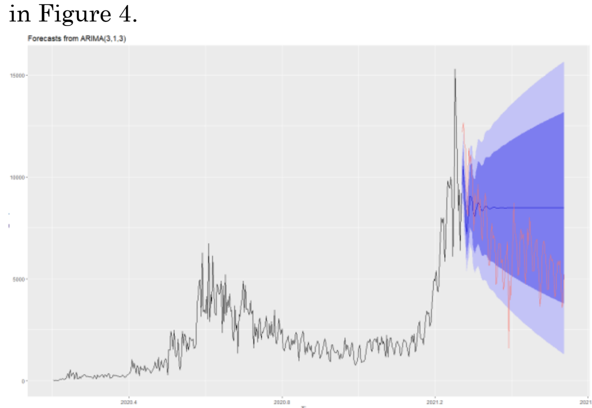

Fig. 4. Time Series Plot for The Training Set and Test Set with The Forecasted Values for Daily Cases from ARIMA

3.3 Prophet Forecasting

Using Prophet Forecasting in R, with yearly and weekly seasonality considered. Checking for accuracy of the results for daily cases, RMSE = 663.874 and MAPE = 90.55279 for the training set and RMSE = 447.3837 and MAPE = 4.200445 for the test set. The daily cases forecasted plot with yearly and weekly seasonality can be seen in Figure 5.

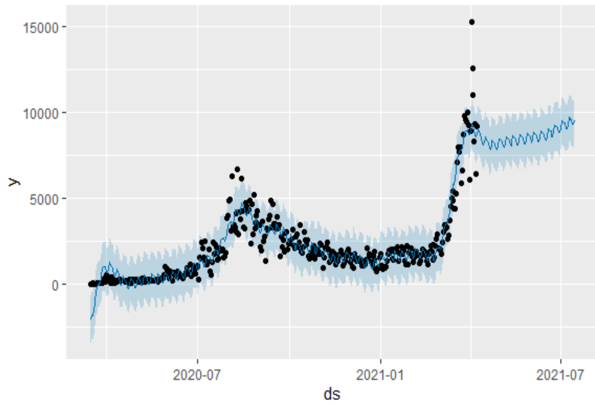

Fig. 5. Plot for the Training Set and Test Set With The Forecasted Values for Daily Cases from Prophet

Using Prophet plot components, it can be observed in Figure 6 that there is a steady upward trend throughout. In the weekly graph there is a spike in cases on Friday, Saturday, and Sunday. It can be assumed that the reason for this is because these are the non-working days where people have the time to go out and get tested. Yearly graph shows that there is a clear spike every middle of the year.

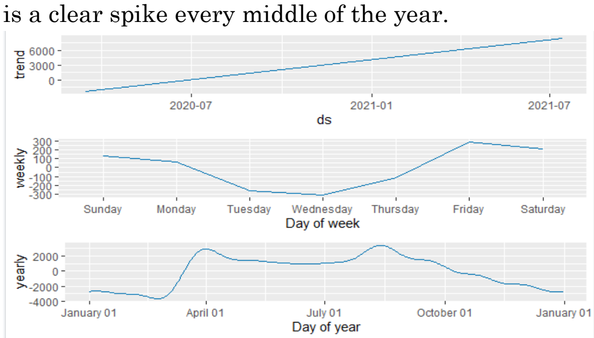

Fig. 6. Prophet Plot Components for Daily Cases

Out of the three methods, the Prophet forecasting model performed better for daily cases, giving the lowest error measure. It can be implied that the reason for Prophet performing best is because the method has easy to interpret parameters, and these parameters can be modified to improve the model fit. When we plot the model fit over the historical data, it can easily be seen if change points were missed by the automatic changepoint selection under trend model, and from there the parameters can easily be modified to get a better fit. From the results for daily cases using the Prophet Method, modifying the seasonality component for daily cases showed a better model fit, this is because the historical data or actual data showed signs of seasonality.

4. CONCLUSIONS

In this study, three forecasting techniques were compared, Simple Exponential Smoothing (SES), Auto Regressive Integrated Moving Average (ARIMA) and Prophet Method. In this research the Prophet Method was used to compare to ARIMA and SES as it was a new forecasting method developed by Sean J. Taylor and Ben Letham. Based on the results of the study, the Prophet Method performed gave a significant performance difference compared to SES and ARIMA. It is better for \predicting future values for daily cases with the method having the lowest Root Mean Square Error and Mean Absolute Percentage Error

From the results, an upward trend can be observed for daily cases. Therefore, it can be advised to the Government and hospitals to prepare, as the number of cases will only keep growing as the time goes. However, there may be different factors that can affect the number of cases, examples can be the number of individuals vaccinated, fully vaccinated or the future variants of the virus. Therefore, it is recommended to further study the spread of the virus and to consider the aforementioned examples in the study.

5. REFERENCES

Amnesty International. (2021, April 26). Health and human rights crisis in the Philippines, one year into pandemic. Amnesty International. https://www.amnesty.org/en/latest/news/2021/04/philippines-faces-health-human-rights-crisis-covid/.

Chaurasia, V., & Pal, S. (2020, October 24). Application of machine learning time series analysis for prediction COVID-19 pandemic. Research on Biomedical Engineering.

https://link.springer.com/article/10.1007/s42600-020-00105-4#Tab1.

"COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University" (n.d.)retrieved from https://github.com/CSSEGISandData/COVID-19

Gulla, V. (2021, March 17). Occupancy rate in at least 20 private Ncr hospitals 'critical.' *Occupancy Rate in at Least 20 Private NCR Hospitals 'Critical' - Hospital Group.* https://news.abs-cbn.com/news/03/17/21/occupancy-rate-in-at-least-20-private-ncr-hospitals-critical-hospital-group.

Ismail, L., Materwala, H., Znati, T., Turaev, S., & Khan, M. A. B. (2020, September 24). Tailoring time series models for forecasting coronavirus spread: Case studies of 187 countries. Computational and Structural Biotechnology Journal. https://www.sciencedirect.com/science/article/pii/S2001037020303998.

Malki, Z., Atlam, E.-S., Ewis, A., Dagnew, G., Alzighaibi, A. R., ELmarhomy, G., Elhosseini, M. A., Hassanien, A. E., & Gad, I. (2020, October 23). ARIMA models for predicting the end of COVID-19 pandemic and the risk of second rebound. Neural Computing and Applications. https://link.springer.com/article/10.1007/s00521-020-05434-0.

Maron, D., (2020)Corona virus linked to chinese wet markets. retrieved from https://www.nationalgeographic.com/animals/article/coronavirus-linked-to-chinese-wet-markets

COVID-19 Tracker: Department of Health website. COVID-19 Tracker | Department of Health website. (n.d.). https://doh.gov.ph/covid19tracker.

Maurya, S., & Singh, S. (2020). Time Series Analysis of the Covid-19 Datasets. 2020 IEEE International Conference for Innovation in Technology (INOCON). https://doi.org/10.1109/inocon50539.2020.9298390

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D., & Roser, M. (2020, March 5). Coronavirus (COVID-19) cases. Our World in Data. Retrieved March 4, 2022, from https://ourworldindata.org/covid-cases

Satrio, C. B. A., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021, February 19). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. Procedia Computer Science. https://www.sciencedirect.com/science/article/pii/S1877050921000417.

Taylor, S. J., & Letham, B. (2017, September 27). Forecasting at scale. PeerJ Preprints. Retrieved December 15, 2021, from https://peerj.com/preprints/3190/

Word Health Organization (n.d.) retrieved from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

Žunić, E., Korjenić, K., Hodžić, K., & Đonko, D. (2020). Application of facebook's prophet algorithm for successful sales forecasting based on real-world data. International Journal of Computer Science and Information Technology, 12(2), 23–36. https://doi.org/10.5121/ijcsit.2020.12203