E-Jurnal Matematika Vol. 13(3), Agustus 2024, pp. 160-166 ISSN: 2303-1751
DOI: https://doi.org/10.24843/MTK.2024.v13.i03.p457

COMPARISON OF VISITS TO THE FACEBOOK WEBSITE USING THE FACEBOOK PROPHET AND SARIMA MODELS

Ni Pande Luh Tresnawardani[1], I Wayan Sumarjaya[2§], Ratna Sari Widiastuti[3]

[1]Program Studi Matematika, Fakultas MIPA-Universitas Udayana [Email: tresnawardani02@gmail.com]
[2]Program Studi Matematika, Fakultas MIPA-Universitas Udayana [Email: sumarjaya@unud.ac.id]
[3]Program Studi Matematika, Fakultas MIPA-Universitas Udayana [Email: ratnasariwidiastuti@unud.ac.id]
[§]*Corresponding Author*

ABSTRACT

Facebook is widely used across various segments of society and offers significant potential as a marketing tool due to its diverse content and communication features. This study aims to forecast visits to the Facebook website for the period of March 2023 to February 2024 using two forecasting methods: the Facebook Prophet and SARIMA models. We analyzed 167 data points representing the percentage of visits to the Facebook website from April 2009 to February 2023, sourced from Statcounter. Our forecasting results indicate that the Facebook Prophet model yielded a Mean Absolute Percentage Error (MAPE) of 22.3% and a Root Mean Square Error (RMSE) of 16.2%. In comparison, the SARIMA model achieved a MAPE of 6.31% and an RMSE of 5.02%. Based on these accuracy metrics, the SARIMA model is determined to be more suitable for forecasting the percentage of visits to the Facebook social media website in Indonesia. This finding underscores the importance of selecting appropriate forecasting models for accurate predictions in social media analytics

Keywords: *facebook prophet model, forecasting, SARIMA*

1. INTRODUCTION

The Internet today drives advancements in various fields. Social networks have revolutionized communication, allowing information to be shared more diversely and quickly, reaching many people simultaneously without time or distance constraints. Social networks or social media enable users to interact, communicate, share, collaborate, and form virtual social relationships (Widada, 2018).

According to the 2023 report by We Are Social, there are 4.76 billion active social media users globally. In Indonesia alone, there are 167 million active social media users, which accounts for 60.4% of the population. The report also indicates that in 2023, Facebook was the most visited social media website in Indonesia, following Google and YouTube.

Facebook is a social media website that allows each user to join communities such as those based on college, city, workplace, and hobbies, by connecting and interacting with others. Facebook is popular among various segments of society and holds significant potential as a marketing tool and business opportunity due to its diverse content and communication features (Syaharullah et al., 2021). Online business activities facilitated by Facebook include promoting products to consumers by uploading photos or videos, conducting transactions, and communicating with consumers without direct meetings between sellers and buyers. Among teenagers, Facebook is often used to communicate and share daily activities through posts directed at friends or the public. Friend networks can be expanded globally through Facebook, making it easier to exchange information and ideas effectively (Hanafi, 2016).

Considering Facebook's status as widely used social media platform and its diverse benefits for society, it is valuable to analyze and forecast visits to the Facebook website. This analysis can provide insights into the trends and potential impacts of emerging social media platforms.

In this study, a comparison of the forecasted percentage of visits to the Facebook website in Indonesia was conducted using the Facebook Prophet and SARIMA models for the period from October 2023 to October 2024. The Facebook

Prophet model is designed to handle common features of time series data (Taylor & Letham, 2018). The Facebook Prophet model employs a simple modular regression model that generally performs well with default parameters. These default parameters refer to the built-in values used by the model, such as settings for identifying trends, handling seasonality, and managing holiday effects. Moreover, the Facebook Prophet model excels in processing data with recurring daily patterns. Even when there are significant outliers and substantial changes in patterns or trends, this model can understand and depict multiple seasonal patterns within the data simultaneously (Zhao et al., 2018). Thus, this model can produce accurate and reliable forecasts (Chandra & Budi, 2020; Gaur, 2020)

This model is highly suitable for use when the time series data exhibits strong seasonal effects and ample historical data is available. Additionally, the Facebook Prophet model can also perform well on data that does not exhibit seasonal effects (Satrio et al., 2021). The Facebook Prophet model is categorized as a semi-automatic forecasting tool. In this model, tasks are essentially divided between human inputs and automated processes (Prakoso et al., 2023).

In addition to the Facebook Prophet model, this study also employs the Seasonal Autoregressive Integrated Moving Average (SARIMA) method as a comparison to determine the best model for forecasting visits to the Facebook website. The SARIMA method is an extension of the ARIMA method that incorporates seasonal effects. One of the advantages of this method is its ability to accommodate all types of time series patterns, although the data must first be made stationary. SARIMA uses past and present values of the dependent variable to generate accurate forecasts. In SARIMA, seasonal differencing is used to transform non-stationary data into stationary data (Box, 2013).

The findings of this research indicate a monthly decline in the percentage of Facebook visits in Indonesia. The results of this study provide valuable insights for social media account owners and business owners on the Facebook platform, enabling them to better understand the platform's popularity trends and optimize their promotional and communication strategies accordingly.

2. RESEARCH METHOD

2.1 Data Type and Source

The variable used for forecasting is the percentage of visits to the Facebook social media website in Indonesia. This percentage of visits is calculated based on the comparison between visits to the Facebook social media website and the total visits to social media websites recorded on the Statcounter website accessed at the following link: [https://gs.statcounter.com/social-media-stats].

The type of data used in the research is secondary data on the percentage of visits to the Facebook social media website in Indonesia from April 2009 to February 2023. This dataset includes the percentage of visits to various social media websites in Indonesia. However, for this study, only the data on the percentage of visits to the Facebook website in Indonesia will be utilized.

2.2 Exploration, Processing, and Plotting of Data

In this stage, the data undergoes an exploratory process before being used in modeling. Data exploration can help identify patterns, trends, outliers, and relationships between variables in the data. Then, data processing is conducted to classify and summarize the data for further analysis. This processing includes steps such as data cleaning (handling missing values, removing duplicates), data transformation (normalizing or scaling features), and data integration (combining data from different sources if necessary). Subsequently, data is plotted to visualize it in the form of graphs, tables, diagrams, or maps for easier understanding. Data plots depict relationships, patterns, or trends within the data.

2.3 Modeling with the Facebook Prophet Model

In this stage, modeling begins with the Facebook Prophet model. The steps for modeling with the Facebook Prophet model are as follows:

a. Plot the data.
b. Convert the data type to a tsibble data type to facilitate modeling in R.
c. Modeling with Facebook Prophet by using the fit method on the model object with data as an argument.
d. Perform forecasting.

2.4 Modeling with the SARIMA

In this stage, modeling with the SARIMA model begins. The steps for modeling with the SARIMA model are as follows:

a. Plot the data.
b. Determine the stationarity of the data in terms of both variance and mean. Checking stationarity of the data with ACF and PACF plots and conduct the augmented Dickey-Fuller (ADF) test to identify stationarity more effectively.
c. Apply Box-Cox transformation to non-stationary data in variance, perform non-seasonal differencing on non-stationary data against non-seasonal mean, and perform seasonal differencing on non-stationary data against seasonal mean.
d. Identify the SARIMA model based on ACF and PACF plots of stationary data, then conduct significance tests for temporary model parameters.
e. Perform diagnostic tests on the model by testing residual white noise assumptions and checking residual normal distribution.
f. Determine the best SARIMA$[(p,d,q)(P,D,Q)^s]$ model for forecasting based on AIC values.
g. Conduct forecasting using the best model.

2.5 Forecasting and Forecast Evaluation

This stage occurs after obtaining the best Facebook Prophet and SARIMA models. These models will be used for forecasting the percentage of visits to the Facebook website in Indonesia for the period from March 2023 to February 2024. Subsequently, an evaluation of the forecasting results is conducted. The forecasting results from the Facebook Prophet and SARIMA models will be compared, and conclusions will be drawn to explain the results obtained. The accuracy of the forecasts will be assessed using root mean square error (RMSE) and mean absolute percentage error (MAPE) tests (Makridakis et al., 1999). Root Mean Square Error (RMSE) is derived from MSE and represents the square root of MSE. Mean Square Error (MSE) calculates the extent to which the forecasted values differ from the actual values. These differences arise due to random elements in the data or inaccuracies in the obtained forecasts. RMSE is used to normalize the units similarly to MAPE. The formula for RMSE can be written as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(x_t - f_t)^2}{n}}$$

Mean Absolute Percentage Error (MAPE) is a calculation method used to obtain the mean absolute value of the differences between the forecasted values and the actual values. The forecasting results will be more accurate if the MAPE value is smaller. Generally, the formula for MAPE can be written as follows.

$$MAPE = \frac{\sum_{t=1}^{n}\left|\frac{x_t - f_t}{x_t}\right|}{n}\ x\ 100\%$$

3. RESULT AND DISCUSSION

3.1 Data Plot

Data plotting is performed to visualize and analyze patterns and trends in the data. The dataset consists of the percentage of visits to the Facebook social media website in Indonesia from April 2009 to February 2023, totaling 167 data points. Data visualization is conducted using R software to help identify these patterns and trends.

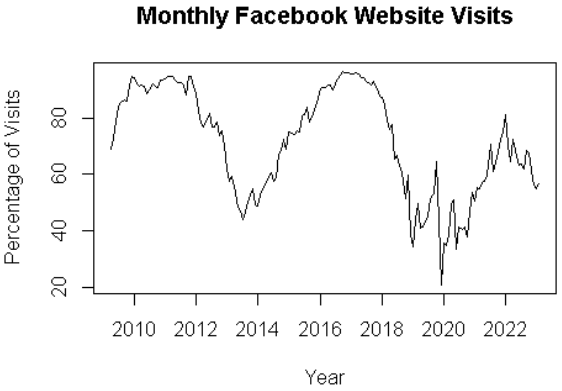

Figure 1. Graph of the Percentage of Visits to Facebook in Indonesia

3.2 Facebook Prophet Model

As the Facebook Prophet model will be used for forecasting, the data type will be converted to a tsibble format.

After the data is converted into tsibble format, modeling is conducted with Facebook Prophet using the fit method on the model object with data as the argument. The decomposition results of the Facebook Prophet model is shown in Figure 2.

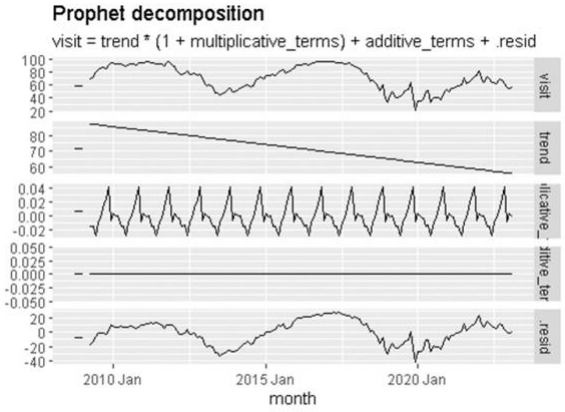

Figure 2. Graph of the Components of the Facebook Prophet model

After obtaining the model on Figure 2, it will be used to forecast the percentage of visits to the Facebook website in Indonesia for the period from March 2023 to February 2024. The forecasted results are as follows:

Table 1. Forecast Results with the Facebook Prophet Model

Month	Forecast (%)
March 2023	55,8
April 2023	54,9
Mei 2023	54
June 2023	53,6
July 2023	54
August 2023	54,6
September 2023	54,6
October 2023	55,8
November 2023	55,9
December 2023	53,4
January 2024	54
February 2024	53,5

with the forecast results graph displayed in Figure 3.

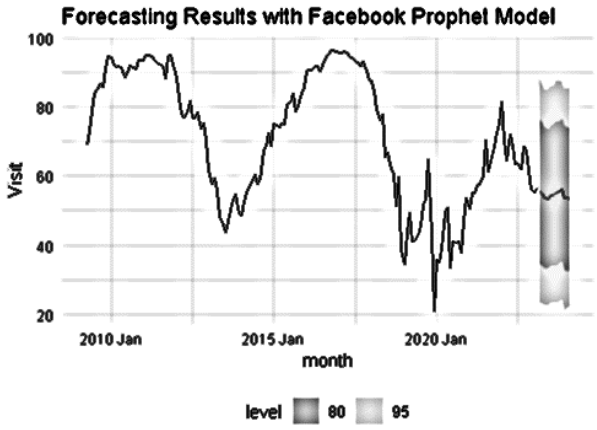

Figure 3. Plot of the forecast results with the Facebook Prophet model

3.3 SARIMA Model

a. Data Stationarity

Before modeling using SARIMA, the data must exhibit stationarity. Data can be considered stationary if it does not change over time. As seen in Figure 3, the data displayed is not stationary in both mean and variance.

To ensure data stationarity more effectively, the Augmented Dickey-Fuller test was conducted in R, yielding a p-value of 0,7182 with a significance level of 0,05. This indicates that the data is not significant because the p-value is greater than 0,05, indicating that the data is not stationary. Therefore, differencing process is conducted on the data to obtain stationary results. After differencing process on the trend component, the data plot obtained is displayed in Figure 4.

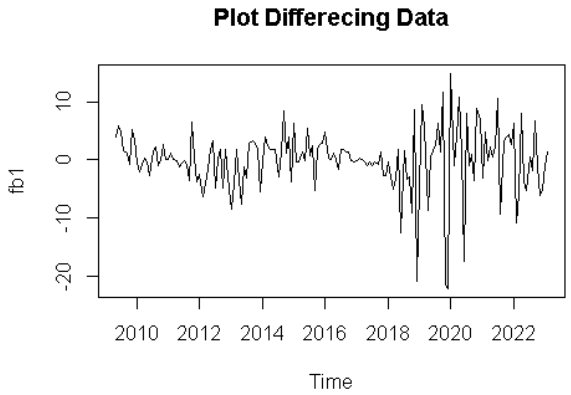

Figure 4. Plot of the data after differencing on the trend

In the data plot, stationarity is still not apparent. Therefore, differencing was performed again on the seasonal component, resulting in the following plot:

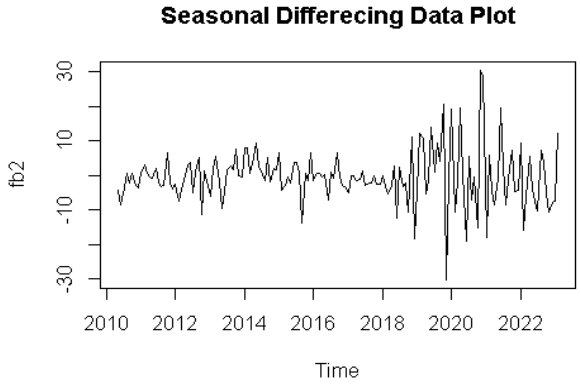

Figure 5. Plot of the data after seasonal differencing

The data can be seen as stationary with a graph pattern centered around the mean. An ADF test was conducted again to observe the stationarity of the data, yielding a p-value of 0,01 at a significance level of 0,05. This implies that

the data is significant; since the p-value is less than 0,05, the data is stationary and can be utilized for modeling purposes.

b. Significance Test of Temporary Model Parameters

The next step is to determine the temporary model parameters. The order of the SARIMA$(p,d,q)(P,D,Q)^S$ can be observed in the ACF and PACF plots of the differenced data that displayed in Figure 6 and Figure 7.

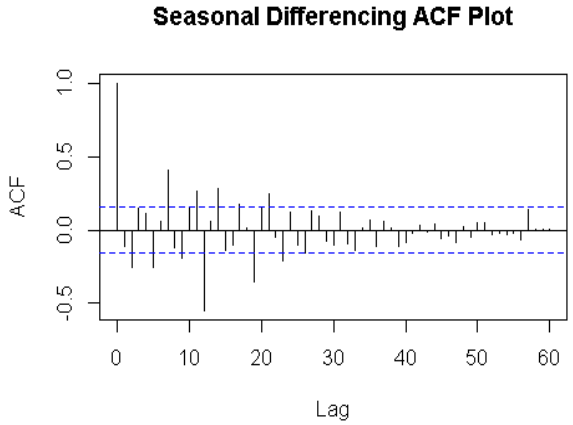

Figure 6. ACF Plot of the data after seasonal differencing

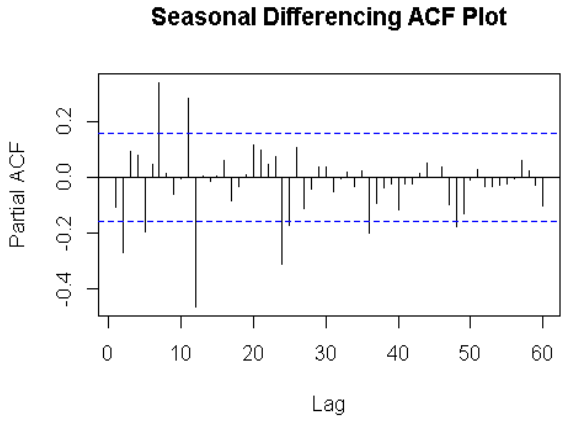

Figure 7. PACF Plot of The Data After Seasonal Differencing

Based on the ACF plot in the figure, it can be observed that the seasonal order experiences a cut-off after the first lag, and the PACF plot gradually dies down after the first lag, resulting in an SMA(1) model with 1 representing the Q order in the model.

After obtaining the Q order, trial and error are conducted to determine a model that is significant for the data. The candidate models used are SARIMA$(0,1,1)(0,1,1)^{12}$, SARIMA$(0,1,2)(0,1,1)^{12}$, SARIMA$(1,1,0)(0,1,1)^{12}$, SARIMA$(1,1,1)(0,1,1)^{12}$, SARIMA$(1,1,2)(0,1,1)^{12}$, SARIMA$(2,1,0)(0,1,1)^{12}$, SARIMA$(2,1,1)(0,1,1)^{12}$, and SARIMA$(2,1,2)(0,1,1)^{12}$.

Each model will determine the significance of each tested parameter in R. The results indicate significance if the significance value of each parameter is less than 0,05. The test results obtained can be seen in Table 2.

Table 2. Results of the significance test of the model parameters

No.	Model	Parameter Significance
1	SARIMA$(0,1,1)(0,1,1)^{12}$	Not significance
2	SARIMA$(0,1,2)(0,1,1)^{12}$	Not significance
3	SARIMA$(1,1,0)(0,1,1)^{12}$	Not significance
4	SARIMA$(1,1,1)(0,1,1)^{12}$	**Significance**
5	SARIMA$(1,1,2)(0,1,1)^{12}$	Not significance
6	SARIMA$(2,1,0)(0,1,1)^{12}$	Not significance
7	SARIMA$(2,1,1)(0,1,1)^{12}$	Not significance
8	SARIMA$(2,1,2)(0,1,1)^{12}$	**Significance**

Based on the results in the table, two models were obtained where each parameter is significant, namely the SARIMA$(1,1,1)(0,1,1)^{12}$ model and the SARIMA$(2,1,2)(0,1,1)^{12}$model.

c. Model Diagnostic Test

After obtaining significant models, a diagnostic test of the model is conducted to determine whether the model is suitable for use or not. The tests used are the normality test and the white noise assumption test.

Referring to Table 3, it's evident that the p-values from the normality test and the white noise assumption test are greater than 0,05, indicating that each model is normally distributed and meets the white noise assumption.

Table 3. Results of the diagnostic test for both models

Model	Test	*p-value*
SARIMA$(1,1,1)(0,1,1)^{12}$	Normality	0,0842
	White noise	0,4501
SARIMA$(2,1,2)(0,1,1)^{12}$	Normality	0,2299
	White noise	0,5088

d. Model Selection and Forecasting

Before making forecasts, one of the best-fit models will be chosen because both models pass the diagnostic test. To determine the best model,

the AIC values of both model obtained can be seen in Table 4.

Table 4. AIC values of SARIMA models

Model	AIC Values
$SARIMA(1,1,1)(0,1,1)^{12}$	992,1734
$SARIMA(2,1,2)(0,1,1)^{12}$	981,2297

According to the findings presented in the table, the $SARIMA(2,1,2)(0,1,1)^{12}$ model, with the lowest AIC value of 981.2297, stands out as the most suitable model and will be employed for future forecasts. The $SARIMA(2,1,2)(0,1,1)^{12}$ model can be written as follows:

$$(1-\phi_1 B-\phi_2 B^2)(1-B)^1(X_t-\mu) = (1+\theta_1 B)(1-B^{12})e_t$$

Next, the forecasting of visits to the Facebook social media website in Indonesia for the period from March 2023 to February 2024 is carried out. The forecast results are as follows:

Table 5. Forecast Results with the SARIMA Model

Month	Forecast (%)
March 2023	57,28320
April 2023	58,90494
Mei 2023	57,35637
June 2023	55,12007
July 2023	58,23460
August 2023	57,65489
September 2023	57,79701
October 2023	59,41402
November 2023	60,47615
December 2023	55,35049
January 2024	56,17947
February 2024	56,84108

The forecast results obtained are depicted in Figure 8, showing the forecasted graph.

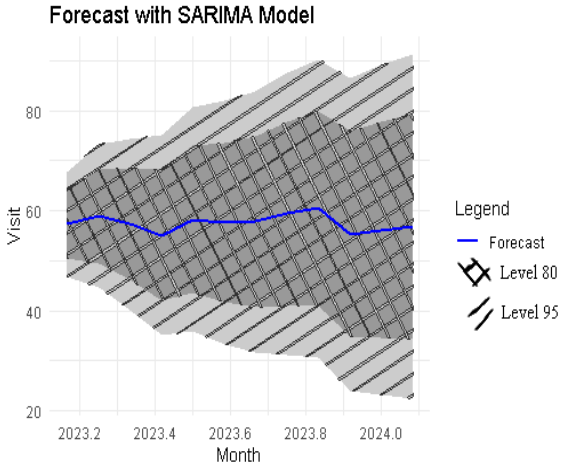

Figure 8. Forecasted Graph Using The SARIMA Model

3.4 Comparison between Facebook Prophet and SARIMA

After analyzing the data using the Facebook Prophet and SARIMA models to forecast the percentage of visits to the Facebook website in Indonesia for the period from March 2023 to February 2024, accuracy testing is conducted. Accuracy testing is performed to assess how accurate the models are in forecasting. To test the accuracy of the models, the forecasted data from the Facebook Prophet and SARIMA models are evaluated for their errors using MAPE and RMSE. The accuracy test results for each model detailed in the subsequent table:

Table 6. Accuracy Test Results for Both Facebook Prophet and SARIMA Models

Model	Accuracy Test	Results (%)
Facebook Prophet	MAPE	22,3
	RMSE	16,2
$SARIMA(2,1,2)(0,1,1)^{12}$	MAPE	6,313498
	RMSE	5,023458

As seen in the table, the accuracy test results for the Facebook Prophet model show a MAPE value of 22.3%, indicating it is sufficiently accurate, and an RMSE value of 16.2%. The accuracy test results for the $SARIMA(2,1,2)(0,1,1)^{12}$ model show a MAPE value of 6.31%, signifying very good accuracy, and an RMSE value of 5.02%.

4. CONCLUSION AND SUGGESTION

4.1 Conclusion

Based on the analysis results, the SARIMA(2,1,2)(0,1,1) 12 model demonstrates significantly better accuracy compared to the Facebook Prophet model. This is evidenced by the notable differences in both MAPE and RMSE values. Facebook Prophet model yielded a Mean Absolute Percentage Error (MAPE) of 22.3% and a Root Mean Square Error (RMSE) of 16.2%. In comparison, the SARIMA model achieved a MAPE of 6.31% and an RMSE of 5.02%. Therefore, it can be concluded that the SARIMA(2,1,2)(0,1,1) 12 model is suitable for forecasting the percentage of visits to the Facebook social media website in Indonesia for the period from March 2023 to February 2024.

4.2 Suggestion

For further research using the Facebook Prophet model, it is advisable to conduct a more in-depth analysis by incorporating holiday effects into the Facebook Prophet model. This addition can assist the model in accounting for patterns that emerge during holiday periods, thus enhancing the model's performance in forecasting data more accurately.

REFFERENCES

Box, G. (2013). Box and Jenkins: Time Series Analysis, Forecasting and Control. In *A Very British Affair* (pp. 161–215). Palgrave Macmillan UK. https://doi.org/10.1057/9781137291264_6

Chandra, C., & Budi, S. (2020). Analisis Komparatif ARIMA dan Prophet dengan Studi Kasus Dataset Pendaftaran Mahasiswa Baru. *Jurnal Teknik Informatika Dan Sistem Informasi*, *6*(2). https://doi.org/10.28932/jutisi.v6i2.2676

Gaur, D. S. (2020). GLOBAL FORECASTING OF COVID-19 USING ARIMA BASED FB-PROPHET. *International Journal of Engineering Applied Sciences and Technology*, *5*(2), 463–467. https://doi.org/10.33564/IJEAST.2020.v05i02.077

Hanafi, M. (2016). Pengaruh penggunaan media sosial facebook terhadap motivasi belajar mahasiswa FISIP Universitas Riau. *Jurnal Online Mahasiswa Fakultas Ilmu Sosial Dan Ilmu Politik Universitas Riau*, *3*(2), 1–12.

Makridakis, S., Wheelwright C, S., & McGee, V. E. (1999). Metode dan Aplikasi Peramalan. In *Erlangga*.

Prakoso, F. B., Darmawan, G., & Bachrudin, A. (2023). Penerapan Metode Facebook Prophet Untuk Meramalkan Jumlah Penumpang Trans Metro Bandung Koridor 1. *ARMADA : Jurnal Penelitian Multidisiplin*, *1*(3), 133–147. https://doi.org/10.55681/armada.v1i3.416

Satrio, C. B. A., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science*, *179*(2020), 524–532. https://doi.org/10.1016/j.procs.2021.01.036

Syaharullah, Yahya, M., & Syarif, A. (2021). Penggunaan Facebook Dalam Promosi Produk Barang Jadi. *Jurnal Kajian Islam Kontemporer*, *12*(2), 27–38.

Taylor, S. J., & Letham, B. (2018). Forecasting at Scale. *American Statistician*, *72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

Widada, C. K. (2018). MENGAMBIL MANFAAT MEDIA SOSIAL DALAM PENGEMBANGAN LAYANAN. *Journal of Documentation and Information Science*, *2*(1), 23–30. https://doi.org/10.33505/jodis.v2i1.130

Zhao, N., Liu, Y., Vanos, J. K., & Cao, G. (2018). Day-of-week and seasonal patterns of PM2.5 concentrations over the United States: Time-series analyses using the Prophet procedure. *Atmospheric Environment*, *192*(August), 116–127. https://doi.org/10.1016/j.atmosenv.2018.08.050