Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 8, No. 2, May, Pp. 571-580

Forecasting model for lighting electricity load with a limited dataset using XGBoost

Maman Abdurohman*[1], Aji Gautama Putrada[2]
School of Computing, Telkom University, Indonesia[1]
Advanced and Creative Networks Research Center, Telkom University, Indonesia[2]

Article Info
Keywords:
Extreme Gradient Boosting, Electricity Load, Forecasting, COVID-19, Limited Dataset.

Article history:
Received: February 22, 2023
Accepted: May 22, 2023
Published: May 31, 2023

Cite:
M. . Abdurohman and A. G. Putrada, "Forecasting Model for Lighting Electricity Load with a Limited Dataset using XGBoost", KINETIK, vol. 8, no. 2, May 2023. https://doi.org/10.22219/kinetik.v8i2.1687

* Corresponding author.
Maman Abdurohman
E-mail address:
abdurohman@telkomuniversity.ac.id

Abstract
Energy forecasting is an important application of machine learning in renewable energy because it is used for operational, management, and planning purposes. However, using the electricity load dataset during COVID-19 is a research challenge in the forecasting model due to the limited dataset and non-stationarity. This paper proposes an extreme gradient boosting (XGBoost) forecasting model for a limited dataset. Forecasting models require large amounts of data to create high-accuracy models. We conduct research using the PT Biofarma office electricity usage dataset for eight months during the COVID-19 period. Because office activities were limited during the pandemic, the datasets obtained were few. Several methods are used for modeling limited datasets, namely XGBoost, multi-layer perceptron (MLP), autoregressive integrated moving average (ARIMA), and long short-term memory (LSTM). We have conducted several experiments using a limited dataset with these four methods. The test results with the t-test show that the electricity load data for work-from-office (WFO) and work-from-home (WFH) periods have a significant average difference. Then the test results with the augmented Dickey–Fuller (ADF) test show that our data is non-stationary. Compared to the benchmark method, the XGBoost method shows the best forecasting performance with mean absolute percentage error (MAPE), root mean squared error (RMSE), mean absolute error (MAE), and R^2 of 0.48, 5.00, 3.09, and 0.61 respectively.

1. Introduction

Indonesia sparked the development of a green economy where one of the steps is renewable energy for electricity production [1]. In this case, using machine learning for predictions in energy production becomes a vital component [2]. Energy forecasting is an important application of machine learning in renewable energy because it is used for operational, management, and planning purposes [3]. Several studies have explored energy forecasting for several different reasons. Shao *et al.* [4] researched predicting electricity usage for pricing applications. Robinson *et al.* [5] created a model to estimate expenses due to electricity use. Arvanitidis *et al.* [6] perform electricity load prediction for marketing decision-making.

Several previous studies have used several different methods for electricity load forecasting. Askari *et al.* [7] mentioned that electricity load data is usually non-stationary, so it becomes challenging to forecast. This research used multi-layer perceptron (MLP) forecasting and achieved the best mean absolute percentage error (MAPE) of 1.38%. Nepal *et al.* [8] used electricity load monitoring in order to prevent excess electricity usage in the future. This method used an autoregressive-integrated moving average (ARIMA) for forecasting and achieved the best MAPE of 3.80%. Alden *et al.* [9] said that the need for electricity load forecasting grows along with the increasingly complex smart grid and the conditions of consumer housing. This study used long short-term memory (LSTM) forecasting and concluded that the best performance of LSTM is for predictions with an hourly horizon.

Donnat *et al.* [10] said that the exponential property in the COVID-19 data made datasets taken during the pandemic difficult to predict. Pane *et al.* [11] also said that, in the economic sector, COVID-19 affects the ability of machine learning to predict the customer price index (CPI). Chandra *et al.* [12] also stated that the limited dataset during COVID-19 presented a challenge in forecasting models. Hence, using electricity load datasets during COVID-19 is a research challenge in forecasting models.

Extreme gradient boosting (XGBoost) is a type of ensemble learning and can also be used for forecasting [13]. Zhao *et al.* [14] said that deep learning methods such as LSTM need to improve forecasting, which requires many data for training. They propose using XGBoost for short datasets with hourly timeframes. Li *et al.* [15] explained that methods such as ARIMA are weak when the data is non-stationary. This research used the XGBoost hybrid method and produced a MAPE of 0.57%. Xue *et al.* [16] said that the application of electricity load forecasting on electric vehicles could anticipate the supply of electric fuel in the future. They claim to use the XGBoost hybrid method for the first time for forecasting electric refueling. There is a research opportunity to make XGBoost a forecasting method in the limited dataset collected during the pandemic.

Our research aim is to forecast lighting electricity load in offices during COVID-19 with XGboost. Our first step is to collect the lighting electricity load dataset from PT Biofarma's office for eight months. Then we apply autocorrelation analysis to the dataset, including seasonal decomposition, the autocorrelation function (ACF) method, the partial autocorrelation function (PACF), and the augmented Dickey–Fuller (ADF) test. We then implement the XGBoost model, among others, by optimizing the test length and lag parameters. Finally, we benchmarked our proposed model with three state-of-the-art methods: MLP, ARIMA, and LSTM.

To the best of our knowledge, there has never been a study that has applied XGBoost in electricity load forecasting during COVID-19. Here are our contributions:

- A forecasting model using the XGBoost method, which has good performance on non-stationary data obtained during WFO
- A novel limited dataset regarding electricity load forecasting from PT Biofarma
- An electricity load forecasting model that has good performance on limited datasets

We arrange the remainder of our paper with the following systematics: Chapter 2 contains the research design. Chapter 3 reports the results of our tests and discusses them concerning state-of-the-art papers. Chapter 4 is the conclusion of our research.

2. Research Method

We propose a methodology for doing research with our proposed method. Our first step is to collect the lighting electricity load dataset from PT Biofarma's office for eight months. Then we apply autocorrelation analysis to the dataset using seasonal decomposition, ACF, PACF, and ADF test methods. We then implement the XGBoost model, then optimize the model by the test length and lag parameters. Finally, we benchmarked our proposed model with three state-of-the-art methods: MLP, ARIMA, and LSTM. Figure 1 is a flow chart showing our proposed methodology.

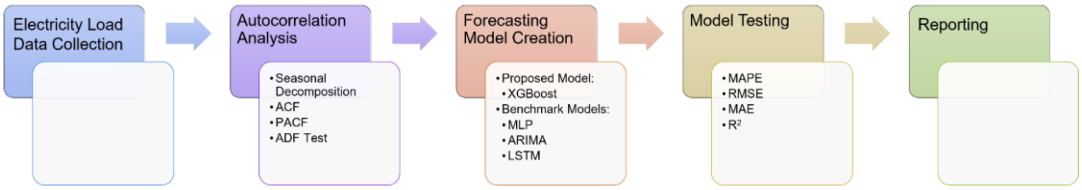


Figure 1. Our Proposed Research Methodology

2.1 Office Lighting Electricity Load Dataset

PT Biofarma is a state-owned enterprise (BUMN) whose core business is vaccine production [17]. PT Biofarma also produces and distributes COVID-19 vaccines such as Pfizer, Astra Zeneca, Moderna, and Sinovac. We measured the lighting system's power on the fifth floor of a building at PT Biofarma for eight months: from January 1, 2020, to August 31, 2020 [18]. On March 16, 2020, Indonesia started implementing work-from-home (WFH), resulting from the entrance of COVID-19 into the national territory [19].

We used Eastron SDM630-Modbus to measure the lighting system [20]. We mounted the tool on the power distribution box on the fifth floor [21]. The output of the measurement tool is an excel file (.xlsx) with attributes, namely date and running total of energy each day. The unit of energy is in kilo Watt hour (kWh). The sampling period of the data collected is one day for one data. Figure 2 shows the implementation and installation of our load data logger.

(a) (b) (c)

Figure 2. The Implementation of the Electricity Load Logger: (a) The Device (b) The Measurement Environment (c) The Device Installed and Running

At the end of the measurement, the data collected was 244 data. We get the daily energy ($e(x)$) from running total energy by the following Equation 1.

$$e(x) = \frac{E(x) - \sum_{t=0}^{x-1} e(t)dt}{\mathrm{d}t} \tag{1}$$

Where x is the index data time series, $E(x)$ is the total running energy in the time series x, and $\mathrm{d}t$ is the smallest unit of time, in our case it is one day [22]. Figure 3 shows the measurement results for eight months.

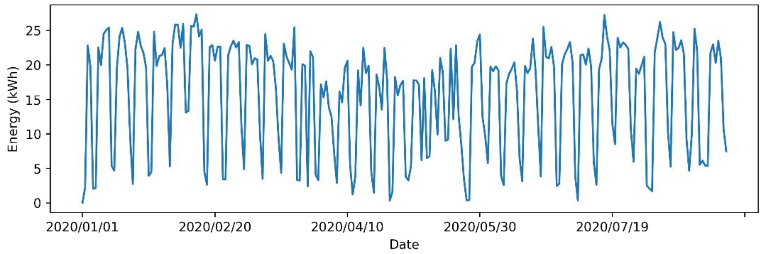


Figure 3. The Energy Data Gathered from the Load Logging Process

Before the next stage, namely modeling, the dataset went through several stages of analysis, including seasonal decomposition, ACF, PACF, and ADF. Seasonal decomposition divides a signal into seasonal, trend, and residual [23]. In seasonal decomposition, a function Y_v is written with the following Equation 2.

$$Y_v = T_v + S_v + R_v \tag{2}$$

where T_v is a trend function, S_v is a seasonal function, and R_v is a residual function [24].

The first step in obtaining these three functions is to obtain T_v using the locally estimated scatterplot smoothing (LOESS) algorithm [25]. At level 0, the LOESS function resembles the weighted moving average function. After the algorithm gets T_v, the next step is to get C_v by $Y_v - T_v$. Then C_v through the moving average again to get a smooth signal [26]. The result is L_v which is the result of the low pass filter from C_v. The following equation is $S_v = C_v - L_v$. Lastly is the equation $R_v = Y_v - T_v + S_v$.

ACF is the correlation of a signal with the signal itself with a certain lag, then removes other values at lower lags [27]. The ACF (R) Equation 3 is presented.

$$R_{XX}(T_1, T_2) = E\left[X_{T_1} \bar{X}_{T_2}\right] \tag{3}$$

Where X is a signal, T_1 is lag 0, T_2 is the next lag, X_{T_1} is the value of X in T_1, $\bar{X}_{T_2}$ is the value of the complex conjugation of X in T_2, and finally E is the expected value function.

Meanwhile, PACF is the partial correlation between a signal and itself at a certain lag, where the signal is stationary [28]. The formula for PACF (ϕ) is as follows Equation 4 and Equation 5.

$$\phi_{XX}(T_1, T_2) = corr(X_{T_1}, X_{T_2}), for\ T_2 = T_1 \tag{4}$$

$$\phi_{XX}(T_1, T_2) = corr(X_{T_1} - \hat{X}_{T_1}, X_{T_2} - \hat{X}_{T_2}), for\ T_2 \neq T_1 \tag{5}$$

Where $\hat{X}$ is the linear combination of all values between X_{T_1} and X_{T_2}.

ADF test is a test to measure the presence or absence of stationarity in a signal. The ADF test Equation 6 is presented.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \ldots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_1 \tag{6}$$

Where Δ is the first difference operator, p is the lag, δ is the coefficient minus one, α is a constant, \beta is the β of the time trend, and ε_1 is the error. The null hypothesis is when $\gamma = 0$ while $\gamma < 0$ is for the alternate hypothesis.

2.2 XGBoost Forecasting

Autoregression is a special regression in which the predicted output is based on the previous values in a stochastic function [29]. With these capabilities, autoregression is a useful method for forecasting [30]. The general formula for autoregression for the x_t output is as follows Equation 7.

$$x_t = \sum_{i=1}^{p} \varphi_i x_{t-i} + \varepsilon_t \tag{7}$$

Where p is the lag and φ_p is the model parameter. With this general formula, all machine learning methods that apply to regression are capable of forecasting, including XGBoost.

XGBoost is a boosting-type learning ensemble that uses the Newton–Raphson equation as its gradient method [31]. Like gradient boosting, XGBoost performs learning iterations with a weak learner where in each iteration, the model tries to reduce errors in misclassified data with a gradient function [32]. The essence of the Newton–Raphson method is to find a slope with a certain formula that utilizes the tangent line at a random point on a curve and its intersection with the x-axis [33]. Suppose a tangent line at point $x = x_n$ intersects the x-axis at point x_{n+1}, then getting that point from the slope Equation 8 is presented.

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{8}$$

Where $f(x_n)$ is the loss function, and $f'(x_n)$ is the slope formula.

2.3 Benchmark Methods and Performance Metrics

We benchmark our proposed method with three state-of-the-art methods, namely MLP [7], LSTM [8], and ARIMA [9]. MLP, or any feedforward artificial neural network (ANN), is a graph that resembles a sentient being's neuron network [34]. Each input and output from a node in one of the layers in the graph forwards information to the node in the next layer with a non-linear function [35]. The training process forms each of these non-linear functions iteratively through many epochs, where each epoch fixes the weight and bias of each equation based on its error value with a gradient function [36]. So, the output equation form of each node (N_i) is as follows Equation 9.

$$N_{ij} = \sum_{k \in K} w_{ijk} x_k + \beta_{ij} \tag{9}$$

Where i is the index layer, j is the index neuron, K is the number of neurons in one layer, $x_k = N_{(i-1)j}$ is the output of the previous layer, w is the weight assigned to each x, β is the bias assigned to each neuron. The training process gets w and β [37].

LSTM is an extension of the recurrent neural network (RNN), where the RNN is a neural network with memory that can be used for sequential data [38]. The advantage of LSTM is its adjustment to long-term and short-term patterns, so there is no need to determine a precise lag to get optimum results [39]. However, deep learning methods such as LSTM need to improve forecasting, which requires many data for training [14]. The four gates in one LSTM node are the forget gate, input gate, output gate, and cell gate [40]. The formula for each gate is as follows Equation 10, Equation 11, Equation 12, Equation 13, and Equation 14.

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \tag{10}$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \tag{11}$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \tag{12}$$

$$\tilde{c}_t = \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \tag{13}$$

$$c_t = f_t \odot c_t + i_t \odot \tilde{c}_t \tag{14}$$

Where f_t is the forget gate, i_t is the input gate, o_t is the output gate, c_t is the cell gate, x_t is the input vector, and h_{t-1} is the hidden state of the previous time step. Then W, U, and b are the weight and biases for each variable. Finally $\tilde{c}_t$ is the input activation function, σ_g is the sigmoid activation function, and σ_c is the tangent activation function [41].

ARIMA is a combination of autoregression (AR), integral (I), and moving average (MA) [42]. Autoregression in ARIMA predicts the future based on the previous series of values. The integral part transforms non-stationary data into stationary data. The moving average reduces errors whose form is a linear function. There are three important variables in ARIMA: p, d, and q. AR affects the value of p, which is the lag in ARIMA training. The I part affects the value of d, namely the degree of differencing. If the value d = 0, then the method is instead an ARMA method. Lastly, MA affects the value of q, which is the window value in the moving average.

For test metrics, we use MAPE, RMSE, MAE, and R^2 as presented in Equation 15, Equation 16, Equation 17, and Equation 18.

$$MAPE = \frac{1}{N}\sum_{t=0}^{N}\frac{|x_t' - x_t|}{xt} \times 100\% \tag{15}$$

$$RMSE = \frac{1}{N}\sum_{t=0}^{N}\sqrt{(x_t' - x_t)^2} \tag{16}$$

$$MAE = \frac{1}{N}\sum_{t=0}^{N}|x_t' - x_t| \tag{17}$$

$$R^2 = 1 - \frac{\sum_{t=0}^{N}(x_t' - x_t)^2}{\sum_{t=0}^{N}(\bar{x}_t - x_t)^2} \tag{18}$$

Where x_t is the actual output, x_t' is the predicted output, $\bar{x}_t$ is the average of the actual values, and N is the amount of data in the test. Small MAPE, RMSE, and MAE values indicate good performance. Furthermore, the value of R^2 has a range of 0 to 1. A value closer to 1 indicates good model performance.

3. Results and Discussion

3.1 Results

The methodology of creating a forecasting model for electricity load, which contains the training process design, is by conducting aucorrelation analysis with four tests: seasonal decomposition, the t-test and ADF test, the ACF, and lastly the PACF. The latter tests are for determining the p, d, and q of ARIMA, respectively. Training is conducted after finalizing the hyperparameters for each model. Three metrics are determined in this process: the training size, the testing size, and the future dates size. We determine that each value are 196, 48, and 48, respectively. Lastly we evaluate and compare each model.

The first stage in our test is autocorrelation analysis, where we conduct seasonal decomposition. Figure 4 shows the results. Within this measurement range, the government implemented WFH on March 16, 2020, due to COVID-19 cases that entered Indonesia. In correlation to the mentioned date, the data has a value decrease in the trend curve compared to previous months. From July to August 2020, PT Biofarma changed the WFH rules because the company is obliged to produce the COVID-19 vaccine; they started to re-enforce work-from-office (WFO).

In summary, there are two sub-datasets. The first sub-dataset is office conditions during WFO, consisting of January 2020 – February 2020 and June 2020 – August 2020. The second sub-dataset is office conditions during WFH, March 2020 – May 2020.

We use the t-test to observe whether the decrease in energy load was significant between the months of WFH and not. The results of our t-test show that the t-statistic value is 2.70 and has a p-value <0.01. This result shows that the electricity load dataset during WFH and not WFH has a significant average difference. We perform ADF testing to determine whether our data is stationary or non-stationary. The results from the ADF show a t-distribution of -2.25 and a p-value of 0.19, which means that our electricity load data is non-stationary. Table 1 summarizes the results of the t-test and ADF test.

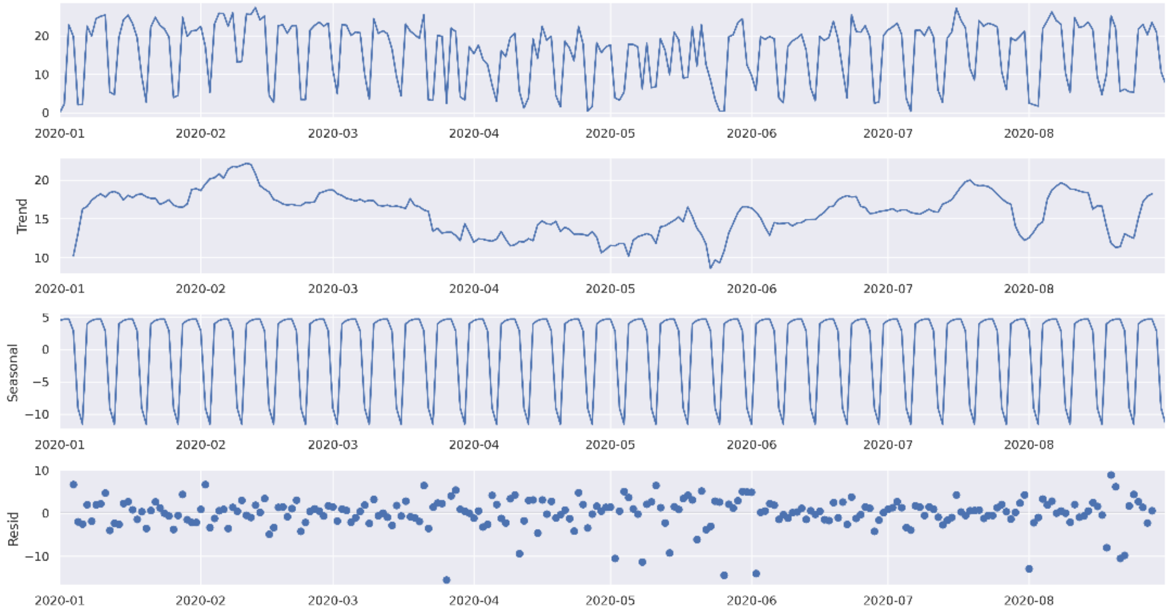
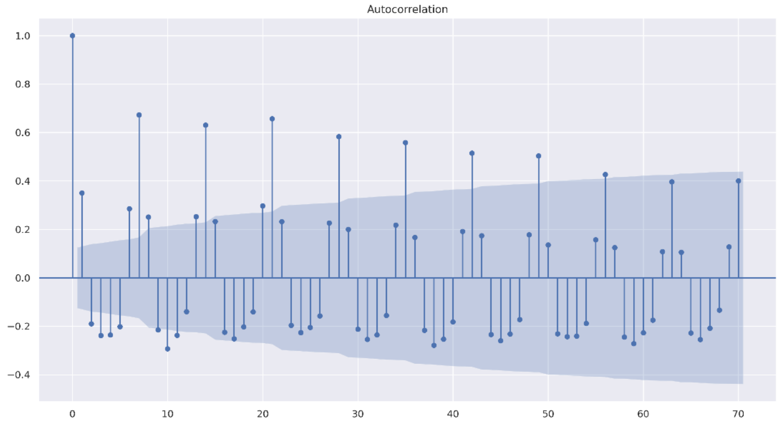


Figure 4. The Seasonal Decomposition Results

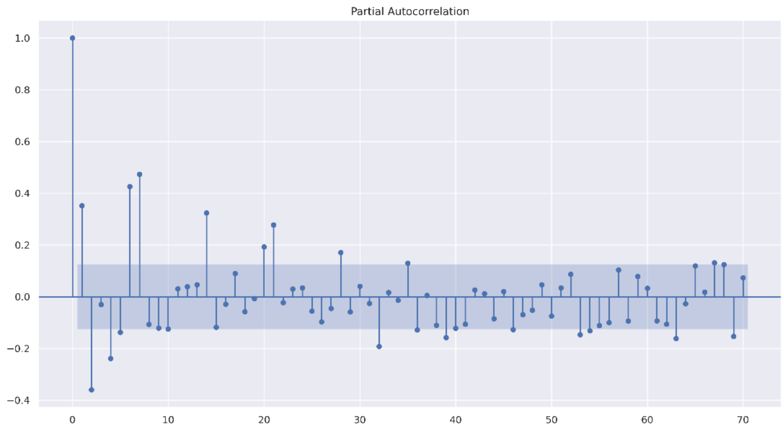
Table 1. Statistical Analysis on the ELECTRICITY LOAD DATASET

No.	Test Name	Parameters	Values
1	T-Test	T-Statistics	2.70
		P-Value	< 0.01
		Conclusion	H_0 rejected H_1 accepted
		Interpretation	The WFH sub-dataset average is significantly different compared to the WFO sub-dataset average
2	ADF Test	T-Statistics	-2.25
		P-Value	0.19
		Conclusion	H_0 accepted H_1 rejected
		Interpretation	The electricity load dataset is non-stationary

To determine the lag values and hyperparameters of the ARIMA model, we perform ACF and PACF tests. Figure 5 shows images of the two test results. Based on the results of the PACF test, the p-value for AR in ARIMA is 2. Based on the results of the ADF test, the value I for ARIMA is 1. Finally, based on the results of the ACF test, the MA value for ARIMA is 11.



(a)



(b)

Figure 5. The Autocorrelation Analysis of the Electricity Load Dataset (a) The ACF (b) The PACF

Meanwhile, because autoregressive XGBoost does not use autocorrelation analysis, we use an empirical approach. Figure 6 shows the performance of the XGboost model based on the amount of lag used in the model. The test results show that a model produces the lowest RMSE of the XGBoost model with a lag = 7. The RMSE value is 4.93. In addition, the XGBoost model at lag 7 has MAPE = 0.48 and R^2 = 0.62. This performance is related to the ACF dataset graph. Our dataset gets the highest correlation when lag = 7. XGboost gets the worst RMSE when lag = 1. We adopt the XGBoost model with lag = 7.

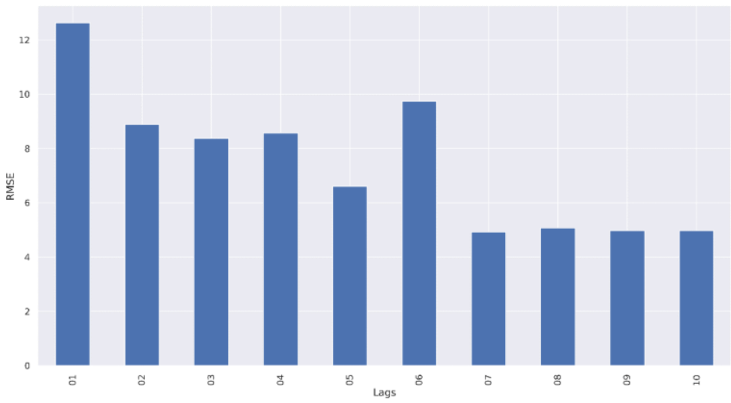


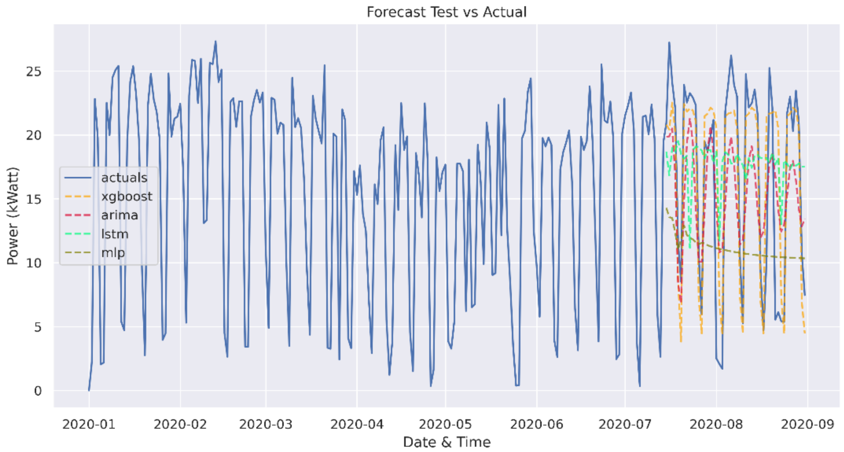
Figure 6. The Performance of the XGBoost Autoregressive Model Based on the Number of Lags

We compare XGBoost with lag = 7 against MLP, LSTM, and ARIMA. Table 2 shows the parameters we used for each model. Parameters for XGBoost, LSTM, and MLP are empirical results. At the same time, the parameters for ARIMA are the results of autocorrelation analysis using ACF, PACF, and ADF.

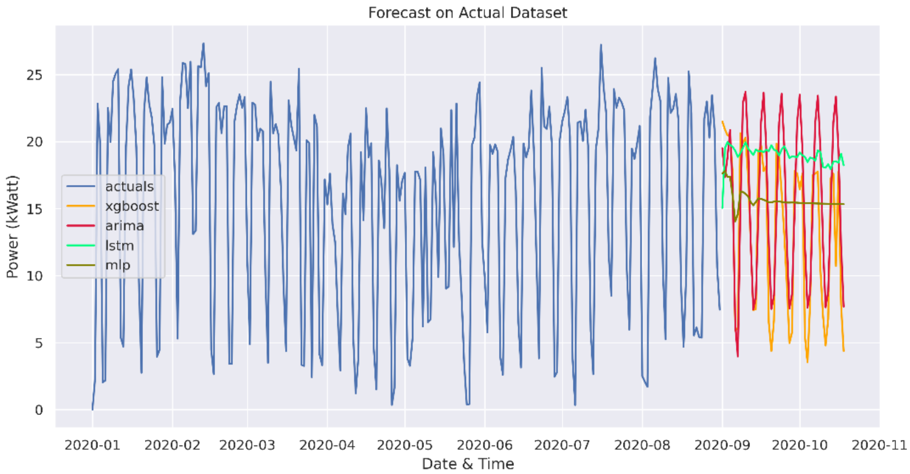
Table 2. Models and Parameters Explanation

No.	Model	Parameters	Values
1	XGBoost	Lags	7
2	ARIMA	p	2
		d	0
		q	11
3	LSTM	Lags	7
		Epochs	25
		Validation Split	0.2
		Shuffle	True
		Early Stopping	True
		Layer Size	16,16,16
		Dropout Layers	0,0,0
4	MLP	Lags	7

We first conduct a qualitative analysis to compare the four models. Figure 7.a shows a comparison of the forecasting results with the actual data. There are 244 data items in our dataset. We use 48 data for test data and the rest for training data. Visualization of the forecasting results on the train data shows that XGBoost has the most approximate curve with the test data. ARIMA is the second most approximate. Lastly, MLP is the least approximate curve. Figure 7.b shows the forecasting model using all the data. We arrange for each model to forecast for the next 48 steps. Qualitatively, XGBoost and ARIMA are the two most approximate models.



(a)



(b)

Figure 7. The Test Results of the Four Models: (a) The Comparison of Test Data with the Forecasting Result of the Four Models (b) The Forecasting Result of the Four Models

3.2 Discussion

We use the MAPE, RMSE, MAE, and R^2 metrics to compare the four models. Table 3 shows the results of the comparison. The best result of each metrics is presented in bold text. XGboost excels in all four metrics compared to the other three models. XGBoost's MAPE, RMSE, MAE, and R^2 values are 0.48, 5.00, 3.09, and 0.61. R^2 values above 0.50 have a moderate category [43]. Among the four models, only XGBoost has a moderate category. ARIMA's performance is categorized as weak. Then LSTM's performance is categorized as very weak. Lastly, the R^2 value of the MLP model is negative. It shows that the predicted results are worse than the average of the actual values [44].

Table 3. Model Performance Comparison and Evaluation

Reference	Model	MAPE	RMSE	MAE	R^2
[7]	MLP	0.75	9.57	8.58	-0.43
[8]	LSTM	1.03	8.02	6.99	0.00
[9]	ARIMA	0.69	6.16	5.44	0.41
Proposed Model	**XGBoost**	**0.48**	**5.00**	**3.09**	**0.61**

ARIMA is a great method for forecasting [9]. However, the paper [15] mentions that ARIMA could be better at non-stationary data. In this study, we prove that the load forecasting data during WFH has a significant average difference from that during WFO. Then we also prove that our dataset is non-stationary. In our research, our proposed model, XGBoost, performs better on forecasting the dataset compared to ARIMA. Our research contribution is a forecasting model using the XGBoost method, which performs well on non-stationary data obtained during the WFO.

Deep learning methods such as LSTM have shown their advantages in various cases of sequential data [9]. However, the paper [14] mentioned that deep learning methods require high volumes of data to perform well. We used a large secondary dataset from Huggingface Datasets for electricity load forecasting [45]. The dataset contains electricity load information with 1344 data items. LSTM proved to be a superior model compared to other methods on the dataset. We use the same hyperparameter in the PT Biofarma dataset, which is only 18.15% of the Huggingface dataset. The result is that LSTM's performance is outperformed by our XGBoost model. Our research contribution is two-fold. First, we produced a limited dataset regarding electricity load forecasting from PT Biofarma. Secondly, we contribute with an electricity load forecasting model with good performance on limited datasets.

Several other studies have used seasonal ARIMA (SARIMA), where the model can add a seasonal component to its autoregressive model [46]. Future works may try SARIMA to increase performance on limited datasets mixed with COVID-19 datasets.

4. Conclusion

We make a system to measure the electrical energy spent on office lighting systems. Our test subject is the office of PT. Biofarma. The range of data collection covers the WFO season due to COVID-19. We propose an autoregressive XGBoost forecasting model that can be used for limited and non-stationary datasets. We use LSTM, ARIMA, and MLP as benchmarks. The test results with the t-test show that the electricity load data during WFO and WFH have a significant average difference. Then the test results with the ADF test show that our data is non-stationary. Compared to the benchmark method, the XGBoost method shows the best forecasting performance with MAPE, RMSE, MAE, and R^2 values of 0.48, 5.00, 3.09, and 0.61, respectively.

Acknowledgement

We thank PT Biofarma for agreeing to become the environment target for installing our electrical measuring devices. Furthermore, we thank PT INTI for helping to collect the raw electricity load dataset for this research. We also thank RISPRO LPDP for fully funding our research. Finally, we would like to thank Rajiv Shah for the electricity load dataset that he shared on Huggingface Datasets.

References

[1] X. Wei, H. Ren, S. Ullah, and C. Bozkurt, “Does environmental entrepreneurship play a role in sustainable green development? Evidence from emerging Asian economies,” *Econ. Res.-Ekon. Istraživanja*, vol. 36, no. 1, pp. 73–85, 2023. https://doi.org/10.1080/1331677x.2022.2067887.

[2] L. Zhang *et al.*, “A review of machine learning in building load prediction,” *Appl. Energy*, vol. 285, p. 116452, Mar. 2021. http://doi.org/10.1016/j.apenergy.2021.116452

[3] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, “A review of deep learning for renewable energy forecasting,” *Energy Convers. Manag.*, vol. 198, p. 111799, 2019. https://doi.org/10.5958/2278-4853.2021.00832.6.

[4] Z. Shao, S. Yang, F. Gao, K. Zhou, and P. Lin, “A new electricity price prediction strategy using mutual information-based SVM-RFE classification,” *Renew. Sustain. Energy Rev.*, vol. 70, pp. 330–341, Apr. 2017. http://doi.org/10.1016/j.rser.2016.11.155.

[5] C. Robinson *et al.*, “Machine learning approaches for estimating commercial building energy consumption,” *Appl. Energy*, vol. 208, pp. 889–904, Dec. 2017. http://doi.org/10.1016/j.apenergy.2017.09.060.

[6] A. I. Arvanitidis, D. Bargiotas, D. Kontogiannis, A. Fevgas, and M. Alamaniotis, “Optimized Data-Driven Models for Short-Term Electricity Price Forecasting Based on Signal Decomposition and Clustering Techniques,” *Energies*, vol. 15, no. 21, Art. no. 21, Jan. 2022. https://doi.org/10.3390/en15217929.

[7] M. Askari and F. Keynia, “Mid-term electricity load forecasting by a new composite method based on optimal learning MLP algorithm,” *IET Gener. Transm. Distrib.*, vol. 14, no. 5, pp. 845–852, 2020. https://doi.org/10.1049/iet-gtd.2019.0797.

[8] B. Nepal, M. Yamaha, A. Yokoe, and T. Yamaji, “Electricity load forecasting using clustering and ARIMA model for energy management in buildings,” *Jpn. Archit. Rev.*, vol. 3, no. 1, pp. 62–76, 2020. https://doi.org/10.1002/2475-8876.12135.

[9] R. E. Alden, H. Gong, C. Ababei, and D. M. Ionel, “LSTM forecasts for smart home electricity usage,” in *2020 9th International conference on renewable energy research and application (ICRERA)*, 2020, pp. 434–438. https://doi.org/10.1109/icrera49962.2020.9242804.

[10] C. Donnat and S. Holmes, “Modeling the heterogeneity in COVID-19’s reproductive number and its impact on predictive scenarios,” *J. Appl. Stat.*, pp. 1–29, 2021. https://doi.org/10.1080/02664763.2021.1941806.

[11] S. F. Pane, Heriyanto, A. G. Putrada, N. Alamsyah, and M. N. Fauzan, “The Influence of The COVID-19 Pandemics in Indonesia On Predicting Economic Sectors,” in *2022 Seventh International Conference on Informatics and Computing (ICIC)*, Dec. 2022, pp. 1–6. http://doi.org/10.1109/ICIC56845.2022.10006897.

[12] R. Chandra, A. Jain, and D. S. Chauhan, “Deep learning via LSTM models for COVID-19 infection forecasting in India,” *PLOS ONE*, vol. 17, no. 1, p. e0262708, Jan. 2022. http://doi.org/10.1371/journal.pone.0262708.

[13] M. H. A. Zamzuri, N. Sofian, and R. Hassan, “The Forecasting of Poverty using the Ensemble Learning Classification Methods,” *Int. J. Perceptive Cogn. Comput.*, vol. 9, no. 1, pp. 24–32, 2023. https://doi.org/10.31436/ijpcc.v9i1.326.

[14] X. Zhao, Q. Li, W. Xue, Y. Zhao, H. Zhao, and S. Guo, “Research on Ultra-Short-Term Load Forecasting Based on Real-Time Electricity Price and Window-Based XGBoost Model,” *Energies*, vol. 15, no. 19, p. 7367, 2022. https://doi.org/10.3390/en15197367.

[15] C. Li *et al.*, "Power load forecasting based on the combined model of LSTM and XGBoost," in *Proceedings of the 2019 the international conference on pattern recognition and artificial intelligence*, 2019, pp. 46–51. https://doi.org/10.1145/3357777.3357792.
[16] M. Xue, L. Wu, Q. P. Zhang, J. X. Lu, X. Mao, and Y. Pan, "Research on load forecasting of charging station based on XGBoost and LSTM model," in *Journal of Physics: Conference Series*, 2021, vol. 1757, no. 1, p. 012145. https://doi.org/10.1088/1742-6596/1757/1/012145.
[17] T. U. Wiganarto, A. Asenar, and E. Gultom, "Legal Aspects of Business Competition in the Procurement of Covid-19 Vaccine by Bio Farma Ltd," *Kanun J. Ilmu Huk.*, vol. 23, no. 2, pp. 193–209, 2021. https://doi.org/10.24815/kanun.v23i2.20416.
[18] A. G. Putrada, N. G. Ramadhan, and M. A. Makky, "An evaluation of activity recognition with hierarchical hidden Markov model and other methods for smart lighting in office buildings," *ICIC Int.*, 2022. http://doi.org/10.24507/icicel.16.01.91.
[19] S. Midesia, "Dampak Covid-19 Pada Pasar Saham Syariah Di Indonesia," *J. Penelit. Ekon. Akunt. JENSI*, vol. 4, no. 1, pp. 68–79, 2020. https://doi.org/10.33059/jensi.v4i1.2663.
[20] F. Nataliia, H. Yevgen, K. Artem, S. Denys, and H. Iryna, "Electric Meters Monitoring System for Residential Buildings," *Adv. Intell. Syst. Comput. Sci. Digit. Econ. IV*, pp. 173–185, 2023. https://doi.org/10.1007/978-3-031-24475-9_15.
[21] M. Abdurohman, A. G. Putrada, S. Prabowo, C. W. Wijiutomo, and A. Elmangoush, "M2M device connectivity framework," *Int. J. Electr. Eng. Inform.*, vol. 9, no. 3, pp. 441–454, 2017. https://doi.org/10.15676/ijeei.2017.9.3.2.
[22] M. Abdurohman, A. G. Putrada, S. Prabowo, C. W. Wijiutomo, and A. Elmangoush, "Integrated lighting enabler system using M2M platforms for enhancing energy efficiency," *J. Inf. Process. Syst.*, vol. 14, no. 4, pp. 1033–1048, 2018. http://dx.doi.org/10.3745/JIPS.03.0103.
[23] J. Wu, M. Yunus, P. K. Streatfield, and M. Emch, "Association of climate variability and childhood diarrhoeal disease in rural Bangladesh, 2000–2006," *Epidemiol. Infect.*, vol. 142, no. 9, pp. 1859–1868, 2014. https://doi.org/10.1017/s095026881300277x.
[24] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: A seasonal-trend decomposition," J Stat, vol. 6, no. 1, pp. 3–73, 1990.
[25] Z. Shen, G. Qin, P. Zuo, F. Wei, and X. Xu, "A Study of Variations of Galactic Cosmic-Ray Intensity Based on a Hybrid Data-processing Method," *Astrophys. J.*, vol. 900, no. 2, p. 143, 2020. https://doi.org/10.3847/1538-4357/abac60.
[26] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, "CIMA: A Novel Classification-Integrated Moving Average Model for Smart Lighting Intelligent Control Based on Human Presence," *Complexity*, vol. 2022, pp. 1–19, Sep. 2022. http://doi.org/10.1155/2022/4989344.
[27] M. H. Ferdous, U. Hasan, and M. O. Gani, "CDANs: Temporal Causal Discovery from Autocorrelated and Non-Stationary Time Series Data," *ArXiv Prepr. ArXiv230203246*, 2023. https://doi.org/10.48550/arXiv.2302.03246.
[28] S. Pasari and A. Shah, "Time series auto-regressive integrated moving average model for renewable energy forecasting," in *Enhancing Future Skills and Entrepreneurship: 3rd Indo-German Conference on Sustainability in Engineering*, 2020, pp. 71–77. https://doi.org/10.1007/978-3-030-44248-4_7.
[29] N. Abu Bakar and S. Rosbi, "Autoregressive integrated moving average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction," *Int. J. Adv. Eng. Res. Sci.*, vol. 4, no. 11, pp. 130–137, 2017. https://doi.org/10.22161/ijaers.4.11.20
[30] N. Maleki, A. Nikoubin, M. Rabbani, and Y. Zeinali, "Bitcoin price prediction based on other cryptocurrencies using machine learning and time series analysis," *Sci. Iran.*, 2020. https://doi.org/10.24200/sci.2020.55034.4040.
[31] A. G. Putrada, N. Alamsyah, S. F. Pane, and M. N. Fauzan, "XGBoost for IDS on WSN Cyber Attacks with Imbalanced Data," in *2022 International Symposium on Electronics and Smart Devices (ISESD)*, Nov. 2022, pp. 1–7. http://doi.org/10.1109/ISESD56103.2022.9980630.
[32] A. G. Putrada, M. Abdurohman, D. Perdana, and H. H. Nuha, "Machine Learning Methods in Smart Lighting Toward Achieving User Comfort: A Survey," *IEEE Access*, vol. 10, pp. 45137–45178, 2022. http://doi.org/10.1109/ACCESS.2022.3169765.
[33] N. Engelke Infante, K. Murphy, C. Glenn, and V. Sealey, "How concept images affect students' interpretations of Newton's method," *Int. J. Math. Educ. Sci. Technol.*, vol. 49, no. 5, pp. 643–659, 2018. https://doi.org/10.1080/0020739x.2017.1410737.
[34] R. R. Pamungkas, A. G. Putrada, and M. Abdurohman, "Performance improvement of non invasive blood glucose measuring system with near infra red using artificial neural networks," *Kinet. Game Technol. Inf. Syst. Comput. Netw. Comput. Electron. Control*, pp. 315–324, 2019. https://doi.org/10.22219/kinetik.v4i4.844.
[35] M. F. Akbar, A. G. Putrada, and M. Abdurohman, "Smart light recommending system using artificial neural network algorithm," in *2019 7th International Conference on Information and Communication Technology (ICoICT)*, 2019, pp. 1–5. https://doi.org/10.1109/icoict.2019.8835192.
[36] R. Irvan, M. Abdurohman, and A. G. Putrada, "Designing a monitoring and prediction system of water quality pollution using artificial neural networks for freshwater fish cultivation in reservoirs," in *Proceedings of Sixth International Congress on Information and Communication Technology*, 2022, pp. 469–476. https://doi.org/10.1007/978-981-16-2380-6_41.
[37] A. M. Syukur, A. G. Putrada, and M. Abdurohman, "Increasing accuracy of power consumption using artificial neural network," in *2019 7th International Conference on Information and Communication Technology (ICoICT)*, 2019, pp. 1–6. https://doi.org/10.1109/icoict.2019.8835371.
[38] M. A. G. Putrada, M. Abdurohman, D. D. Perdana, and D. H. H. Nuha, "Recurrent Neural Network Architectures Comparison in Time-Series Binary Classification on IoT-Based Smart Lighting Control," p. 6. https://doi.org/10.1109/icoict55009.2022.9914831.
[39] H. C. Kilinc, "Daily streamflow forecasting based on the hybrid Particle Swarm Optimization and Long Short-Term Memory model in the Orontes Basin," *Water*, vol. 14, no. 3, p. 490, 2022. https://doi.org/10.3390/w14030490
[40] A. G. Putrada, N. Alamsyah, S. F. Pane, and M. N. Fauzan, "GRU-MF: A Novel Appliance Classification Method for Non-Intrusive Load Monitoring Data," in *2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, 2022, pp. 200–205. https://doi.org/10.1109/comnetsat56033.2022.9994409.
[41] A. G. Putrada, "Gated Recurrent Unit for Fall Detection on Motorcycle Smart Helmet with Accelerometer Sensor," *Indones. J. Comput. Indo-JC*, vol. 7, no. 3, Art. no. 3, Dec. 2022. http://doi.org/10.34818/INDOJC.2022.7.3.672.
[42] E. Sener and I. Demir, "Wavelet-Denoised ARIMA and LSTM Method to Predict Cryptocurrency Price," in *2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 2022, pp. 577–582. https://doi.org/10.1109/ismsit56059.2022.9932727.
[43] A. Narotama, "The Effect of Service Quality on Customer Loyalty with Brand Images and Customer Satisfaction as Mediation," *J. Manaj. DAN KEWIRAUSAHAAN*, vol. 7, no. 1, p. 86, Jun. 2019. http://doi.org/10.26905/jmdk.v7i1.2692.
[44] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, p. e623, 2021. https://doi.org/10.7717/peerj-cs.623.
[45] Q. Lhoest *et al.*, "Datasets: A community library for natural language processing," *ArXiv Prepr. ArXiv210902846*, 2021. https://doi.org/10.48550/arXiv.2109.02846.
[46] J. Cong, M. Ren, S. Xie, and P. Wang, "Predicting Seasonal Influenza Based on SARIMA Model, in Mainland China from 2005 to 2018," *Int. J. Environ. Res. Public. Health*, vol. 16, no. 23, Art. no. 23, Jan. 2019. http://doi.org/10.3390/ijerph16234760.