Vol. 4, No. 2, December 2025, pp. 39-49
p-ISSN: 2963-8577
e-ISSN: 2964-3511
DOI: 10.57102/jescee.v4i2.101

SHORT-TERM ELECTRICAL LOAD FORECASTING USING LINEAR REGRESSION AND TIME SERIES MODELS: A CASE STUDY AT PT. PLN (PERSERO) TARAKAN

Kartika Putri Wardani[1*], Ismit Mado[1], Achmad Budiman[1], Sugeng Riyanto[1], Ghusaebi[2], Rustam Effendy[2]
[1]Department of Electrical Engineering, Universitas Borneo Tarakan, Jl. Amal Lama No.1, Tarakan, 77115, Indonesia
[2]PT.PLN (Persero) Kota Tarakan, Jl. P. Diponegoro No.18 Kelurahan Sebengkok, Kota Tarakan, 77114, Indonesia

Abstract
Electrical energy is a fundamental necessity and constitutes one of the most critical needs for society. Since electricity demand over a specific period cannot be determined with absolute certainty, PT. PLN (Persero) Tarakan, as the energy provider, must accurately predict daily load requirements. This study applies Time Series and Linear Regression methods. The optimal result using the ARIMA(1,1,0) model yielded a MAPE value of 14.68%. By incorporating seasonal components, the best model obtained was SARIMA(1,1,0)(1,0,0) with a MAPE of 6.13%. Meanwhile, load forecasting using the Linear Regression method resulted in the equation Y = 44,217.8 - 6.9X with a MAPE of 7.87%.

Keywords:
Linear Regression, Electrical Load, Time Series

Article History:

Corresponding Author:
Kartika Putri Wardani
Department of Electrical Engineering, Universitas Pertamina, Indonesia
Email:: kartikap.w@icloud.com

1. Introduction

Electrical energy is indispensable to modern society. Human activities aimed at fulfilling daily needs constitute a primary factor driving the increasing demand for electrical energy. Consequently, the electricity supply sector must anticipate this trend in terms of both future quantity and capacity. The magnitude of electricity demand over a specific period cannot be calculated with absolute certainty, leading to the challenge of operating power generation systems continuously to meet real-time demand. Conversely, if the power supplied significantly exceeds load demand, it results in the wastage of generation costs. On the other hand, if the generated power is insufficient to meet consumer needs, it leads to power outages. Therefore, efforts to predict electrical load demand are essential as a foundation for both operational planning and the future development of electric power systems.

The application of short-term electrical load forecasting based on Linear Regression and Time Series models at *PT. PLN (Persero)* in Tarakan is expected to assist the company in achieving more effective energy planning and management. With more accurate predictions, *PT. PLN* can optimize power plant operations, mitigate the risk of outages, and enhance energy usage efficiency. This not only positively impacts the continuity of electricity service in Tarakan but also contributes to the overall operational performance of *PLN*.

This study performs short-term electrical load forecasting modeling based on power generation data at *PT. PLN (Persero) Tarakan*. The data utilized in this research consists of hourly electrical load data from November 1 to November 15, 2023.

2. Experimental Section

A. Short-Term Electrical Load Forecasting

Forecasting is the process of estimating future events based on relevant historical data and projecting them forward using mathematical models. To achieve forecasting objectives, it is essential to employ methods compatible with the specific data and information being analyzed. This is particularly critical in electric power system operations, which hold significant importance for both corporate management and operational efficiency. Short-term forecasting is defined as the prediction of conditions over a time horizon ranging from daily to hourly intervals. The objective of short-term forecasting in this context is to serve as a comparative study between forecasted and actual electrical loads.

1. *ARIMA Modeling*

The ARIMA method is also known as the Box-Jenkins method, a model extensively developed by George Box and Gwilym Jenkins in 1970. However, the Box-Jenkins method continues to dominate many research fields to this day. ARIMA utilizes two algorithms, namely autoregressive (AR) and moving average (MA), and combines an integrated element to address non-stationarity through a differencing method. A key prerequisite for stationarity is that the data pattern must not exhibit a significant trend. To identify AR properties, the Autocorrelation Function (ACF) is used, which represents the relationship between a series of observations in a time series. Conversely, the Partial Autocorrelation Function (PACF) is employed to identify MA properties. Generally, significant ACF and PACF values are observed at lag 1 or 2; it is rarely found that AR and MA properties possess values greater than 2.

ARIMA can be applied to data regardless of whether it exhibits seasonal patterns. In general, if an ARIMA model has an AR order of p, a degree of differencing d, and an MA order of q, the model is denoted as ARIMA(p,d,q), expressed by the following equation:

$$Z_t = (1+\phi_1)Z_{t-1} + (\phi_2 - \phi_1)Z_{t-2} + \cdots + (\phi_p - \phi_{p-1})Z_{t-p} + \phi_p Z_{t-p-1} + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} \quad (1)$$

or

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t \quad (2)$$

where

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$
$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q.$$

A stationary time series possesses an Autocorrelation Function (ACF) that declines linearly and slowly. The same applies to the estimated ACF derived from the data. If there is a tendency for the ACF and the estimated ACF (f_k) not to decay rapidly, the function is classified as non-stationary.

The seasonal time series is described as follows:

Seasonal Autoregressive Moving Average (SARIMA) Model

The general form of SARIMA is

$$\phi_p(B)\Phi_{P_1}(B^{s_1})(1-B)^d(1-B^{s_1})^{D_1}Z_t = \theta_0 + \theta_q(B)\Theta_{Q_1}(B^{s_1})a_t \quad (3)$$

where

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$
$$\Phi_P(B^s) = 1 - \Phi_{1_1} B^{s_1} - \Phi_{2_1} B^{2s_1} - \cdots - \Phi_{P_1} B^{P_1 s_1}$$

$$\theta_q(B) \quad = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

$$\Psi_{Q_1}(B^{s_1}) = 1 - \Theta_{1_1} B^{s_1} - \Theta_{2_1} B^{2s_1} - \cdots - \Theta_{Q_1} B^{Q_1 s_1}$$

1) Box-Jenkins ARIMA Model Procedure

Data analysis was performed using the ARIMA method assisted by statistical software, specifically MINITAB 14. The sequential steps for applying the ARIMA method are as follows:

a. Data preparation, including stationarity checks.
b. Model identification. Through ACF and PACF plots, the appropriate model for ARIMA prediction can be determined.
c. Determination of ARIMA parameters p, d, and q.
d. Formulation of the ARIMA model equation. The coefficients utilized are derived from model analysis parameters yielding the lowest Mean Squared Error (MSE) value.
e. Validation of predictive parameters.
f. Forecasting.

The subsequent step involves utilizing the optimal model for prediction. Once the best model has been determined, it is ready to be employed for daily electrical load forecasting at *PT. PLN (Persero) Tarakan*.

2. Linear Regression Modeling

The linear regression model is a statistical framework used to understand and predict the relationship between a dependent variable (target) and one or more independent variables (predictors).

The relationship between the dependent and independent variables can be expressed as a function. The simple linear regression equation is defined as follows:

$$Y = \alpha + bX \tag{4}$$

Where:

Y = Dependent Variable

X = Independent Variable

α = Constant (Intercept)

b = Magnitude of the response induced by the predictor (Slope)

To obtain Equation (4), the initial step is to determine the constant and the regression coefficient. The formulas used to calculate these variables are as follows:

$$\alpha = \frac{\sum y - b\sum x}{n} \tag{5}$$

$$b = \frac{n\sum(xy) - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \tag{6}$$

3. Forecasting Model Evaluation

Forecasting accuracy is essential in data analysis to evaluate the suitability of the utilized forecasting models. In this study, the criterion employed to evaluate forecasting accuracy is the Mean Absolute Percentage Error (MAPE). The equation for calculating MAPE is as follows:

$$MAPE = \sum_{t=1}^{n} \left(\frac{Yreal - Yforcast}{Yforcast}\right) \times 100\% \tag{7}$$

 p-ISSN: 2963-8577
e-ISSN: 2964-3511

3. Result and Discussion

The data utilized in this study is secondary data, specifically electrical load data obtained from *PT. PLN (Persero) Tarakan* spanning from November 1 to November 15, 2023. This dataset consists of hourly measurements. The training set comprises 336 data points covering the period from August 1 to August 14, 2023, while the testing set consists of 24 data points recorded on November 15, 2023.

A. *Electrical Load Forecasting Study Based on Time Series Models*

1. *Data Stationarity*

The initial step is to assess the stationarity of the electrical load data using a time series plot. Based on visual observation, the data exhibits trend components, as it tends to fluctuate (increase and decrease) over time. Consequently, the data is considered non-stationary. Figure 1 shows the plot after data differencing.

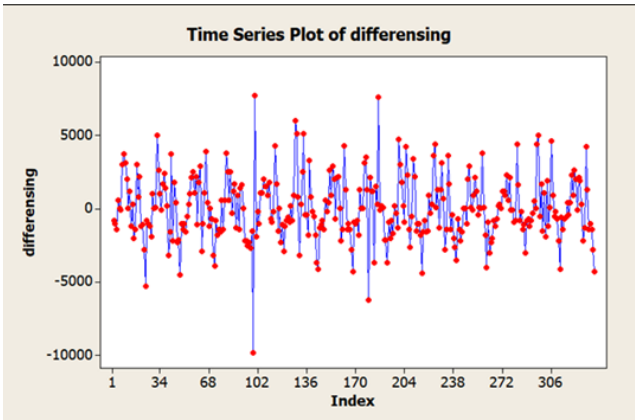

Figure 1. Time series plot after differencing.

Meanwhile, the ACF plot after differencing is presented in Figure 2.

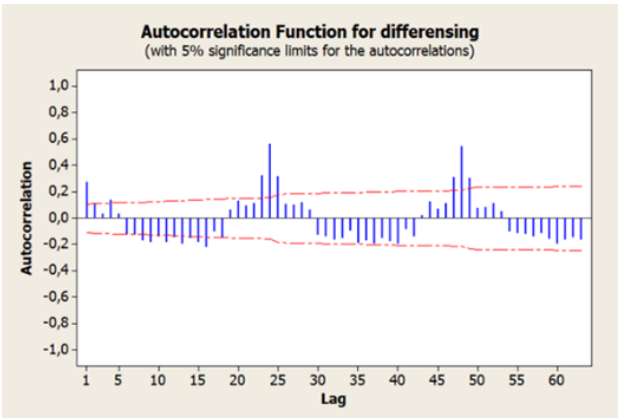

Figure 2. Autocorrelation Function (ACF) plot after differencing.

In addition to the time series plot, data stationarity can also be assessed using the ACF plot. The initial ACF plot indicates that the data is non-stationary, as it decays slowly and exhibits high correlation at early lags. Based on this identification, differencing is required. Figure 2 demonstrates that the data has become stationary, as the correlations at various lags rapidly approach zero, indicating the absence of any remaining trend patterns. The PACF plot is also necessary for model analysis, as presented in Figure 3.

p-ISSN: 2963-8577
e-ISSN: 2964-3511

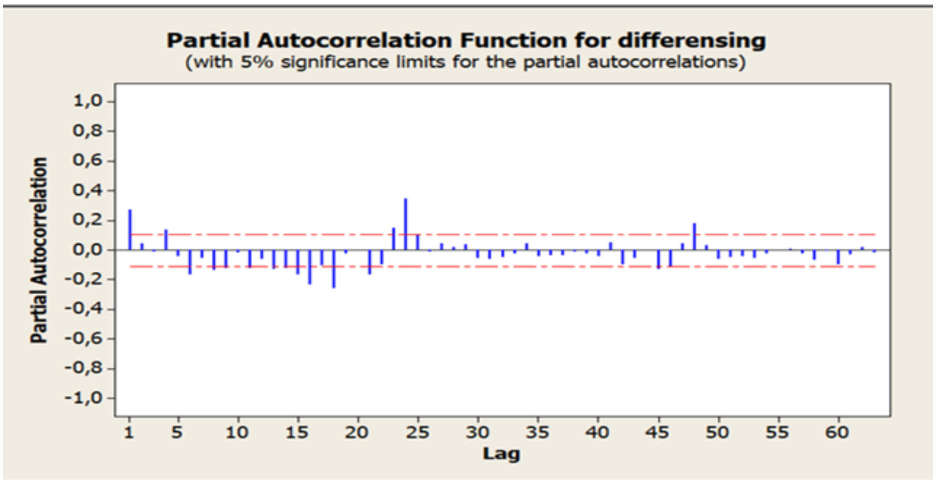

Figure 3. Partial Autocorrelation Function (PACF) plot after differencing.

The criteria used for model determination are as follows:

a. If the ACF exhibits a "dying down" (tailing off) pattern and the PACF shows a "cut off," the ARIMA model is identified as a pure Autoregressive (AR) process.

b. If the ACF shows a "cut off" and the PACF exhibits a "dying down" pattern, the ARIMA model is identified as a pure Moving Average (MA) process.

c. If both the ACF and PACF exhibit "dying down" patterns, the ARIMA model is considered a mixed AR and MA process.

The ACF plot in Figure 2 demonstrates a dying down pattern, characterized by lags that gradually decrease towards zero. Additionally, the T-value at the first lag is 5.08, which slowly declines and approaches zero. With significant spikes observed initially at lags 1 and 3, the parameter p can be determined to be between 1 and 2. Consequently, it can be concluded that $p = 1$ (denoted as AR 1) or $p = 2$ (denoted as AR 2).

Meanwhile, the PACF plot in Figure 3 exhibits a "dying down" pattern, declining towards zero with insignificant T-values after the first lag; thus, the parameter q is determined to be 1. Regarding the parameter d (differencing), the value is 1, as the data achieved stationarity after the first differencing. Consequently, the model to be utilized is the ARIMA(p,d,q) model.

To validate the determination of the ARIMA model, it is necessary to evaluate several candidate models by selecting the one with the lowest Mean Squared Error (MSE) value derived from the identification results.

Table 1. ARIMA Model determination using MSE.

No	ARIMA Model	MSE
1	(1,1,1)	4403641
2	(1,1,2)	4415008
3	(1,1,0)	4400358
4	(0,1,1)	4454668
5	(0,1,2)	4399166
6	(2,1,0)	4404417

7	(2,1,1)	4414802
8	(2,1,2)	4428689

Based on the eight models evaluated above, the ARIMA(1,1,0) model yielded the lowest MSE value of 4,400,358; therefore, this model is selected for forecasting. Based on the forecasting results of the selected model, the resulting MAPE is 14.68%.

Although the plot of the differenced data is stationary, it exhibits indications of a Seasonal ARIMA (SARIMA) process. This is evidenced by the presence of multiple significant lags or repetitive significant spikes at specific intervals, occurring at lags 24, 48, and so forth. Consequently, a re-analysis incorporating the seasonal pattern is required. To understand the model's behavior, the ACF and PACF plots are interpreted with reference to Table 2 as follows.

Table 2. Seasonal ACF and PACF Patterns.

Model	ACF	PACF
AR (P)	*Dying down* (exponential decay) at seasonal lags	*Cut off* (abrupt drop)
MA (Q)	*Cut off* (abrupt drop)	*Dying down* (exponential decay) at seasonal lags
ARIMA (P,d,Q)	*Dying down* (exponential decay) at seasonal lags	*Dying down* (exponential decay) at seasonal lags

The SARIMA model is denoted as $(p,d,q)(P,D,Q)_{24}$, where the non-seasonal parameters (p,d,q) are derived from the previously determined optimal ARIMA model, specifically (1,1,0). Consequently, eight estimated SARIMA models are compared. To facilitate analysis, the MSE values for each model are presented below.

Table 3. SARIMA Model determination using MSE.

SARIMA Model	MSE
(1,1,0)(1,0,0)24	2932520
(1,1,0)(1,0,1)$_{24}$	-
(1,1,0)(0,0,1)24	3674686
(1,1,0)(2,0,0)24	-
(1,1,0)(2,0,1)24	-
(1,1,0)(2,0,2)24	-
(1,1,0)(0,0,2)$_{24}$	3194503

(1,1,0)(1,0,2)24	3545555

Berdasarkan Among the eight SARIMA models evaluated above, the model with the lowest MSE is the SARIMA(1,1,0)(1,0,0)$_{24}$ model, with an MSE value of 2,932,520. Consequently, this model is selected for forecasting. Based on the forecasting results of the selected model, the Mean Absolute Percentage Error (MAPE) is 6.136030879%, or approximately 6.13%. When comparing the MAPE values of the seasonal and non-seasonal models, the seasonal model demonstrates superior performance, as indicated by its lower MAPE value compared to the non-seasonal counterpart.

B. Short-Term Electrical Load Forecasting Study Based on Linear Regression Models

1. Determination of Normal Equation Coefficients

To obtain the coefficients for the linear regression equation, the normal equations must be constructed. Prior to forming these equations, the value of each element is calculated by organizing the variables as shown in the following table.

Table 4. Determination of normal equation elements.

N	X	Y	Xy	x^2
1	1	40.600	40.600	1
2	2	39.800	79.600	4
.	.	.	.	.
.	.	.	.	.
335	335	45.700	15.309.500	112.225
336	336	41.400	13.910.400	112.896
n = 336	Σx= 56.616	Σy = 14.463.700	Σxy = 2.415.155.500	$Σx^2$ = 12.700.856

Subsequently, the values of coefficients *α* and *b* are determined using Equations 5 and 6, yielding:

$$b = \frac{336(2.415.155.500) - 56.616\ .14.463.700}{336(12.700.856) - (56.616)^2}$$

$$b = -6,95$$

Subsequently, to determine the value of coefficient *α* using the equation above, yielding:

$$\alpha = \frac{14.463.700 - (-6,95).56.616}{336}$$

$$\alpha = 44.217,8$$

2. *Forecasting*

Calculating the short-term electrical load forecast using the linear regression model via Equation 4 yields:

$Y = \alpha + b \cdot X$

$= 44.217,8 + (-6,95) \cdot 337$

$= 41.875,65$

Subsequently, the comparison between the actual data (specifically from November 15, 2023) and the results obtained from the linear regression model is presented in a tabulated format in Table 5 below.

Table 5. Overall results of $|Y_{real} - Y_{forecast}|$ values.

No	Date	Time	X	Y_{real}	$Y_{forecast}$	$\|Y_{real} - Y_{forecast}\|$
1	15-Nov-23	01.00	337	40.000	41.875,65	-1.875,65
2	15-Nov-23	02.00	338	38.300	41.868,7	-3.562,70
3	15-Nov-23	03.00	339	36.900	41.861,75	-4.961,75
4	15-Nov-23	04.00	340	35.400	41.854,8	-6.454,80
5	15-Nov-23	05.00	341	36.300	41.847,85	-5.547,85
6	15-Nov-23	06.00	342	37.000	41.840,9	-4.840,90
7	15-Nov-23	07.00	343	37.700	41.833,95	-4.133,95
8	15-Nov-23	08.00	344	41.900	41.827	73
9	15-Nov-23	09.00	345	45.600	41.820,05	3.779,95
10	15-Nov-23	10.00	346	46.700	41.813,1	4.886,90
11	15-Nov-23	11.00	347	47.600	41.806,15	5.793,85
12	15-Nov-23	12.00	348	48.500	41.799,2	6.720,80
13	15-Nov-23	13.00	349	50.600	41.792,25	8.807,75
14	15-Nov-23	14.00	350	51.200	41.785,3	9.414,70
15	15-Nov-23	15.00	351	51.700	41.778,35	9.921,65
16	15-Nov-23	16.00	352	48.300	41.771,4	6.528,60
17	15-Nov-23	17.00	353	47.500	41.764,45	5.735,55
18	15-Nov-23	18.00	354	51.900	41.757,5	10.142,50

19	15-Nov-23	19.00	355	52.900	41.750,55	11.149,45
20	15-Nov-23	20.00	356	51.700	41.743,6	9.956,40
21	15-Nov-23	21.00	357	50.600	41.736,65	8.863,35
22	15-Nov-23	22.00	358	48.800	41.729,7	7.070,30
23	15-Nov-23	23.00	359	46.900	41.722,75	5,177,25
24	15-Nov-23	00.00	360	43.200	41.715,8	1.484,20
	Total		8364	1.087.200	1.003.097,4	78.951,35

3. *MAPE Accuracy Test*

After obtaining the absolute values, the MAPE accuracy test for the short-term electrical load on November 15, 2023, is performed as follows:

$$MAPE = \sum_{t=1}^{n} \left(\frac{78.951,35}{1.003.097,4}\right) \times 100\%$$

= *7,87075612*

= *7,87%*

Based on the accuracy testing conducted using MAPE, an error value of 7.87% was obtained.

4. Conclusion

Following the forecasting analysis conducted on electrical load data from PT PLN (Persero) Tarakan utilizing both time series and linear regression methods, the following conclusions are drawn:

a. The MAPE for the ARIMA model is 14.67% and for the SARIMA model is 6.13%. These results generally fall within the 10–20% range, indicating "Good" forecasting accuracy.
b. The MAPE for the Linear Regression model is 7.87%, which is also interpreted as "Good".
c. Forecasting using the SARIMA model proved to be the best approach, demonstrating the smallest error margin among the three methods utilized.

Acknowledgment

The authors express their gratitude for the collaboration between the Department of Electrical Engineering at Universitas Borneo Tarakan and PT PLN Kota Tarakan, which serves as a manifestation of the synergy between academia and practitioners.

References

[1]. Putra, A A. & Budiono, H. (2017). Peramalan Beban Listrik dengan Metode Time Series. Jurnal Teknik Elektro, Vol 9 No 2, 58-65.

[2]. Dewi, I. A. M. C., Nrartha, M. A., & Suksmadana, M. B. (2016). Peramalan Beban Listrik Jangka Pendek pada Sistem Kelistrikan Jawa Timur dan Bali Menggunakan Fuzzy Time Series. Jurnal Ampere Vol 3 No 1, 1–6.

p-ISSN: 2963-8577
e-ISSN: 2964-3511

[3]. Aswathappa. (2005). Time Series Forecasting for Financial Applications. Jurnal of Applied Business Research. Vol 21 No 4, 115-130 .

[4]. Khair, Aulia. (2011). Peramalan Beban Listrik Jangka Pendek Menggunakan Kombinasi Metode Autoregresif Integrated Moving Avarage (ARIMA) Dengan Regresi Linier Antara Suhu Dan Daya Listrik". Universitas Indonesia.

[5]. Helmi Wibowo, Yadi Mulyadi, Ade Gafar Abdullah. (2012). Peramalan Beban Listrik Jangka Pendek Terklasifikasi Berbasis Metode Autoregressive Integrated Moving Average. Jurnal, Electrans, Vol 11 NO.2, 44-50.

[6]. Juliasari, Yossi. (2020). Peramalan Beban Listrik Jangka Pendek Menggunakan Metode Regresi Linier-Algoritma Genetika. Universitas Hasanuddin.

[7]. Hartono, D. (2021). Penerapan Model Regresi Linier untuk Peramalan Beban Listrik. Jurnal Elektronika Indonesia.

[8]. Haryanto, R. (2019). Analisis Peramalan Beban Listrik Menggunakan Model ARIMA pada PT. PLN (Persero). Prosiding Seminar Nasional Teknik Elektro, 243-250. Surabaya: Universitas X.

[9]. I Mado, A Soeprijanto, S Suhartono. (2018). Applying of Double Seasonal ARIMA Model for Electrical Power Demand Forecasting at PT. PLN Gresik Indonesia. International Journal of Electrical and Computer Engineering 8 (6), 4892.

[10]. I Mado, A Rajagukguk, A Triwiyatno, A Fadllullah. (2014). Short-Term Electricity Load Forecasting Model Based DSARIMA. International Journal of Electrical, Energy and Power System Engineering 5(1), 6-11.

[11]. I Mado, A Budiman, A Triwiyatno. (2023). Short-Term Forecasting Daily Electricity Loads Using Seasonal Arima Patterns of Generation Units at PT. PLN (PERSERO) Tarakan City. Jurnal Ilmiah Kursor 12 (2), 93-100.

[12]. I Mado. (2020). Electric Load Forecasting an Application of Cluster Models Based on Double Seasonal Pattern Time Series Analysis. Forecasting in Mathematics-Recent Advances, New Perspectives and Applications

Biographies of Authors

Kartika Putri Wardani is an Electrical Engineering Student at the University of Borneo Tarakan in 2019. The author was born in Tarakan on April 07 1999. Before becoming a student, the author had completed his education at SMA Negeri 1 Pulau Bunyu.

Ismit Mado graduated from the ITS Surabaya Doctoral Program in 2019. Currently he serves as Head of the Power System Stability Laboratory at the Department of Electrical Engineering, University of Borneo, Tarakan. He is interested in stability and control studies of power generation systems, forecasting studies based on time series models, and fuzzy modeling.

Achmad Budiman is an Electrical Engineering Lecturer since 2002. He obtained his Bachelor's degree in Electrical Engineering from Gadjah Mada University, Yogyakarta. He obtained his Master's degree in Electrical Engineering from ITS Surabaya. Currently, he serves as the Head of Electrical Engineering Study Program, Borneo Tarakan University. He is interested in the field of stability in power distribution and transmission systems.

Sugeng Riyanto is an Electrical Engineering Lecturer since 2002. Currently, he serves as a lecturer in the Electrical Engineering Study Program, University of Borneo Tarakan. He is interested in the fields of drawing electrical engineering, basic electrical installations and transformers.

Ghusaebi graduated with a Bachelor's degree in Electrical Engineering from the University of Indonesia in 2015. He has been working at PT PLN (Persero) since 2016, specializing in power distribution. Currently, he serves as the Manager of PT PLN (Persero) ULP Tarakan.

Rustam Effendy Graduated from Hang Tuah Tarakan High School in 1998. He has been working at PT PLN (Persero) since 2010. Currently, he serves as a Junior Load Regulator Technician at PT PLN (Persero) ULP Tarakan.