International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

The FB Prophet Model Application to the Growth Prediction of International Tourists in Indonesia during the COVID-19 Pandemic

Suryaningsih Patandung[1], Ihsan Jatnika[2]
[1, 2]Information Systems Management, Postgraduate Program, University of Gunadarma
Jl. Margonda Raya No. 100, Depok 16424, West Java
Email address: [1]suryaningsihpatandung @ gmail.com, [2]ihsan @ staff.gunadarma.ac.id

Abstract— *COVID-19 declared as pandemic by the World Health Organization (WHO) on March 11, 2020. Based on data by the COVID-19 Task Force website the number of COVID-19 cases reached 126,921,284 cases in the world in March 29, 2021. This has an impact on various sectors are supports the Indonesian economy, such the tourism sector. This research predicted the growth of foreign tourists in Indonesia during COVID-19 by applying the Prophet Model. This model was developed by Facebook using a forecasting approach scheme, namely inputting the dataset, preprocessing, processing, post-processing, and output. The results of this research indicated that the growth rate of foreign tourists in Indonesia tends to decline until February 2022 and will begin to increase in March 2022. On the other side, the number of COVID-19daily cases tends to decline until December 2021. This research can be taken into consideration in decision making by parties who needed.*

Keywords— *COVID-19, FB Prophet, forecasting, time series data, tourist.*

I. Introduction

The COVID-19 outbreak declared as pandemic by the World Health Organization (WHO) on March 11, 2020. Based on data from the COVID-19 Task Force website, the number of COVID-19 cases reached 126,921,284 in the world as of March 29, 2021. Pandemic becomes a threat to various countries in the world, such Indonesia. One of the sectors affected the tourism sector is one of the pillars of the sustainability of the country's economy. The data presented by the Central Bureau of Statistics regards on monthly foreign tourist visits in Indonesia start from December 2019 to January 2021, the number of foreign tourist arrivals is actually tends to show a decline. Therefore, supporting data is needed to find out the plans and strategies can be carried out since pandemic so that the Indonesia tourism sector can slowly recover in the future. Information is needed on the number of tourists visit since the pandemic through forecasting tourist dynamics by using one of the forecasting models is the Prophet model by Facebook.

In previously, research by forecasting or prediction methods had been conducted by Sakib Mahmud, the prediction of COVID-19 in Bangladesh decreased from day to day after the maximum number of cases was seen on July 3, 2020. The prediction by the Prophet model showed that if the trend continues until July 21, cases will continue increasingly and predicted until the last day of September 19, 2020, daily cases will reach 5,750 [1]. Then, research conducted by K. Krishna Rani Samal et al. which shows the feasibility of using the Prophet forecasting model to predict future pollution levels and establish an early warning system for public safety. Research can be expanded by analyzing health care data to establish a correlation between health and future pollution levels [2].

In this study, researchers will apply the FB Prophet Model to predict the growth of the tourism sector in Indonesia during the COVID-19 pandemic based on data on the number of foreign tourist visits presented by the Central Statistics Agency. The results of the forecast based on the number of foreign tourist visits can be used to identify the growth rate of the tourism sector in Indonesia during the COVID-19 pandemic. In addition, this research can be used as a basis for consideration for decision-making regarding alternative policies, programs and activities by parties working in the tourism sector in Indonesia during the COVID-19 pandemic.

II. Theoretical Basic

A. Forecasting

Forecasting is about how to predict the future as accurately as possible supported by available information, including historical data and understanding / knowledge of future events that are likely to affect forecasts [3].

Forecasting has purpose to determine the accuracy and strength required by the forecasting technique chosen to determine forecasts will allow entry into the business area based on the size of the existing market. At the same time, the purpose of forecasting is to analyze the company's past, current policies and understand the extent of their impact in the future [4].

B. Time Series Analysis

Time series is a series of data sets sorted in chronological order. The time series units of the data include years, months, hours, or even milliseconds. As long as the data is stored in sequence based on the time, then the data is time series data. Therefore, it can be called that time series data are fundamentally different by other data. Techniques to analyzing time series data are also different. One of the most important techniques for describing and obtaining time series data patterns is to decompose the data.

There are 4 components in the decompose time series analysis technique. First, base / level, Base is the value of the data if the data set is a straight line. Second, a trend in which there are several cyclic components are different from the trend, but are usually combined with a trend. Third, seasonality which is due to seasonal factors, unique patterns are seen in time intervals. The reason for this is because it is a month of the year, a day of the month, or an hour of the day. Fourth, residual / noise / error is a variation of data cannot be explained.

C. Facebook Prophet

Prophet is an open source library (free) which is based on a decomposable model [5]. Prophet is a model for forecasting time series based on the additive model developed by the Facebook Data Science team. This model has the ability to make time series predictions with good accuracy using simple parameters. One of the advantages of Prophet is that it has support for including seasonality and irregular components. Researcher [6] in his research made a scalable forecasting approach which can be described as in Figure 1.

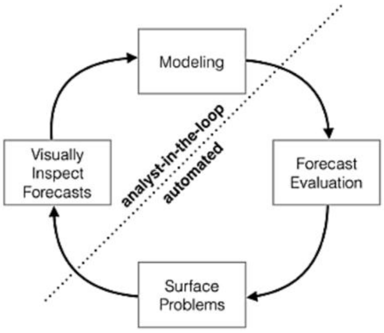

Fig. 1. Scale Forecasting Approach

The Prophet works with time series which can be broken down by three main components: trend, seasonality, and irregular components. The simplest form:

$$y_{(t)} = g_{(t)} + s_{(t)} + h_{(t)} + \varepsilon_{(t)}$$

The components of equation (1) are:

- g (t): piecewise linear or logistical growth curve for modeling non-periodic changes in time series S (t): periodic changes (eg weekly / annual seasonality)
- h (t): effect of holidays (user provided) with irregular schedules
- $\varepsilon\ (t)$: the error term accounts for any unusual changes are un-accommodated by the model.

D. Forecasting Accuracy

Accuracy represents the accuracy of the model used. Accuracy measurement methods can be divided [7] into measurements based on scale, measurement based on percentage error, scale error [8], and measurement based on relative error. Scale-dependent errors include mean error (ME), mean scale error (MSE), root-mean-squared error (RMSE), and mean absolute error (MAE). The method of measurement accuracy depends on the data scale used and should not be used to compare data sets with different scales.
In this research, the accuracy used is:

1) Mean Absolute Error (MAE): showed the average error value is the error of the actual value by the predicted value. MAE is generally used to measuring error prediction in time series analysis. The formula for MAE itself is defined as follows:

$$MAE = \sum \frac{|Y' - Y|}{n}$$

2) Mean Square Error (MSE): to calculate the MSE value as well as RMSE. It's just not using a root process. At this stage, if the error value is getting bigger, the MSE value will be bigger.

$$MAE = \sum \frac{|Y' - Y|2}{n}$$

3) R-Square: the coefficient of determination or R-Squared represents the proportion of variance on the dependent variable described by the linear regression model. It is a scale-free score i.e. regardless of the small or large value, the R-Square value will be less than one.

$$R^2 = 1 - \frac{\sum(yi - y)^2}{\sum(yi - y)^2}$$

III. REQUIREMENTS AND MODELS

A. Research Methodology

The research methodology carried out by the researcher can be seen in Figure 2. This research consists of five stages. The first stage is the input process for the tourist visit dataset, the second stage is the preprocessing process, the third stage is the modeling and forecast / prediction process, the fourth stage is post-process, and the last stage is displaying the results of the time series forecast as output.

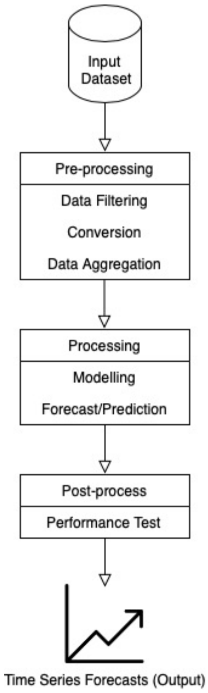

Fig. 2. Research Methodology

- The data used is time series data collected from a certain time series and describes the progress of an ongoing activity.

- Secondary data is data obtained from certain institutions or agencies that support research objectives.

Based on Figure 2, this research was carried out by going through the input dataset stage, preprocessing, processing, to post-process, it is necessary to test the performance of the model used for forecasting. The test in this study is a performance test of the forecasting results by looking at the error value calculated using the Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-Squared (R2).

B. *Data Sources*

Data Sources used in this study were obtained from the official website of the Central Statistics Agency (BPS). This data is data on foreign tourist visits to Indonesia from January 2018 to January 2021. The data presented by BPS is published in monthly periodicity based on the nationality of origin of the foreign tourists. There are 240 nationalities presented by BPS.

IV. RESULT

A. *Input Dataset*

Based on the research method has been described in, start from the forecasting application in this research, the data has been collected by the BPS website is inputted first. The data is contained in a table with two columns (variables), namely the Month column and the Visitor column. The Month column contains time data in monthly periodicity (YYYY-MM), while the Visitor column contains data on the number of tourist visits. Data is stored in .CSV format according to the standard Prophet model. The table of foreign tourist visits in Indonesia from January 2018 to January 2021 which is saved with the file name Wisata_Mancanegara.csv.

To be able to read the contents of the Tourist_Mancanegara.csv file, the researcher used the read_csv function from the pandas package "pd" object and stored it in the dataframe variable "tourism_df". Through this read in process, each line of data automatically gets an index starting at 0. The input and read in dataset processes can be seen in Figure 3.

```
data_to_load = files.upload()

Choose Files  Wisatawan_Ma...anegara.csv
• Wisatawan_Mancanegara.csv(text/csv) - 630 bytes, last modified: n/a - 100% done
Saving Wisatawan_Mancanegara.csv to Wisatawan_Mancanegara (1).csv

[275] tourism_df = pd.read_csv(io.BytesIO(data_to_load['Wisatawan_Mancanegara.csv']))
      tourism_df.head()
```

Fig. 3. Process of Input and Read in Dataset

B. *Preprocessing*

1) Filtering Data: from several types of existing filtering data, the researcher uses the Pandas dataframe filtering by rows position and column names. This filtering works by selecting data according to the row position and column name to be used. The researcher filtered the dataset by selecting the top 3 data rows to be displayed which were in the 0 to 3 index. In accordance with the location of the index and after defining the column / variable name selected, the results of the filtering data can be seen in Figure 4.

2) Conversion: the researcher made two conversions on the tourism_df dataframe. First, change the month time series variable data type to datetime64 [ns] by using the "to_datetime" function from the pandas package. The second conversion is to change the names of the Month and Visitor variables with the command "tourism_df.columns = ['ds', 'y']". The prophet model cannot work if the variable names containing time series data are not "ds" and "y" which contain constant values for forecasting. The process and results of the tourism_df dataframe conversion can be seen in Figure 5.

```
tourism_df.loc[tourism_df.index[0:3],["Month","Visitor"]]
```

	Month	Visitor
0	2018-01	1097839
1	2018-02	1197503
2	2018-03	1363426

Fig. 4. Filtering Data Results

```
[11] tourism_df.columns = ['ds', 'y']
     tourism_df['ds'] = to_datetime(tourism_df['ds'])

tourism_df.tail()
```

	ds	y
32	2020-09-01	148984
33	2020-10-01	152293
34	2020-11-01	144476
35	2020-12-01	164079
36	2021-01-01	141264

```
[13] tourism_df.dtypes

     ds    datetime64[ns]
     y              int64
     dtype: object
```

Fig. 5. Conversion Data

3) Aggregation Data: there is no need to group data based on time periodicity. However, in testing the function, data aggregation in this forecasting process can be included in the preprocessing series if needed at any time. Here, the researcher tested the application of data aggregation by grouping the data in quarters with the command "df.resample ('QS'). Sum ()". In this command, "QS" is the data collection for the periodicity of the quarter.

C. *Processing*

1) Modeling: The researcher makes graphical plotting using the matplotlib package. The plot of the data trend of foreign tourist visits in Indonesia from January 2018 to January 2021 can be seen in Figure 6.

```
tourism_df.plot()

<matplotlib.axes._subplots.AxesSubplot at 0x7fcdbda25710>
```

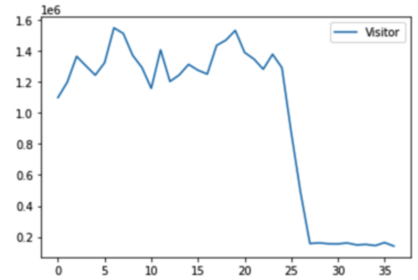

Fig. 6. Trend of Foreign Tourist Visits in Indonesia from January 2018 to January 2021

Suryaningsih Patandung and Ihsan Jatnika, "The FB Prophet Model Application to the Growth Prediction of International Tourists in Indonesia during the COVID-19 Pandemic," *International Research Journal of Advanced Engineering and Science*, Volume 6, Issue 2, pp. 110-115, 2021.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

The prophet model is referred to as the classifier for classification problems and a regressor for regression problems. The model is an object that is created to store the Prophet model's functions, then calls the .fit () function to make adjustments to the historical data available in the tourism_df dataframe using the "model.fit (tourism_df)" command. The function of .fit (tourism_df) is to train the Prophet model as a classifier with the tourism_df dataframe. After running the program code, Google Colab displays information on the status of the adjustments that have been made. The results of adjusting historical data in the tourism_df dataframe can be seen in Figure 7.

```
[49] model = Prophet()
     model.fit(tourism_df)

     INFO:fbprophet:Disabling weekly seasonality. Run prophet with weekly_seasonality=True to override this.
     INFO:fbprophet:Disabling daily seasonality. Run prophet with daily_seasonality=True to override this.
     <fbprophet.forecaster.Prophet at 0x7f26213db4d0>
```

Fig. 7. Fitting Model by Dataframe

After the model or classifier is formed, the researcher uses this model to create a predictable length of time. Based on information from the Ministry of Health of the Republic of Indonesia, vaccination is targeted to be completed in March 2022, so this study predicts for 15 months. With the built-in helper function make_future_dataframe (periods = N_month, freq = 'M') from the Prophet, we enter the value for N_month with 15 (15 months) and 'M' for the type of frequency. The variable "ds" shown in Figure 7. shows that the time series for the time to be predicted has been created. The actual time series data from January 2018 to January 2021 has increased by 15 months until March 2022.The data line index has also increased to the 51st index with a data length of 50.

```
[67] future = model.make_future_dataframe(periods=15, freq='M')
     future.tail()
```

	ds
47	2021-11-30
48	2021-12-31
49	2022-01-31
50	2022-02-28
51	2022-03-31

Fig. 8. New Time Series Framework

2) Forecasting: Prophet provides a .predict (future) function, a function that is called to perform predictive value calculations on future dataframes by inheriting the modeling properties that exist in the "model" in the form of an array. With that, create the command "forecast = model.predict (future)". The prediction method that has been executed, returns a row of predictive data in the future. The predicted value returned by the .predict (future) function is named YHAT. Some objects whose values are obtained in forecasting are YHAT, YHAT_LOWER, and YHAT_UPPER. YHAT contains the predicted value, while YHAT_LOWER and YHAT_UPPER contains the error range (the possibility of incorrect prediction data being in that interval). The results of the calculation of the predicted value and error range can be seen in Figure 9.

```
[68] forecast = model.predict(future)
     forecast[['ds','yhat','yhat_lower','yhat_upper']].tail()
```

	ds	yhat	yhat_lower	yhat_upper
47	2021-11-30	127555.185408	-2.637514e+05	543015.402775
48	2021-12-31	84922.742512	-3.157233e+05	470896.164300
49	2022-01-31	-492548.455621	-8.991365e+05	-120390.341139
50	2022-02-28	-712902.896779	-1.078528e+06	-328371.315115
51	2022-03-31	-427472.745218	-8.045362e+05	-43426.585788

Fig. 9. YHAT, YHAT_LOWER, YHAT_UPPER Calculation Result Framework

By calling the ".plot ()" function, the YHAT, YHAT_LOWER, YHAT_UPPER values are converted into a visualization that is easier to analyze. The visualization in graphic form can be seen in Figure 10. The black dot shows the actual dataset that has been input and is undergoing preprocessing. The blue line that has the image of the interval is the value of YHAT_LOWER (lowest point) and YHAT_UPPER (top point).

```
[170] model.plot(forecast)
      pyplot.show()
```

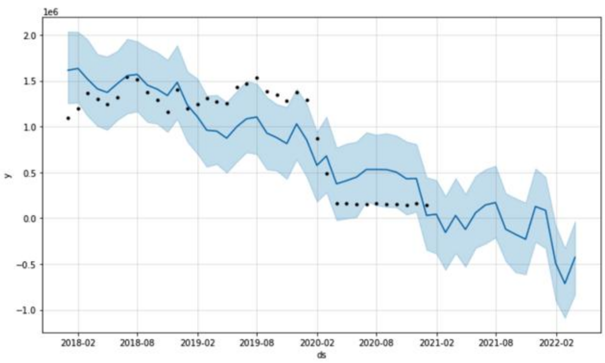

Fig. 10. Visualization of Forecasting Result

The researcher uses the function ".plot_components ()" where the parameter is the forecast variable to see the forecasting component which by default displays a graph of the trend and a graph in the annual group (yearly). The forecasting trend and yearly component graphs can be seen in Figure 11.

```
[171] model.plot_components(forecast);
```

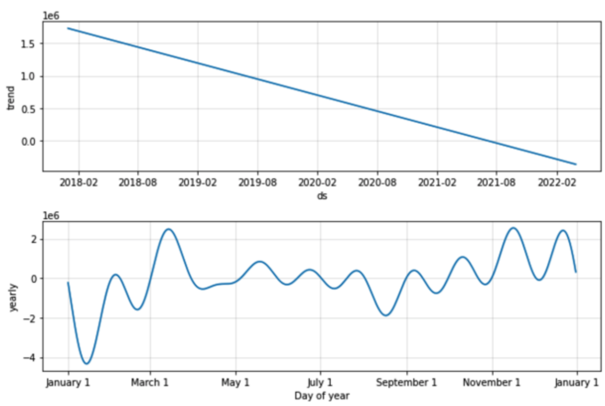

Fig. 11. Graphic of Trend and Forecasting Yearly Components

D. Postprocessing

```
metric_tourism = forecast.set_index('ds')[['yhat']].join(tourism_df.set_index('ds').y).reset_index()
metric_tourism.tail()
```

	ds	yhat	y
47	2021-11-30	127555.185408	NaN
48	2021-12-31	84922.742512	NaN
49	2022-01-31	-492548.455621	NaN
50	2022-02-28	-712902.896779	NaN
51	2022-03-31	-427472.745218	NaN

Fig. 12. metric_tourism function

The value of the variable "y" reads as Not a Number in the dataframe so it is changed by using the .dropna (inplace = True) function call. With this function, the actual "y" value is contained in numbers. Since the values needed to calculate the error value for each method are available, namely the values for "yhat" and "y", the researcher calculates each method by calling the objects from sklearn.metrics. The results of each method's calculation can be seen in Figure 12, Figure 13, and Figure 14.

```
[176] mean_absolute_error(metric_tourism.y, metric_tourism.yhat)

      271494.5740878813
```

Fig. 12. *Mean Value of Absolute Error*

```
[177] mean_squared_error(metric_tourism.y, metric_tourism.yhat)

      94368423915.85638
```

Fig. 13. *Mean Value of Squared Error*

```
[178] r2_score(metric_tourism.y, metric_tourism.yhat)

      0.6647893230649078
```

Fig. 14. *Value of R-Squared Error*

The results of the calculation of the error value of the three methods can be said to be quite large. Researchers realize this can occur because the COVID-19 pandemic has occurred for approximately 1 year which resulted in a lack of historical data to make calculation accuracy better.

E. Output

Output Forecasting research closes by comparing the results of the calculation of the value "y" which in this case represents the number of tourist visits in Indonesia from the original value / actual value with the predicted / predicted value. The visualization of the comparison of the "y" values can be seen in Figure 15.

```
pyplot.plot(metric_tourism.y, label='Actual')
pyplot.plot(metric_tourism.yhat, label='Predicted')
pyplot.legend()
pyplot.show()
```

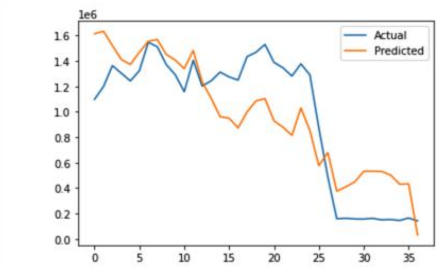

Fig. 15. Comparison Graph of Actual and Predicted "y"

Values From the visual graph, it can be seen that the line pattern between the actual line and the predicted line does not make a big difference in the distance. This shows that the calculation of the predictive value that applies the Prophet model can be an option for making predictions in the future.

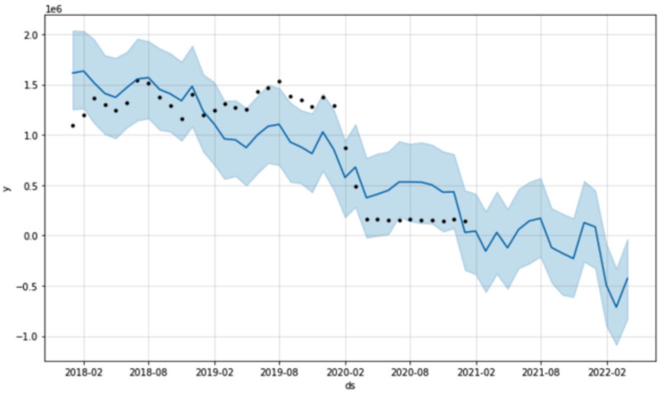

Fig. 16. Forecasting Results of the Growth Rate of Foreign Tourists from February 2021 to March 2022

Based on forecasting with the Prophet model, the results showed that in the next 15 months, namely February 2021 (after the last foreign tourist visit data) to March 2022, foreign tourist visits to Indonesia will tend to decline. However, predictions also showed that prior to February 2022 there were several quite high increases in the number of tourists, namely in May and November 2021. This occurs because these two months are the periods when countries enter the holiday season. After December 2021, predictions again show a decline in the number of tourists until February 2022 and starting to increase in March 2022. The visualization of the results of forecasting the growth rate of foreign tourists from February 2021 to March 2022 can be seen in Figure 16.

The parameters that can be used for forecasting the growth rate of foreign tourist arrivals in Indonesia are limited due to the univariate nature of the Prophet model, so that it is only determined by time series parameters and the number of visitors. Therefore, researchers use the results of forecasting of daily COVID-19 case data in Indonesia to see and compare the two forecasts as supporting information that can strengthen the prediction results, so that conclusions can be taken into consideration for those needs the information by this research.

Based on daily COVID-19 case data as of March 29, 2021 from https://covid19.go.id, forecasting is carried out with the same stages and processes as forecasting the previous rate of foreign tourist arrivals in Indonesia. The prediction results show that the daily cases of COVID-19 starting in April 2021 will experience a tendency for the number of cases to decline until December 2021, although there have been several increases. This is indicated by the value of "y" which reached the 0 line in December 2021. In Figure 17, it can be noticed that there are areas highlighted in light blue. The area is the shaded area because it is at the lowest and highest point of the predicted value. Therefore, the number of daily cases of COVID-19 has the opportunity to increase or decrease in the interval area. Visualization of daily COVID-19 case predictions from April 2021 to December 2021 can be seen in Figure 17.

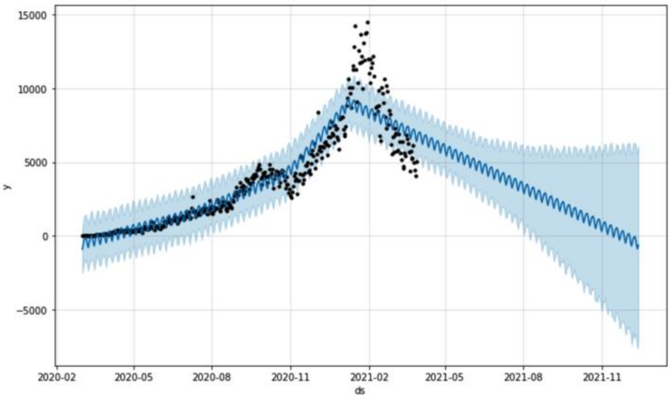

Fig. 17. Prediction of COVID-19 Daily Cases April 2021 to December 2021

The prediction of COVID-19daily cases in Indonesia that has been carried out is then used to compare the predicted value as a support with the results of forecasting the growth rate of foreign tourists in Indonesia during the COVID-19 pandemic. The comparison of forecasting results carried out from the growth rate of foreign tourists in Indonesia during the COVID-19 pandemic with the daily cases of COVID-19 in Indonesia showed that the growth rate of foreign tourists in Indonesia during the COVID-19 pandemic has increased quite significantly, especially in May - August 2021 and November - December 2021. This is in line by the number of daily cases of COVID-19 in Indonesia is tends to decline until December 2021. The rate of tourist growth will start to pick up again in March 2022. It should be noted that in the predicted time intervals of the two spheres, although the trend is decreasing, the increase in COVID-19daily cases has occurred several times.

V. Conclusions And Suggestions

A. Conclusions

Some of the things that can be concluded from this research are as follows:

- Forecasting methods has using the Prophet model on a dataset of the number of foreign tourist visits in Indonesia can be done and go through each process stage of the Prophet model. Starting from preparing and entering datasets, preprocessing data by filtering, conversion, aggregation. Then, model the dataset that is owned into the Prophet model and calculate predictions, to measure the accuracy of the predictions made.
- Performance test is performed using the Mean Absolute Error, Mean Square Error, and R-Squared methods. The error test value of the three methods can be said to be quite large. MAE is 271494.6, MSE is 94368423915.9, and R-Squared is 0.67. Researchers are aware that this can occur because the COVID-19 pandemic occurred for approximately one year which resulted in a lack of historical data to make calculation accuracy better.

B. Suggestions

From the research that has been done, the researcher can provide the following suggestions:

- The results of this research can be further developed by certain elements or parties who need to predict the growth rate of tourists in Indonesia during the COVID-19 pandemic as a consideration in making decisions, either for personal or organizational needs.
- Future research should use a multivariate forecasting model to correlate the rate of foreign tourists in Indonesia with the number of COVID-19daily cases.

References

[1] S. Mahmud, “Bangladesh COVID-19 Daily Cases Time Series Analysis using Facebook Prophet Model,” *SSRN Electron. J.*, vol. 19, no. May 2020, 2020, doi: 10.2139/ssrn.3660368.

[2] E. Žunić, K. Korjenić, K. Hodžić, and D. Đonko, “Application of Facebook’S Prophet Algorithm for Successful Sales Forecasting Based On Real-World Data,” *arXiv*, vol. 12, no. 2, pp. 23–36, 2020, doi: 10.5121/ijcsit.2020.12203.

[3] G. Athanasopoulos, R. J. Hyndman, N. Kourentzes, and F. Petropoulos, “Forecasting with temporal hierarchies,” *Eur. J. Oper. Res.*, vol. 262, no. 1, pp. 60–74, 2017, doi: 10.1016/j.ejor.2017.02.046.

[4] A. Kolková, “And Methods Based On Technical Indicators In The Field Of Measuring The Accuracy Of Quantitative Prognostic Methods And Methods Based On Technical Indicators In The Field Of Tourism,” no. August, 2018.

[5] M. Navratil and A. Kolková, “Business Economy Using Prophet Forecasting,” pp. 26–39, 2019.

[6] S. J. Taylor & B. Letham, "Forecasting at Scale," *PeerJ Preprints*, 2017, doi: 10.7287/peerj.preprints.3190v2.

[7] I. Yenidogan, A. Cayir, O. Kozan, T. Dag, and C. Arslan, “Bitcoin Forecasting Using ARIMA and PROPHET,” *UBMK 2018 - 3rd Int. Conf. Comput. Sci. Eng.*, no. September, pp. 621–624, 2018, doi: 10.1109/UBMK.2018.8566476.

[8] K. K. R. Samal, K. S. Babu, S. K. Das, and A. Acharaya, “Time series based air pollution forecasting using SARIMA and prophet model,” *ACM Int. Conf. Proceeding Ser.*, pp. 80–85, 2019, doi: 10.1145/3355402.3355417.