ICIC Express Letters
Part B: Applications
Volume **13**, Number **2**, February **2022**

ICIC International ©2022 ISSN 2185-2766
pp. **211–218**

COVID-19 FORECASTING IN INDONESIA USING PROPHET MODEL

Aedentrisa Yasmanda Paulindino, Elvan Selvano, Maryanto Pricillia Katarina and Widodo Budiharto*

Computer Science Department
School of Computer Science
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
{ aedentrisa.paulindino; elvan.selvano; maryanto001; pricillia.katarina }@binus.ac.id
*Corresponding author: wbudiharto@binus.edu

Received March 2021; accepted June 2021

Abstract. *The first COVID-19 case in Indonesia occurred on March 2, 2020. By April 9, 2020, the pandemic had spread to all 34 provinces in the country. As of February 27, 2021, Indonesia has reported 1,329,074 cases, the highest in Southeast Asia. With 35,981 deaths, Indonesia ranks third in Asia and 17th in the world. The purpose of this paper is to conduct real-time forecasting of the COVID-19 pandemic that occurred in Indonesia based on publicly available epidemiological data. In this research, we have forecasted the number of cumulative COVID-19 cases for the next 30 days using the Prophet model. The model's testing accuracy values were 0.9746 for the R^2 score and 17553.8462 for the Root Mean Squared Error (RMSE). We hope that this forecasting may facilitate public institutions about future cases.*
Keywords: COVID-19, Indonesia, Forecasting, Prophet

1. **Introduction.** The current outbreak of the novel coronavirus COVID-19, which started in Wuhan, Hubei Province, China, in early December 2019, has spread to many other countries. On January 30, 2020, the WHO Emergency Committee declared a global health emergency based on growing case notification rates at Chinese and international locations [1]. On February 28, 2021, there were 113,745,002 confirmed cases and 2,524,133 deaths [2].

COVID-19 cases started to appear in Indonesia on March 2, 2020. Other countries worldwide introduced lockdown measures during this time, but Indonesia offered incentives to promote tourism instead. This late response from the government has made Indonesian vulnerable to the pandemic. On March 13, 2020, the government set up a Task Force for Rapid Response to COVID-19, two days after the first confirmed death [3].

On February 28, 2021, there were a total of 1,329,074 confirmed cases with 157,039 active cases, 1,136,054 recovered cases, and 35,981 deaths [4]. The Indonesian government has implemented various methods to break the chain of the spread of COVID-19. One of them is Large-Scale Social Restrictions (LSSR), limiting individual residents' activities in an area suspected of being infected with the coronavirus to prevent the possibility of spreading it more widely. For residents affected by COVID-19 with severe symptoms, quarantine is carried out at a COVID-19 referral hospital. In contrast, residents who have interacted or are positive for COVID-19 without severe symptoms are advised to self-quarantine for 14 days [5].

Recognizing the rate of spreading the virus is vital in the battle against this pandemic. Monitoring the degree of spreading pace can help national authorities and government

DOI: 10.24507/icicelb.13.02.211

officials in policymaking to address this pandemic [6]. In this research, we explain Prophet and its methodology. Then, we develop a machine learning model using Prophet to predict the cumulative COVID-19 confirmed cases in Indonesia. In the end, we evaluate its predictions using several scoring metrics.

2. **Materials and Methods.** Prophet is a forecasting library developed by Facebook and is implemented in R and Python [8]. Prophet forecast time series data based on simple linear equations, which fit the non-linear trends by adding the daily, weekly, and yearly seasonality by considering holiday effects [9]. It is usually used for uncertain options, trend options, holiday options, seasonal options, and added regression/model diagnostics. Prophet's limitation is that Prophet does not allow non-Gaussian noise distribution (currently). Prophet did not take account of the autocorrelation of the residuals, and it did not consider stochastic trends.

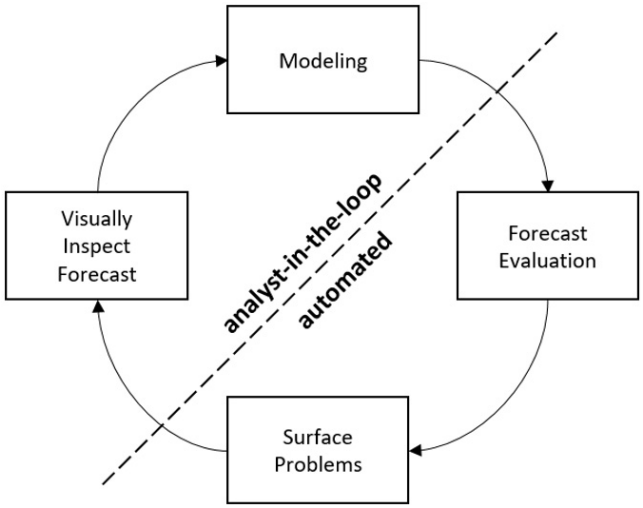

FIGURE 1. Analyst-in-the-loop approach to forecasting at scale [7]

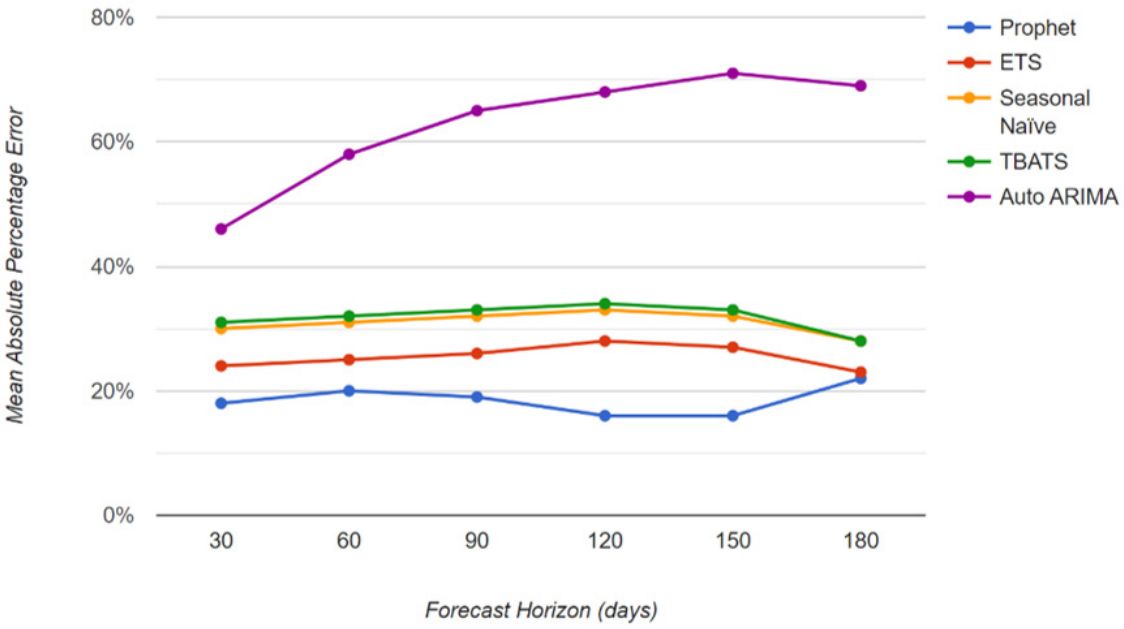

FIGURE 2. Comparison of mean absolute percentage errors for the forecasting methods and time series of Prophet and other forecasts methods [7]

Compared to other forecast horizons such as Auto ARIMA, ETS, snaive, and TBATS, Prophet has a lower prediction error. Prophet forecasts were made with default settings, and tweaking the hyperparameters could further improve performance [7]. The methodology we use for forecasting using Prophet is shown in Figure 3.

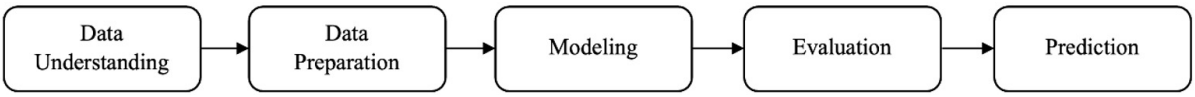

FIGURE 3. Prediction process diagram

3. **Forecasting Using Prophet.** Prophet has two columns, "ds" and "y", to store data time series and corresponding values of the time series. Prophet uses a decomposable time series model with three main components: trend, seasonality, and holidays, which are combined in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{1}$$

In the equation, $g(t)$ is the trend, $s(t)$ represents seasonal changes and $h(t)$ captures irregular effects [10]. The error term ε_t represents any abnormal changes accommodated by the model are considered [11].

Trend modeling can be determined using the logistic growth model, which is represented in the following equation [11]:

$$g(t) = \frac{C}{1 + \exp(-k(t-m))'} \tag{2}$$

where C indicates the growing capacity, k specifies the growth rate, and m is an offset parameter [11]. The piecewise logistic growth model is then [12]

$$g(t) = \frac{C(t)}{1 + \exp\left(-\left(k + \alpha(t)^T\delta\right)\left(t - \left(m + \alpha(t)^T\gamma\right)\right)\right)} \tag{3}$$

where δ and γ could be a vector rate adjustment that defines the modification inside the rate that happens at the time sj, for $j = 1, 2, \ldots, S$ where S is the number of change points. The change points because of a development, which ends up within the rate of growth can be modified, and so the trend model is [13]

$$g(t) = \left(k + \alpha(t)^T\delta\right) + \left(m + \alpha(t)^T\gamma\right) \tag{4}$$

where k is the rate of growth, m is an offset parameter, δ is the rate adjustment, and γj is set to $-sj\delta j$ to create the function continuously. In automatic change points choice, $\delta j \sim Laplace(0, \tau)$. To fit the projected model with seasonality effects and forecast supported it, it uses a Fourier series that provides a versatile model. Seasonal effects may be portrayed as the following equation [13]:

$$s(t) = \sum_{n=1}^{N} \left(\alpha_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)\right) \tag{5}$$

where P could be a regular amount.

4. **Result and Discussion.** Based on the method described, we conducted research on forecasting using a Prophet related to the COVID-19 pandemic that is happening in Indonesia. The following are the steps that were followed by using the methodology in Figure 3.

4.1. **Data understanding.** The daily COVID-19 prevalence data we use are from the official repository of Johns Hopkins University [14]. Based on these sources, data on COVID-19 cases in Indonesia were recorded from January 22, 2020, to February 23, 2021. The data consists of the date the cases were reported and the number of cases that occurred on that date. However, in our study, we deleted several data with a few cases greater than 4. It is intended that the exponential thread of the plot and model can be seen more clearly.

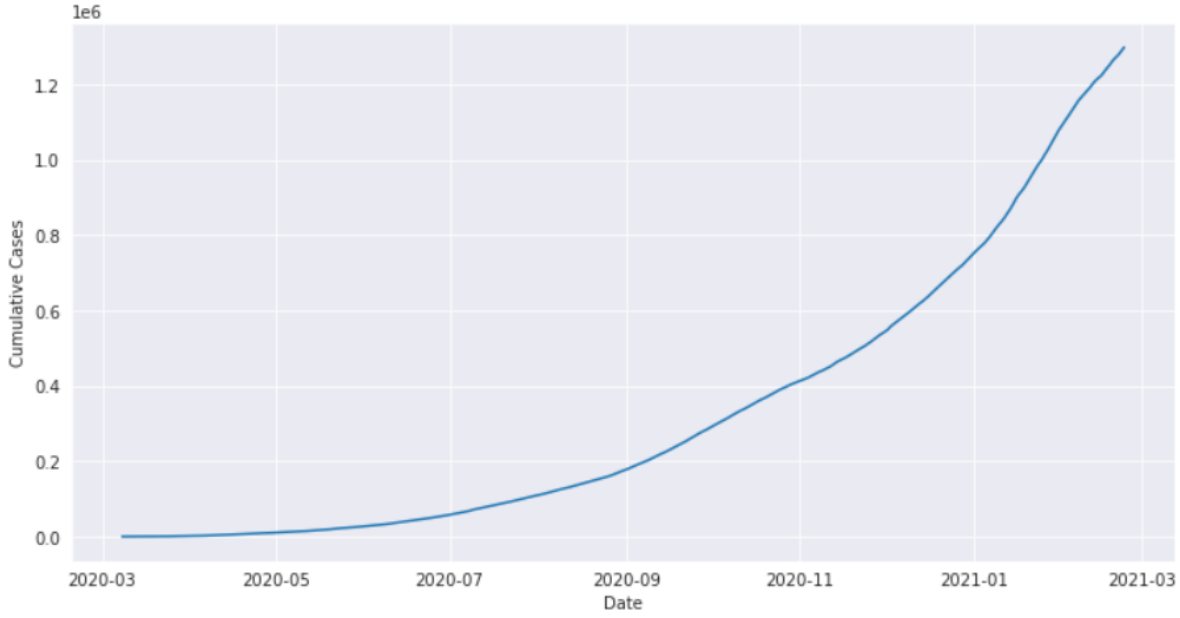

FIGURE 4. Data on the spread of the COVID-19 virus in Indonesia

4.2. **Data preparation.** In the data we use, some factors influence the predictive case. This factor is the day of the prediction symbolized by x and the target case symbolized by y. Before checking the accuracy of the data, the main data will be divided into two parts, which are shown in the table below.

TABLE 1. Data used on Prophet model

Data	Date	Percentage
Training set	March 8, 2020-January 18, 2021	90%
Testing set	January 19, 2021-February 23, 2021	10%

4.3. **Modeling.** Prophet model was built with the help of Scikit-learn. Making this model begins with the creation of its Prophet class instance followed by the fit and predict methods. We will customize the Prophet model by adding a new object Prophet instance. The objects we use are holidays in Indonesia. After the adjustments are complete, all configurations for the forecasting process are transferred to the designer. Then with the fit method, train data will be transferred into a historical data framework to be trained.

To predict future data, we define a target day that we want to predict, where the number of days we want to predict must be in accordance with the test data. In our study, we used *Prophet.make_future_dataframe* to predict 36 days and compared it to test data. The results of these comparisons will produce data accuracy. The amount of accuracy is determined by checking the difference between the two data, and the result of this difference is called the total error.

This total error will be the accuracy value of the predictions that have been made. In the prediction method, each line of future predictive value (margin error) will be assigned, which is called $\hat{y}$. The object of the forecast in the new dataframe contains $\hat{y}$ with the forecast value, $\hat{y}$ with the value of components, and $\hat{y}$ with the value uncertainty intervals.

Table 2. Predicted cumulative cases in the near future ∼ February 23, 2021

ds	$\hat{y}$	$\hat{y}$_lower	$\hat{y}$_upper
2021-02-19	1.228926e+06	1.187464e+06	1.270509e+06
2021-02-20	1.239243e+06	1.195850e+06	1.283743e+06
2021-02-21	1.249097e+06	1.203091e+06	1.294829e+06
2021-02-22	1.258564e+06	1.210752e+06	1.307795e+06
2021-02-23	1.268200e+06	1.219034e+06	1.319443e+06

4.4. **Evaluation.** The table below shows the detailed prediction using $\hat{y}$_upper on each data in testing set.

Table 3. Predicted cumulative cases from January 19, 2021 to February 23, 2021

Date	Cumulative cases
2021-01-19	9.228300e+05
2021-01-20	9.329890e+05
2021-01-21	9.434134e+05
2021-01-22	9.537641e+05
2021-01-23	9.641855e+05
2021-01-24	9.749754e+05
2021-01-25	9.850267e+05
2021-01-26	9.961417e+05
2021-01-27	1.007749e+06
2021-01-28	1.019335e+06
2021-01-29	1.030392e+06
2021-01-30	1.041789e+06
2021-01-31	1.052685e+06
2021-02-01	1.064028e+06
2021-02-02	1.074718e+06
2021-02-03	1.085707e+06
2021-02-04	1.096639e+06
2021-02-05	1.108949e+06
2021-02-06	1.120921e+06
2021-02-07	1.132963e+06
2021-02-08	1.143837e+06
2021-02-09	1.154681e+06
2021-02-10	1.164920e+06
2021-02-11	1.176825e+06
2021-02-12	1.186960e+06
2021-02-13	1.199564e+06
2021-02-14	1.210953e+06
2021-02-15	1.222711e+06
2021-02-16	1.234545e+06
2021-02-17	1.245537e+06
2021-02-18	1.257987e+06
2021-02-19	1.270504e+06
2021-02-20	1.282459e+06
2021-02-21	1.294163e+06
2021-02-22	1.306569e+06
2021-02-23	1.318692e+06

To see how accurate the model is, we do some calculations with several functions that are available in Scikit-learn library [15]. Following are the metrics and results of the calculations used.

- **Max Error**

  Max Error is a metric that captures the worst-case error between the predicted value and the actual value, which is defined as:

$$Max\ Error\,(y, \hat{y}) = \max\left(|y_i - \hat{y}_i|\right) \tag{6}$$

- **Mean Absolute Error**

  Mean Absolute Error (MAE) is a risk metric corresponding to the expected value of the absolute error loss or l1-norm loss, which is defined as:

$$MAE\,(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} |y_i - \hat{y}_i| \tag{7}$$

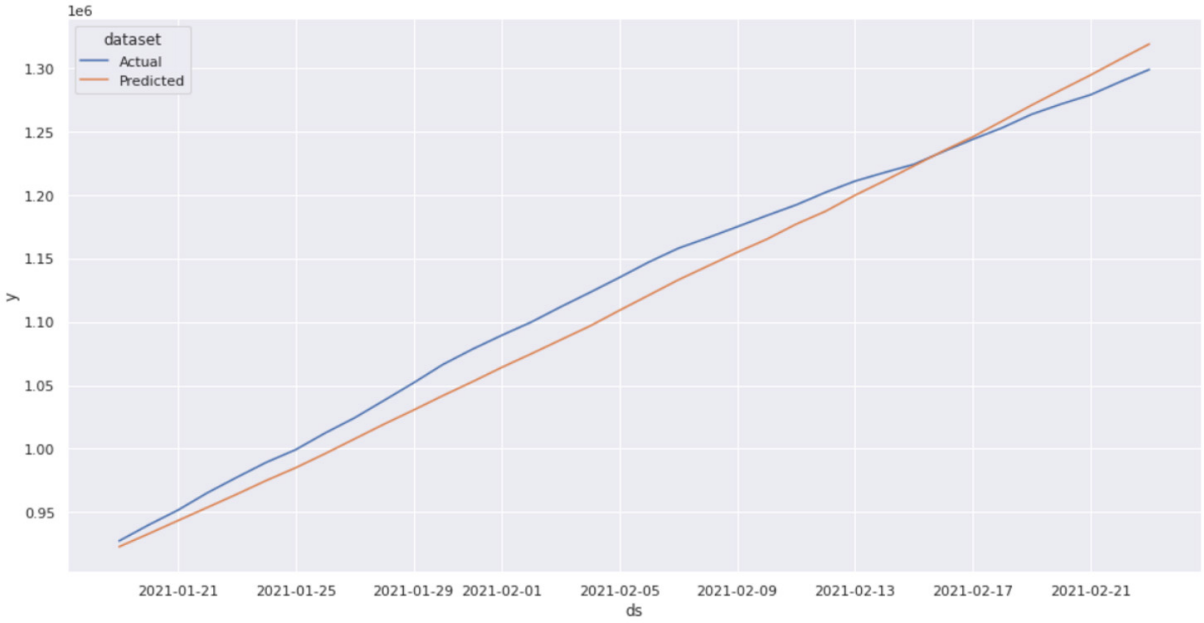

FIGURE 5. Comparison of actual data with prediction data

- **Root Mean Squared Error**
  Root Mean Squared Error (RMSE) is a risk metric corresponding to the expected value of the squared (quadratic) error or loss, which is defined as:

$$RMSE\,(y,\hat{y}) = \sqrt{\frac{1}{n_{samples}}\sum_{i=0}^{n_{samples}-1}(y_i-\hat{y}_i)^2} \tag{8}$$

- **R^2 Score**
  R^2 is a representation of the proportion of variance (of y) that has been explained by the independent variables in the model. It indicates goodness of fit and, therefore, a measure of how well-unseen samples are likely to be predicted by the model, through the proportion of explained variance. R^2 is defined as:

$$R^2\,(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{9}$$

$$\text{where } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ and } \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \sum_{i=1}^{n}\epsilon_i^2 \tag{10}$$

TABLE 4. Prophet model evaluation metrics

Model	Max Error	MAE	RMSE	R^2 Score
Prophet	26465.9045	15692.1276	17553.8462	0.9746

Based on the metrics, it can be seen that the value of R^2 is close to 1, where 1 is the highest accuracy figure. The mean absolute error value indicates that the predicted cases are not much different from the observed cases. The calculations have shown such satisfactory results that the Prophet model predicts a steady increase in the number of confirmed cases in the future [7].

4.5. **Prediction.** After doing evaluation using the testing set, we can predict future cases for the next 30 days using the trained model. We can see from Figure 6 that more cases were reported on Saturday and fewer cases reported on Tuesday. However, there might

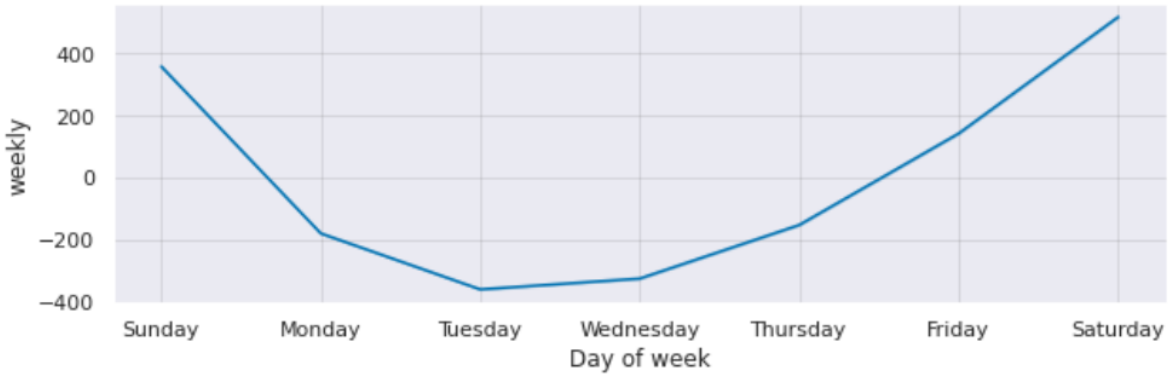

FIGURE 6. Weekly trend of the confirmed cases in Indonesia

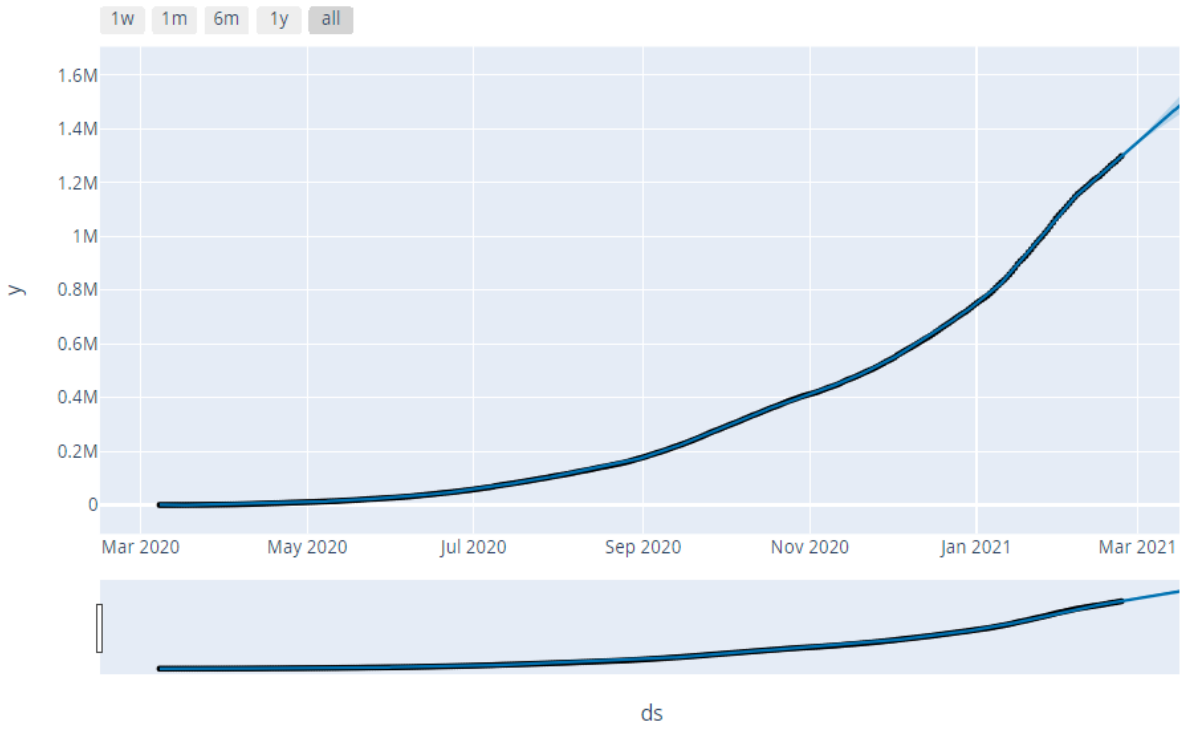

FIGURE 7. Predicted case plots

be a delay in the number of cases reported on time and data entry. The plot diagram in Figure 7 shows that by the end of March 2021, more than 1,400,000 cases might be witnessed.

5. **Conclusions.** The death toll from COVID-19 has risen sharply, particularly during the holidays. Therefore, it is necessary to predict the possibility of cases that will occur in the future as a basis for taking the necessary actions to prevent the spread of COVID-19 in Indonesia. In this paper, a prediction system for the number of COVID-19 cases in Indonesia is proposed based on the Prophet model using data obtained from Johns Hopkins University. The results showed that the value of R^2 is 0.9746, which means the model explains around 97% of the variation in the response variable around its mean. We will continue to refine this research, and for further research, we plan to predict the effects of the COVID-19 vaccine by using more datasets and modifying the model's hyperparameters in more detail [16]. Through this paper, it is hoped that it can help the government in overcoming the COVID-19 crisis in Indonesia.

Acknowledgment. This work is supported by Bina Nusantara University. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] T. P. Velavan and C. G. Meyer, The COVID-19 epidemic, *Tropical Medicine & International Health*, vol.25, no.3, 2020.

[2] Wikipedia contributors, *COVID-19 Pandemic*, https://en.wikipedia.org/wiki/COVID-19_pandemic, Accessed on February 28, 2021.

[3] I. Abdullah, COVID-19: Threat and fear in Indonesia, *Psychological Trauma: Theory, Research, Practice, and Policy*, vol.12, no.5, 2020.

[4] Indonesian COVID-19 Task Force, *COVID-19 Distribution Map in Indonesia*, https://covid19.go.id/, Accessed on February 28, 2021.

[5] R. Nasruddin and I. Haq, Large-scale social restrictions (LSSR) and low-income families, *SALAM: Islamic Journal of Social and Cultural Affairs*, vol.7, no.7, pp.639-648, 2020.

[6] G. Battineni, N. Chintalapudi and F. Amenta, Tropical conditions and outbreak of COVID-19, *Pharmaceutical and Biomedical Research*, vol.6, no.s1, pp.9-16, DOI: 10.18502/pbr.v6i(s1).4396, 2020.

[7] S. J. Taylor and B. Letham, Forecasting at scale, *PeerJ Preprints*, DOI: 10.7287/peerj.preprints.3190v2, 2017.

[8] M. Indhuja and P. P. Sindhuja, Prediction of COVID-19 cases in India using Prophet, *International Journal of Statistics and Applied Mathematics*, vol.5, no.4, 2020.

[9] A. K. Gupta, V. Singh, P. Mathur and C. M. Travieso-Gonzalez, Prediction of COVID-19 pandemic measuring criteria using support vector machine, Prophet, and linear regression models in the Indian scenario, *Journal of Interdisciplinary Mathematics*, pp.1-20, 2020.

[10] K. Abdulmajeed, M. Adeleke and L. Popoola, Online forecasting of COVID-19 cases in Nigeria using limited data, *Data in Brief*, vol.30, DOI: 10.1016/j.dib.2020.105683, 2020.

[11] J. Devaraj, R. M. Elavarasan, R. Pugazhendhi, G. Shafiullah, S. Ganesan, A. K. Jeysree and E. Hossain, Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant?, *Results in Physics*, vol.21, DOI: 10.1016/j.rinp.2021.103817, 2021.

[12] W. Robson, *The Math of Prophet*, https://medium.com/future-vision/the-math-of-prophet-46864fa9c55a, Accessed on March 1, 2021.

[13] R. S. Pontoh, S. Zahroh, H. R. Nurahman, R. I. Aprillion, A. Ramdani and D. I. Akmal, Applied of feed-forward neural network and Facebook Prophet model for train passengers forecasting, *Journal of Physics: Conference Series*, vol.1776, no.1, DOI: 10.1088/1742-6596/1776/1/012057, 2021.

[14] Github repository, *2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE*, https://github.com/CSSEGISandData/COVID-19, Accessed on February 25, 2021.

[15] B. Thirion, O. Grisel, E. Duchesnay et al., Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, vol.12, pp.2825-2830, 2011.

[16] X. Du, H. Xu and F. Zhu, Understanding the effect of hyperparameter optimization on machine learning models for structure design problems, *Computer-Aided Design*, vol.135, 2021.