JOIN (Jurnal Online Informatika)
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 9 Number 2 | December 2024: 269-275
DOI: 10.15575/join.v9i2.1113

Prediction of Solar Radiation Data for Garlic Production in Magelang Regency Using Long Short-Term Memory

Muhammad Safrul Safrudin[1], Imas Sukaesih Sitanggang[2], Hari Agung Adrianto[3], Syarifah Aini[4]
[1234] Department of Computer Science, Faculty of Mathematics and Natural Sciences, IPB University, Indonesia

Article Info

Article history:

Received June 28, 2023
Revised August 09, 2024
Accepted August 26, 2024
Published December, 30 2024

Keywords:

Garlic
LSTM
Prediction Models

ABSTRACT

Garlic importation in Indonesia is frequently carried out to meet the high domestic market demand. To reduce dependency on imports, the development of local garlic production is crucial. This study aims to determine the optimal solar radiation for garlic growth using the Long Short-Term Memory (LSTM) algorithm. This algorithm was selected due to its ability to analyze time-series data and predict long-term patterns. The LSTM model was trained with the Adam optimizer, using a configuration of 1000 epochs, a batch size of 6, and a dropout rate of 2.0 to prevent overfitting. The model evaluation results show an indicating good accuracy with a RMSE of 0.1020, a Mean Squared Error (MSE) of 0.0104, and a correlation coefficient of 0.740, although it still has limitations in capturing extreme data fluctuations. The study found that in Magelang Regency especially in the sub-districts of Windusari, Grabag, Ngablak, Pakis, Dukun, Kaliangkrik, and Kajoran have optimal solar radiation for garlic cultivation between March and May, with a radiation range of 380 W/m² to 440 W/m². These findings provide valuable guidance for farmers in determining the optimal planting period, potentially enhancing local garlic production and reducing import dependency.

Corresponding Author:

Muhammad Safrul Safrudin,
Department of Computer Science, Faculty of Mathematics and Natural Sciences, IPB University
Jl. Raya Darmaga Kampus IPB, Babakan, Kec. Dramaga, Kabupaten Bogor, Indonesia. 16680
Email: muhammadsafrul@apps.ipb.ac.id

1. INTRODUCTION

Garlic is one of the strategic horticultural commodities that is very important and has high economic value because in addition to functioning as a spice in cooking, this plant also has medicinal properties that can inhibit the development of Escherichia coli bacteria. Garlic is also widely cultivated in highland areas to get more sunlight in the mountains. The demand for garlic continues to increase from time to time in Indonesia, but domestic production is still unable to meet the demand.

Therefore, garlic imports are required to fulfil domestic demand. This phenomenon can be explained as the inability of domestic production to fulfil the increasing market demand. This condition requires the government to import in an effort to fulfil domestic market needs and avoid supply shortages and price increases that can harm consumers. Although imports can help fulfil domestic market needs, it should also be noted that garlic imports can have an impact on the growth of the domestic garlic industry. Therefore, the government must also consider long-term strategies in managing the garlic industry in order to meet domestic market needs with optimal domestic production. Garlic consumption in 2020 to 2024 is projected to increase by 1.38% per year, so that the total garlic consumption is predicted to reach 515.74 thousand tonnes in 2021 [1].

This shows that garlic imports cover 80% of national needs, in an effort to achieve garlic self-sufficiency. From 2020 to 2029, the government will plan the development and stabilization of land areas for garlic cultivation. The plan aims for higher domestic garlic production [2]. So, it is necessary to identify suitable land for garlic plants to support the government's goals.

Garlic production in Indonesia comes from the island of Java especially Central Java Province which reaches 40% of domestic production [1]. Magelang district in Central Java has a contribution of 3.15% to national garlic production. Therefore, land development for garlic planting in this area can be a good option. Plant growth factors are strongly influenced by climatic factors, namely solar radiation where plant growth, especially garlic, is needed as a source of energy, especially in photosynthesis and greatly affects plant properties. The higher the light intensity obtained by plants can accelerate the rate of photosynthesis, the amount of energy given to the leaves means that the energy is for the process of synthesizing carbohydrates, but if the light energy is too high it can damage photosynthetic pigments.

The optimal level of solar radiation varies depending on the growth phase of garlic. For example, in the vegetative phase, garlic requires higher solar radiation to promote leaf and stem growth. However, in the bulb formation phase, lower solar radiation is required to promote better bulb growth [7]. In this case, the growth phase of garlic by looking at the intensity of solar radiation in a region can be considered to significantly increase the production and quality of garlic [8].

Direct observation of large amounts of solar radiation data on land in each region manually is impractical. Therefore, a more efficient approach is needed, one of which is a computational approach. In this context, one of the computational approaches applied is the deep learning algorithm. By using this approach, it can minimize the search space and make the processing of solar radiation data efficient. In addition, it is possible to predict the optimal region and time period for planting garlic based on solar radiation data. One of the algorithms in deep learning that can be used is the LSTM algorithm.

The LSTM algorithm has the ability to remember long-term information and handle time series data. The advantage of LSTM is its ability to capture temporal information from sequential data [19], which is very important in predicting short-term wind speed, the LSTM model is proven to be superior to other baseline models in terms of prediction performance [20][3]. Research by modeling land suitability by making prediction using a deep learning algorithm of garlic land spatial decision trees [4]. Research in analyzing the suitability of garlic land using the Spatial Clustering approach. However, both studies have not considered solar radiation data as the most important factor in plant growth [5].

In this study, modeling is carried out on solar radiation data obtained from the NASA website to predict the suitability of garlic land and optimal garlic planting time in Magelang Regency, Central Java using the LSTM algorithm.

2. MATERIAL AND METHOD

2.1. Research flow

This research was conducted through several stages: data collection, data pre-processing, division of training data and test data, modeling, and model evaluation. The research stages can be seen in Figure 1.

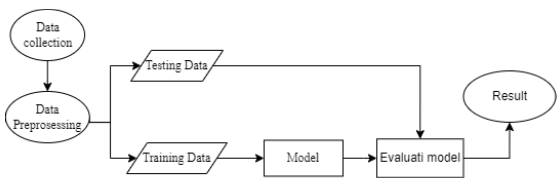

Figure 1. Research flow

2.1.1 Data

The data used in this study are daily solar radiation data in units of W/m^2, taken from the website https://power.larc.nasa.gov. We selected solar radiation data in the Magelang district, Central Java, from 1 January 2010 to 31 December 2022, with a total of 4748 data rows and 3 attributes, namely year, date, and solar radiation value. Table 1 shows the irradiance dataset. The reason for selecting this solar radiation data is because solar radiation is one of the environmental factors that greatly affects plant growth, including garlic which is the focus of this study. Optimal solar radiation is essential for the photosynthesis process, which in turn affects crop productivity.

Magelang district was chosen as the study area because it has significant agricultural potential and has been recognized as one of the garlic production centers in Indonesia. Understanding the pattern of solar radiation in this region will provide important insights to optimize agricultural yields, particularly in garlic cultivation. In addition, the long time span of data collection, 13 years, allows for a more in-depth analysis of the changing trends of solar radiation in this region, which can be used for long-term agricultural planning.

Table 1. Solar Radiation Dataset

Year	Day	Solar Radiation
2010	1	405.73
2010	2	407.41
...	...	...
2022	364	434.84

2.1.2 Preprossesing Data

Data pre-processing begins with data cleaning, including handling of missing values, removal of duplicates, and handling of outliers. Next, categorical data is converted into numerical data through encoding techniques. Next, normalization is performed using mix- max normalization, which is scaled with a value of 0 to 1 or numeric feature standardization. Next, the data is separated into features and targets. Table 2 is the result of the pre-processed data.

Table 2. Normalized Solar Radiation Dataset

Day	Solar Radiation
1	0.587441
2	0.648950
...	...
4747	0.910202

2.1.3 Training and testing data

The data is divided into two parts, namely training data and test data, is an important step in the development of prediction models. In this study, data division is done using the split validation method, where 80% of the total dataset is allocated to training data, and the remaining 20% is allocated to test data. An 80:20 split is considered ideal as more training data allows the model to learn patterns better, while a significant amount of test data allows for an accurate evaluation of the generalization ability of the model [6].

This division process is important to avoid overfitting, which is a condition where the model overfits itself to the training data and thus loses its ability to predict new data well [9]. Figure 2 shows a visualization of the split between training and testing data, where training data (80%) is used to train the prediction model, and testing data (20%) is used to evaluate its performance.

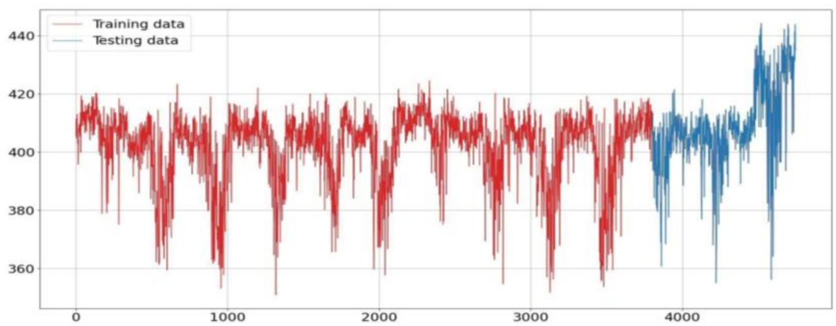

Figure 2. Split of training and testing data

2.1.4 Modeling

At this stage, modeling is carried out using the LSTM method. LSTM is a deep learning method developed from the Recurrent Neural Network (RNN) method [10]. Before modeling, tuning is done on the parameters of the number of neurons, batch size, number of epochs, and optimizer with learning rate. Table 3 is the result of the parameter tuning performed.

Tabel 3. Parameter tuning

Activation	Batch_ Size	Dropout _Rate	Epochs	Neurons	Optimizer	Verbose	Score
sigmoid	8	0.2	1000	10	adam	1	-0.072636
sigmoid	16	0.2	1000	10	adam	1	-0.072849
sigmoid	6	0.2	1000	10	adam	1	-0.073433
sigmoid	6	0.2	1000	10	adamax	1	-0.074046
sigmoid	8	0.2	1000	10	adamax	1	-0.074980
sigmoid	16	0.2	1000	10	adamax	1	-0.075626
relu	16	0.2	1000	10	adamax	1	-0.088249
relu	6	0.2	1000	10	adamax	1	-0.090805
relu	8	0.2	1000	10	adamax	1	-0.091236
relu	6	0.2	1000	10	adam	1	-0.096010
relu	8	0.2	1000	10	adam	1	-0.096074
relu	16	0.2	1000	10	adam	1	-0.096271

From Table 3, it can be seen that the best parameter values obtained are the number of neurons 10, batch size 8, activation sigmoid, epoch 1000 and optimizer adam. These parameter values are used as parameters for modeling with LSTM. Furthermore, predictions are made using the model. LSTM is very effective in modelling time series data, especially compared to other types of artificial neural networks [13]. Firstly, LSTM is designed to overcome the "vanishing gradient" problem that often occurs in traditional Recurrent Neural Network (RNN) [14]. Mechanisms such as forget gate, input gate and output gate, LSTM can retain important information in long-term memory and discard irrelevant information [15].

LSTM has high flexibility in handling data with complex and non-linear patterns [16]. LSTM is able to learn from both long- and short-term temporal patterns, which gives it a significant advantage in predicting phenomena that are highly dependent on time sequences [17].

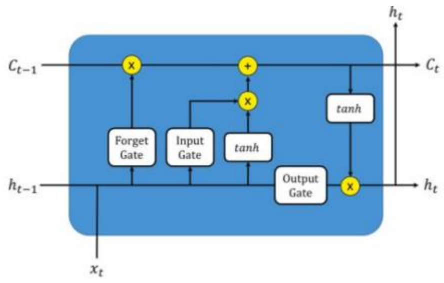

Figure 3. The architecture of LSTM [14]

After the prediction process is complete, the predicted data then undergoes denormalization. The purpose of this denormalization is to get the actual prediction result, which is the value returned to its original scale or format after going through the normalization process.

2.1.5 Model Evaluation

At this stage, an evaluation of the resulting model is carried out. Model evaluation is done by calculating the mean square error (MSE), root mean square error (RMSE), and mean absolute percentage error (MAPE). MSE is used to calculate the average square error in prediction. The smaller the MSE value, the better the quality of the model [11]. RMSE is a derivative of MSE, where RMSE is the square root of

MSE. The smaller the RMSE value, the better the resulting model. MAPE is used to calculate the average error in prediction as a percentage of the actual value. The smaller the MAPE value, the better the quality of the model [12]. Equation 1, Equation 2, and Equation 3, and Equation 4 are formulas used to calculate MSE, RMSE, and MAPE.

$$MSE = \frac{1}{n}\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)2 \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\frac{(y_j-\hat{y}_j)2}{n}} \quad (2)$$

$$MAPE = \frac{1}{n}\sum_{j=1}^{n}\left|\frac{y_j-\hat{y}_j}{n}\right| * 100 \quad (3)$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{Tot}} \quad (4)$$

3. RESULTS AND DISCUSSION

This study used daily solar radiation data measured in watts per square meter for 10 years. The LSTM model used to predict solar radiation intensity showed good performance with the following evaluation matrix: RMSE of 0.1020, MSE of 0.0104, MAE of 0.0760, and R^2 of 0.7400. These values indicate that the LSTM model is able to predict solar radiation with high accuracy, where an R^2 value close to 1 indicates an excellent model in explaining data variability. Figure 3 is the result of the comparison of actual data and predicted data.

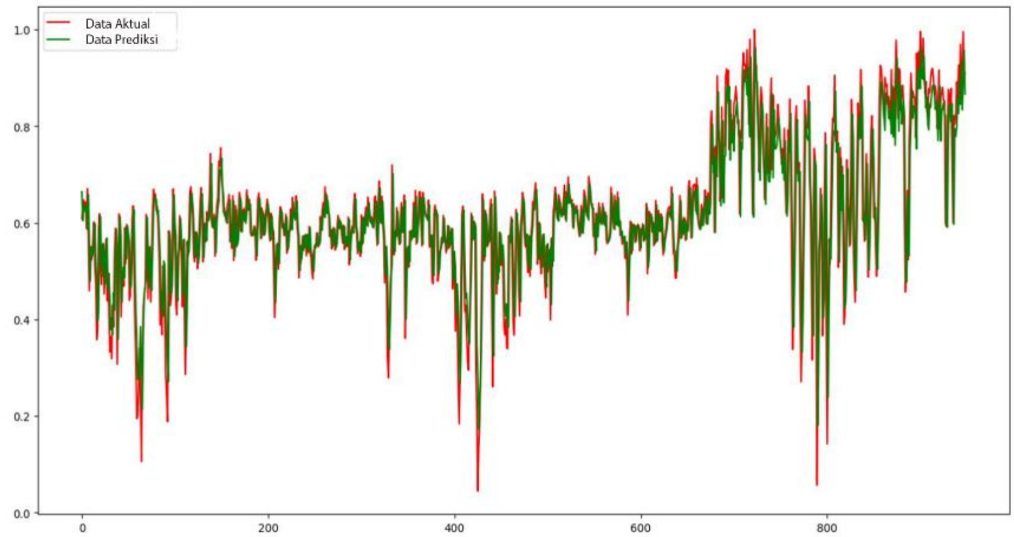

Figure 4. Comparison of actual data and prediction data

LSTM model with good accuracy can help garlic farmers predict the right planting time period for garlic. By knowing the periods of high and low radiation, farmers can plant garlic at the right time so that each phase of plant growth gets optimal radiation conditions. In addition, by knowing the high and low irradiance periods, the use of water, fertilizer and other resources can be optimized. For example, during high irradiance periods, water requirements may increase, so irrigation can be adjusted to avoid drought. This will help maximize the yield and quality of garlic produced, which in turn can increase farmers' income and welfare.

Table 4. Solar radiation condition and range of solar radiation values by month in Magelang District

Month	Solar Radiation Conditions	Range of Solar Radiation Values	Analysis and Recommendations
January	Low radiation with small fluctuations	360 - 400	Conditions are not yet optimal for planting, better to wait for increased radiation.

Month	Solar Radiation Conditions	Range of Solar Radiation Values	Analysis and Recommendations
February	Low radiation, starting to increase	370 - 410	Conditions are improving, but risks remain. Wait for further increase.
March	Radiation is starting to increase	380 - 420	Planting time begins to be ideal, radiation starts to support plant growth.
April	Radiation increases significantly	400 - 430	Excellent planting time, supports vegetative growth of garlic.
May	High and stable radiation	410 - 440	Optimal planting time, plants will receive sufficient light for good growth.
June	High radiation, stable with slight fluctuations	400 - 430	Still good for maturation, suitable for the final growth stage.
July	High enough radiation, fluctuations begin	380 - 420	Conditions are still supportive but monitor emerging fluctuations.
August	High enough radiation, starting to decline	370 - 410	Good for maturation stage, prepare for harvest.
September	Radiation starts to decline	380 - 420	Radiation remains stable, good for final maturation or drying stage.
October	Low, stable radiation	370 - 410	Planting time is less ideal, growth may not be optimal.
November	Low radiation with fluctuations	360 - 400	High risk, growth may not be optimal.
December	Low radiation with small fluctuations	360 - 400	Conditions are not ideal for planting, consider starting in early January.

Table 4 shows the condition of solar radiation and the range of solar radiation values by month in Magelang district. From Table 4, it can be seen that garlic should be planted once a year, with the most optimal planting time starting from March to May. During this period, solar radiation experiences a steady increase, thus supporting the vegetative phase of the plant. High and stable radiation conditions allow garlic to develop strong leaves and roots, to enter the bulb formation phase.

After May, solar radiation remains high but starts to show slight fluctuations until July. This period remains favourable for bulb formation and maturation, although farmers need to be more careful in irrigation management to ensure the plants are not water deficient. Entering August to September, solar radiation starts to decrease, which is an ideal condition for the ripening and drying phase of garlic before harvesting is done in late September or early October.

From October to December, solar radiation is low and stable, so new plantings are not recommended because these conditions do not favour the early growth of garlic plants. Instead, these months are more suitable for harvest storage, as the low radiation and temperature conditions can help maintain the quality of the harvested garlic bulbs.

4. CONCLUSION

This study shows that the solar radiation prediction model developed using the Long Short-Term Memory (LSTM) algorithm with parameter tuning provides good results in following the actual data pattern. The LSTM model with the best parameters (epoch 100, batch size 4, dropout 0.2, and optimizer Adam) produces an RMSE of 0.0852 after tuning. The RMSE value of the model after prediction is 0.1020, the MSE is 0.0104, the MAE is 0.0760, and the coefficient of determination R^2 is 0.7400. These values show that the LSTM model is able to predict solar radiation with high accuracy, although it still has limitations in capturing very extreme patterns from actual data. These solar radiation predictions provide valuable insights for assessing land suitability and the optimal time period for planting garlic in Magelang district. The results show that the most optimal planting time is between March to May, where solar radiation experiences a steady increase, supporting the vegetative phase of the garlic crop. After May, solar radiation remains high but starts to show slight fluctuations until July. During this period, bulb formation and maturation are supported, although irrigation management needs to be considered to avoid water shortages. Entering August to September, solar radiation starts to decrease, which is the ideal condition for the ripening and drying phase of garlic before harvesting in late September or early October. With this prediction model, farmers can optimize the use of water, fertilizer, and other resources based on the estimated solar radiation, which in turn can increase the production yield and quality of garlic as well as farmers' welfare. This model has the potential to be an effective tool in supporting garlic self-sufficiency in Indonesia and reducing dependence on imports.

REFERENCES

[1] "Outlook bawang putih Pusat Data dan Sistem Informasi Pertanian Sekretariat Jenderal Kementerian Pertanian 2020," 2020.

[2] Dirjen Hortikultura. 2017. Pengembangan bawang putih nasional. Kementerian Pertanian.

[3] Qian, P., Zhang, Q., Peng, L., Zhu, Q., Wang, X., & Chen, X. *A novel 2Atts-LSTM model with spatio-temporal attention for short-term wind speed forecasting*. Renewable Energy, 173, 931-945.2021. https://doi.org/10.1016/j.renene.2021.04.064.

[4] Nurkholis A, Sitanggang IS, Annisa, Sobir. Spatial Decision Tree Model For Garlic Land Suitability Evaluation. IAES International Journal of Artificial Intelligence (IJ-AI). 2021.

[5] Purwayoga V. Analisis Pemetaan Kesesuaian Lahan Untuk Bawang Putih dengan Pendekatan Spatial Clustering [tesis] Bogor: Institut Pertanian Bogor. 2019.

[6] L. Ni et al., "Streamflow and rainfall forecasting by two long short-term memory-based models," J Hydrol (Amst), vol. 583, p. 124296, Apr. 2020, doi: 10.1016/j.jhydrol.2019.124296.

[6] I. Renewable Energy Agency, *RENEWABLE POWER GENERATION COSTS IN 2018*. 2019. [Online]. Available: www.irena.org

[7] Gartner L Barbara, "Plant Stems: Physiology and Functional Morphology," 1995.

[8] P. Liu *et al.*, "Distinct Quality Changes of Garlic Bulb during Growth by Metabolomics Analysis," *J Agric Food Chem*, vol. 68, no. 20, pp. 5752–5762, May 2020, doi: 10.1021/acs.jafc.0c01120.

[9] W. Chen, L. Yang, B. Zha, M. Zhang, and Y. Chen, "Deep learning reservoir porosity prediction based on multilayer long shortterm memory network," GEOPHYSICS, vol. 85, no. 4, pp. WA213–WA225, Jul. 2020, doi: 10.1190/geo2019-0261.1

[10] S. Mahjoub, L. Chrifi-Alaoui, B. Marhic, and L. Delahoche, "Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks," *Sensors*, vol. 22, no. 11, Jun. 2022, doi: 10.3390/s22114062.

[11] L. Liu et al., "Transient rotor angle stability prediction method based on SVM and LSTM network," Dianli Zidonghua Shebei/Electric Power Automation Equipment, vol. 40, no. 2, 2020, doi: 10.16081/j.epae.202001009.

[12] G. S. Vidya and V. S. Hari, "LSTM Network Integrated with Particle Filter for Predicting the Bus Passenger Traffic," J Signal Process Syst, vol. 95, no. 2–3, pp. 161–176, Mar. 2023, doi: 10.1007/s11265-022-01831-x.

[13] X. Liu and Z. Lin, "Impact of Covid-19 pandemic on electricity demand in the UK based on multivariate time series forecasting with Bidirectional Long Short Term Memory," *Energy*, vol. 227, Jul. 2021, doi: 10.1016/j.energy.2021.120455.

[14] M. Marani, M. Zeinali, V. Songmene, and C. K. Mechefske, "Tool wear prediction in high-speed turning of a steel alloy using long short-term memory modelling," Measurement, vol. 177, p. 109329, Jun. 2021, doi: 10.1016/j.measurement.2021.109329.

[15] M. Dehghan Manshadi, M. Ghassemi, S. M. Mousavi, A. H. Mosavi, and L. Kovacs, "Predicting the Parameters of Vortex Bladeless Wind Turbine Using Deep Learning Method of Long Short-Term Memory," Energies (Basel), vol. 14, no. 16, p. 4867, Aug. 2021, doi: 10.3390/en14164867.

[16] J. Zhao, F. Deng, Y. Cai, and J. Chen, "Long short-term memory - Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction," *Chemosphere*, vol. 220, pp. 486–492, Apr. 2019, doi: 10.1016/j.chemosphere.2018.12.128.

[17] P. Gong, Y. Ma, C. Li, X. Ma, and S. H. Noh, "Understand Data Preprocessing for Effective End-to-End Training of Deep Neural Networks," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.08925.

[18] J. Qian, M. Zhu, Y. Zhao, and X. He, "Short-term wind speed prediction with a two-layer attention-based lstm," *Computer Systems Science and Engineering*, vol. 39, no. 2, pp. 197–209, 2021, doi: 10.32604/csse.2021.016911.

[19] F. R. Alharbi and D. Csala, "Wind speed and solar irradiance prediction using a bidirectional long short-term memory model based on neural networks," *Energies (Basel)*, vol. 14, no. 20, Oct. 2021, doi: 10.3390/en14206501.

[20] M. Kamal Wisyaldin, G. Maya Luciana, H. Pariaman, and P. Pembangkitan Jawa Bali, "Pendekatan Long Short-Term Memory untuk Memprediksi Kondisi Motor 10 kV pada PLTU Batubara," vol. 9, no. 2, 2020, doi: 10.33322/kilat.v9i2.997.