JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)

Forecasting Photovoltaic Output Power Based on Environmental Parameters Using Artificial Neural Network Methods

Desri Kristina Silalahi[1], Agnes Christy Margareth Rumapea[2], Wahmisari Priharti[3], Bandiyah Sri Aprillia[4]
[1,2,3,4]Bachelor of Electrical Engineering, School of Electrical Engineering, Telkom University, Bandung, Indonesia
[1]desrikristina@telkomuniversity.ac.id, [2]agneschristymargareth@gmail.com, [3]wpriharti@telkomuniversity.ac.id,
[4]bandiyah@telkomuniversity.ac.id

Abstract

Photovoltaic is a system that can convert sunlight into electrical energy. However, photovoltaic efficiency tends to be low and its performance is affected by several environmental parameters such as dust, wind speed, humidity, temperature and other external factors. Because many factors can affect the power generated, we need a power output prediction system that can assist in planning and managing as well as increasing the efficiency of photovoltaic systems. In this research, a system is designed that can predict the photovoltaic output power in the short term using the Artificial Neural Network method or what is often called an artificial neural network. Predictions are made based on the effects of several environmental parameters such as wind speed, dust, humidity, and temperature on a 10 Wp photovoltaic system. Performance data for 7 days is used as a dataset and then processed using ANN with 1 input layer, 3 hidden layers 1 output layer and 3 sample epochs (10, 100, and 1000). The results of the study can predict the output of photovoltaic power for the next 4 days with an error value of Mean Square Error (MSE) of 0.0010, Mean Absolute Error (MAE) of 0.0155, Root Mean Square Error (RMSE) of 0.0229 with an increase in power reach 0.5 to 1 watt.

Keywords: artificial neural network; forecasting; output power; photovoltaic

1. Introduction

Solar cells are often also called photovoltaic and can be defined as a technology that generates direct current electricity from a semiconductor material when exposed to sunlight. As long as the semiconductor material is exposed to sunlight, the solar cell will always produce electrical energy, and when it is not exposed, the solar cell stops producing electrical energy [1]. Solar panels are tools for converting sunlight energy into electrical energy [2].

Indonesia has 40% reserves of solar thermal energy resources [3], sunlight can be used as a substitute for fossil fuels whose use is very critical at this time. Sunlight will be converted into electrical energy using photovoltaics. This photovoltaic system also has many benefits for the environment and can also improve the reliability of electric power systems [3]. With an ever-increasing demand for clean energy sources, the manufacture of solar panels has increased dramatically in recent years. However, there are also some disadvantages of photovoltaic systems, namely high investment costs and low photovoltaic power efficiency, this value can be affected by several

environmental parameters such as wind speed, dust, temperature and humidity [4].

The operation of solar panels is highly dependent on the intensity of the sun and environmental factors such as wind speed, dust, temperature and humidity. Environmental factors that are constantly changing also affect the output value of existing solar panels. Wind speed affects the output power of solar panels. Outdoor wind speed can be affected by weather and altitude. Indonesia has an average annual wind speed of only 3 m/s–6 m/s [5]. Based on the Institute for Essential Service Reform (IESR) report Indonesia has a wind potential of 25 GW at an altitude of 50m and 19.8 GW at an altitude of 100m using a minimum annual average wind speed of 7.25 m/s at 50 m and 7. 99m/s at 100m [6]. The dust deposition rate will affect the transmission of the glass surface on the panel and change the performance of the solar panel, this also results in the efficiency of the solar panel decreasing [7]. When the solar panel operates, the sun's radiant energy is converted into electrical energy and the temperature of the solar cell increases. Temperature affects the current-voltage characteristics of solar panels. With decreasing temperature, the electric current in solar panels also

decreases, even extreme temperature changes can disrupt electricity production in a solar power plant [8]. Humidity is the concentration content of water vapour in the air. The higher the humidity, the less output of solar panels such as voltage, current, and power [9].

A prediction system is needed to determine the output of photovoltaic power in the next few days. This system urgently needs to know when photovoltaics will produce a large power output and a small power output in the future. The timeframe in the prediction is classified into three categories namely short term, with a time range of 48-72 hours, medium term with a time range of 8-30 days and long term, with a time range of 1 month to 1 year [10].

The prediction process requires a method. Prediction methods have been widely used in other studies to predict photovoltaic output. Research on predicting photovoltaic power output was previously carried out by Reza Sabzehgar by predicting photovoltaic power output for the next 30 days using the Artificial Neural Network method. The parameters used are cloud cover, dew point, sun peak angle, precipitation, humidity, temperature and air pressure, located in San Diego, California, United States. From this study, it was obtained that MAPE reached 45.30% and MSE 422.91 [11]. The Support Vector Regression method with the Kernel RBF function can predict the photovoltaic output power with MAPE values in the range of 20%. MAE value is 0.004 and MSE is 0.069 [12]. The evaluation results of the forecasting model can predict the photovoltaic output power well. However, this study has not considered environmental parameters.

In this research, the photovoltaic output power will be predicted through the data obtained. The dataset will be obtained from a tool that has been designed through sensor readings of the environmental parameters used, such as wind speed, dust, humidity and temperature. Data is processed using the ANN method. Environmental factors greatly affect the power generated by solar panels, and also determine whether solar panels can work optimally or not.

2. Research Methods

The method in this study can be seen in Figure 1. The dataset is checked for empty data and discards unnecessary data. Then the data is entered into Google Colab. Next, the architectural development of the artificial neural network method is carried out starting from the input layer, hidden layer and output layer. In the next stage, the data learning or training process is carried out using an epoch. After the model is trained, the predicted and actual data obtained will be evaluated first. After obtaining the error value from the model, the

prediction of the output of photovoltaic power has been successfully obtained.
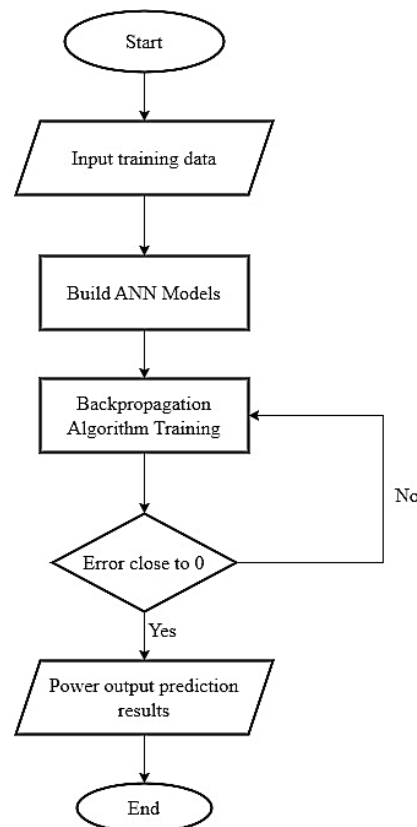

Figure 1. Flowchart of the System

2.1 Input Data

The input data used is in the form of wind speed, dust, temperature and humidity through a sensor that has been installed. Meanwhile, the output power data is photovoltaic, current and voltage. The data obtained was taken from Building P, Floor 3 Telkom University for 7 days from January 19 to January 25, 2022. The time of collection is from 08.00-17.00 WIB, data is collected every 1 minute. Photovoltaic will use the sun's light source which is converted into electrical energy. When photovoltaics capture sunlight, the solar energy is converted into electrical energy. Then data from photovoltaic and other components will be acquired using the Arduino application on a laptop.

2.2 Forecasting Models

Artificial Intelligence according to Andreas Kaplan and Michael Haenlin is a system's ability to interpret external data correctly, learn from that data, and use that learning to achieve certain goals through flexible adaptation [13]. The goal of using artificial intelligence is to limit its principles from the input data. This means that inaccuracies in the data will show up in the final result. So to get accurate results, the input data must also be accurate. The use of AI in this research is an Artificial Neural Network (ANN).

Artificial Neural Network or what is often referred to as an artificial neural network is information processing whose way of working is inspired by the biological nervous system such as the performance of the brain to obtain information [13]. The way ANN works is to model the relationship between input and output, it can also be used to recognize or find patterns in data. ANN has been successfully used as a tool for predicting time series and modelling in various application domains [14]. Artificial Neural Network has 3 layers, namely [15]: The input layer is the amount of data that will be entered to carry out the training process; the Hidden layer, layer serves as a liaison between the input and output layers. Aside from being a connector, hidden also functions as an input value calculation and classifies it into a certain type of value by using an activation function; The output layer becomes the data received from the node and the mash has the value that needs to be classified one last time.
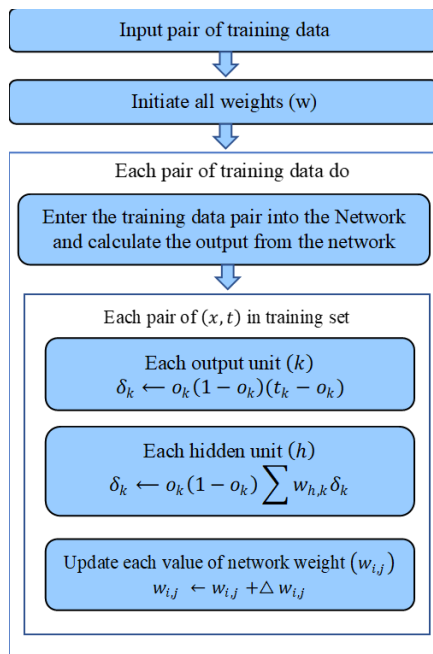

Figure 2. Backpropagation Algorithm

This research uses training in the form of a backpropagation algorithm. ANN-based backpropagation is included in supervised learning algorithms and is usually used by perceptrons with many layers to change the weights connected to the neurons in the hidden layer. The backpropagation algorithm is the most widely used in neural network training. The backpropagation algorithm makes it possible to update the weighting scheme with very small values in complex networks [16].

There are 3 phases in the backpropagation algorithm [17]. The first is the advanced phase. The input layer is first calculated forward by adding the weight and bias values to the output layer using a predetermined activation function. The second phase is the reverse phase, which calculates the difference between the network output and the desired target, which is then referred to as an error. Next, namely the back-propagation phase, the error factor is propagated backwards, starting from the line that is directly related to the units in the output layer. The last phase is modifying the weights to reduce the errors that occur [3]. The backpropagation algorithm can be seen in Figure 2. ANN uses an activation function to limit the Y output to match the output signal limit. The activation function plays a role in providing output to each neuron, several activation functions can be used according to the problem to be solved. The use of the activation function depends on the problem to be solved. For linear cases, the use of step or linear activation functions will be very effective. However, for nonlinear cases using the sigmoid, tanh, ReLU or Gaussian functions will give good results. The target output value is also a consideration in selecting the activation function. In this study using the ReLU activation function.

Figure 3 is a flowchart of the artificial neural network method. The dataset obtained is in the form of a time series which is inputted into Google Colab for processing. The first stage is the data is read then the data is divided into two parts, namely training data and testing data. The data is normalized to equalize the value scale of the environmental parameters used. Then the model architecture starts from the input layer, hidden layer and output layer. Data training on backpropagation is done to get the best modelling. The selected model is used to predict the output of the photovoltaic power. Then, evaluation of the prediction results is carried out by comparing the predicted data with actual data.

2.3. Model evaluation

In the case of predictions, the prediction model obtained must be validated to obtain the best predictive model used. The evaluation used as validation is by looking at the error value based on the statistical parameters.

Root Mean Square Error (RMSE) is a measurement method to measure the difference in the value of a model's prediction as an estimate of the observed value. RMSE is the result of the square root Mean Square Error [18]. The accuracy of the measurement error estimation method is indicated by the presence of a small RMSE value. An estimation method that has a smaller RMSE value is said to be more accurate than an estimation method that has a larger RMSE. Formula 1 is for the RMSE [18].

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(A_t - F_t)^2}{n}} \tag{1}$$

A_t is the actual data values, F_t is the forecasting value; n is the amount of data.
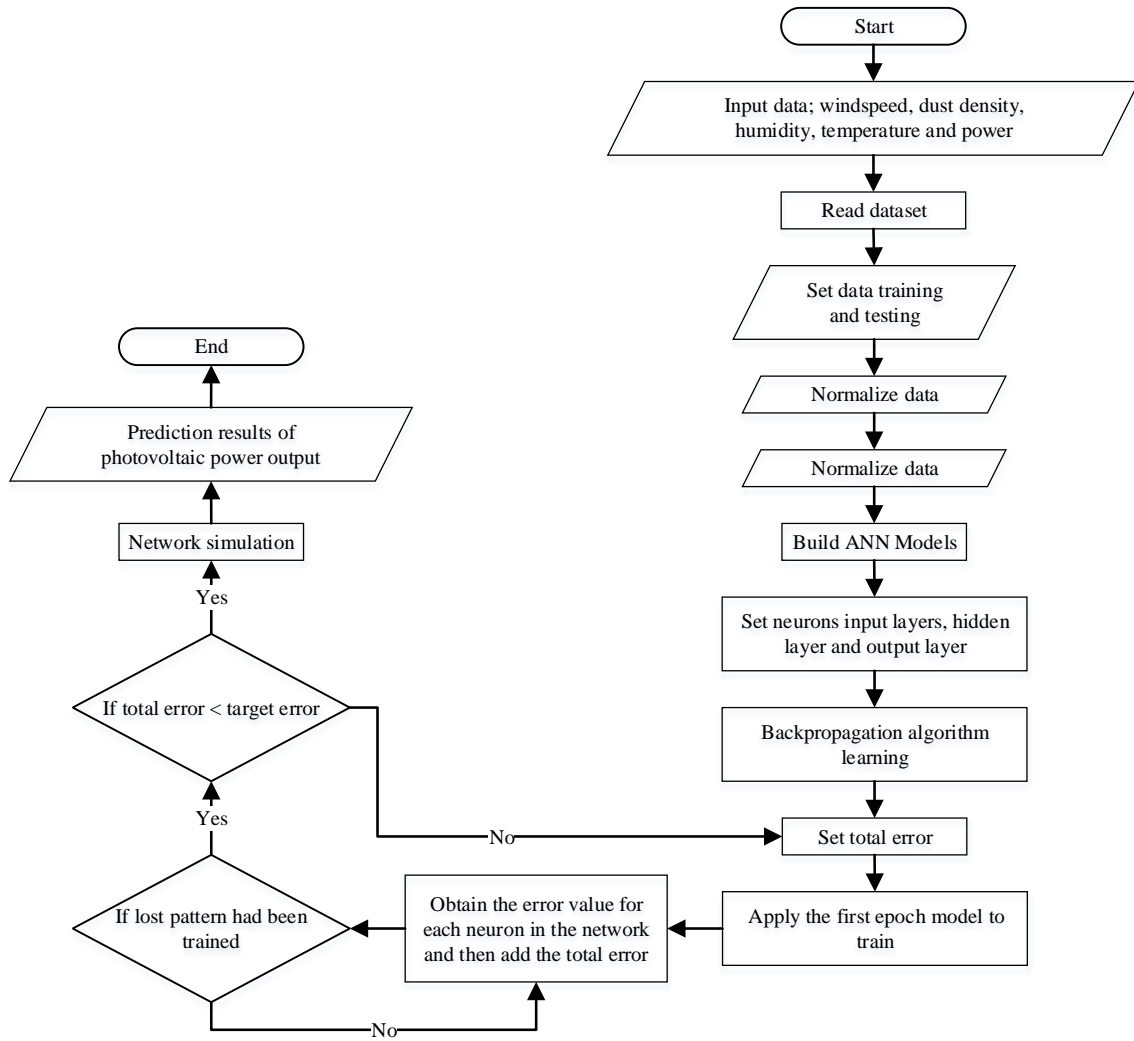
Figure 3. ANN Method Flowchart

Mean Squared Error (MSE) is the average value of the squared error between the actual value and the forecast value [19]. A Mean Squared Error that is low or close to zero indicates that the forecasting results are by the actual data and can be used for forecasting calculations in future periods. Formula 2 is for MSE[19].

$$MSE = \frac{\sum_{t=1}^{n}(A_t - F_t)^2}{n} \qquad (2)$$

A_t is the Actual data values; F_t is the Forecasting value; n is the amount of data.

Mean Absolute Error (MAE) is the average error value which is an error from the actual value with the predicted value [20]. MAE itself is generally used to measure error predictions in time series analysis. The formula for the MAE is as follows [21]:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}(Y_t' - Y_t)^2 \qquad (3)$$

Whereas: $Y_t' =$ Predictive Value; $Y_t =$ actual value; $n =$ amount of data.

3. Results and Discussions

The training data used in Figure 4 is the photovoltaic output power for 7 days. This data is used as training data to obtain a forecasting model using ANN.

Table 1 presents the statistical parameter values of environmental parameter data, including mean, variance, minimum value, first quartile, median, third quartile, maximum, skewness and kurtosis. In addition to environmental parameters, it also includes load voltage, current and photovoltaic output power. Skewness is a distribution characteristic that is used to explain the lack of symmetry in the distribution of data [22]. There are two types of skewness namely negative skew and positive skew [23]. Negative skew has a longer left tail and the mass of the distribution is concentrated on the right so that the distribution is said to be skewed to the left [24]. Conversely, positive skew has a longer right tail and the mass of the distribution is concentrated on the left, so the distribution is said to be skewed to the right [24].
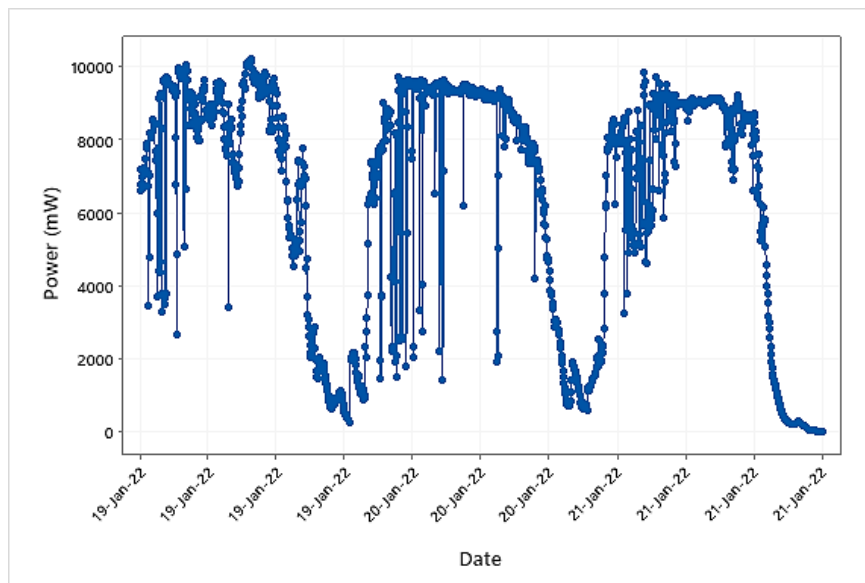
Figure 4. Photovoltaic Power Output

Table 1. Statistical Parameter Value of the data set

Variable	Mean	Variance	Minimum	First Quartile	Median	Third Quartile	Maximum	Skewness	Kurtosis
Wind speed	0.14998	0.10476	0	0	0	0.11	2.43	3.31	12.82
Dust Density	0.2971	0.00492	0	0.26	0.32	0.35	0.74	-0.87	1.86
Humidity	36.797	103.222	24	29	33	42	66	1.11	0.19
Temperature	31.341	19.067	22	28	33	35	39	-0.65	-0.64
Load Voltage	13.814	26.193	0	10.043	16.47	17.58	18.64	-1.14	-0.07
Current	406.29	22553.89	2.9	296.55	484.55	519.2	547.6	-1.12	-0.15
Power	6380.1	11179682	0	2994.9	7991	9115.5	10206.6	-0.76	-1.03

The variables of dust density, temperature, load voltage, current and power have negative skewness values meaning that the data for each variable is negatively distributed with the data spread more over the mean value. While the wind speed and humidity variables have a positive skewness value, meaning that they have a positive distribution, the distribution of data is below the mean value.

Kurtosis is the degree of peakedness of a distribution, usually taken relative to the normal distribution [25]. There are three types of kurtosis which are distinguished based on the height of the curve, namely leptokurtic which has a curve that is more pointed than the normal distribution (kurtosis value greater than 3), platykurtic which has a nearly flat peak (kurtosis value less than 3) and mesokurtic which has a moderate peak and not flat or normal (kurtosis value is equal to 3) [26], [27]. P In Table 1, the variables of dust density, temperature, load voltage, current and power have a kurtosis value of less than 3 meaning that the shape of the platykurtic curve with a rather even distribution of data with a low peak indicates no frequency in an extreme class. Meanwhile, the wind speed variable has a kurtosis value of more than 3, meaning that the shape of the leptokurtic curve with the data distribution is very running. This means that the data distribution is more

economical on average than the corresponding normal distribution [28].

3.1 Correlation test

Correlation testing is a test conducted to determine the relationship between two variables or determine the suitability between the two methods [29]. The correlation coefficient is used as a statistical parameter to determine the correlation between two variables. the correlation coefficient value has an interval between -1 to 1 [30]. correlation coefficient to tell the strength of the relationship between the two variables [31].

In this research, a correlation test to determine the magnitude of the correlation between the power output from the photovoltaic and the environmental parameters used, namely temperature, wind, humidity, and dust. In addition, it is also to determine the correlation between the power output from photovoltaics with current and voltage.

Figure 5 shows the correlation between the two parameters. If the correlation value is close to 1 then the photovoltaic power output and environmental parameters have a strong positive correlation, and vice versa if it is close to -1 then the photovoltaic power output and environmental parameters have a strong negative correlation. Whereas if the correlation value is

0 then the photovoltaic power output and environmental parameters do not correlate at all.

The scatter plot in Figure 5 shows that the power has an almost perfect correlation with current and voltage. Power with current and voltage has a linear distribution with a correlation value above 0.9. Power with

temperature has a strong correlation with a correlation value of 0.750. This shows that the power of photovoltaics is very influential on temperature. While the power with humidity and dust density has a negative correlation, each correlation value is -0.735 and -0.342. A negative correlation means that there is a reversed relationship between these variables.
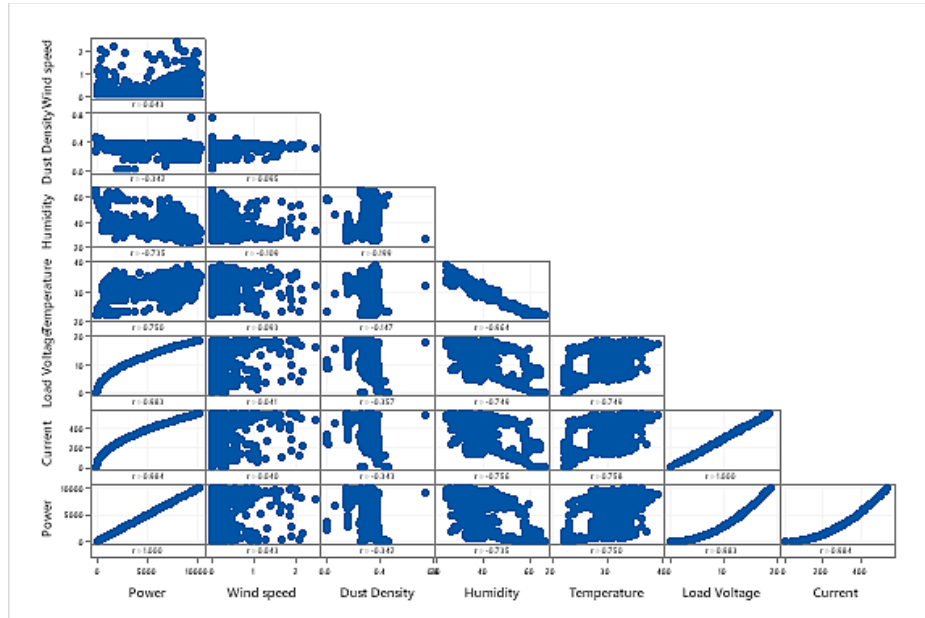

Figure 5. Correlation Test Results

3.2 Training of ANN models

In the training process using ANN, the selected epoch parameters are tested. The epoch values used are 10, 100 and 1000. In addition, the selected layers are 1 input

layer, 3 hidden layers and 1 output layer. Model evaluation is seen based on the MSE, MAE and RMSE parameter values obtained for each number selection. hidden layer neurons. Table 2 shows the results of tests with Epoch Values of 10, Epoch 100 and Epoch 1000.

Table 2. Test results with Epoch Value

Epoch Value	Number of Neurons in Hidden Layer 1	Number of Neurons in Hidden Layer 2	Number of Neurons in Hidden Layer 3	MSE	MAE	RMSE
Epoch 10	10	20	40	0,0042	0,0432	0,0650
	20	20	40	0,0080	0,0626	0,0896
	20	40	80	0,0035	0,0420	0,0593
	30	60	120	0,0025	0,0348	0,0504
	40	80	160	0,0018	0,0290	0,0421
Epoch 100	10	20	40	0,0026	0,0327	0,0507
	20	20	40	0,0029	0,0334	0,0542
	20	40	80	0,0014	0,0244	0,0374
	30	60	120	0,0012	0,0227	0,0350
	40	70	120	0,0012	0,0220	0,0344
Epoch 1000	10	20	40	0,0019	0,0290	0,0431
	20	20	40	0,0022	0,0312	0,0474
	40	70	120	0,0011	0,0206	0,0320
	30	60	120	0,0010	0,0184	0,0273
	40	80	160	0,0010	0,0155	0,0229

Based on Table 2, for each epoch value of 10, epoch value of 100, epoch value of 1000 and the number of neurons in the hidden layer, the best model is obtained, namely the number of neurons in hidden layer 1 is 40, hidden layer 2 is 80 and hidden layer 1 is 160. Thus, the

best epoch value is at epoch 1000 because it has smaller MSE, MAE, and RMSE values compared to epochs 10 and 100. This can be seen from the evaluation of the model the MSE, MAE and RMSE values obtain the lowest values compared to the number of neurons in

each hidden layer that has been selected. Furthermore, the evaluation of the model can also be seen based on the loss value at each selected epoch. The loss value is used to measure the error between the predicted output and the specified target.
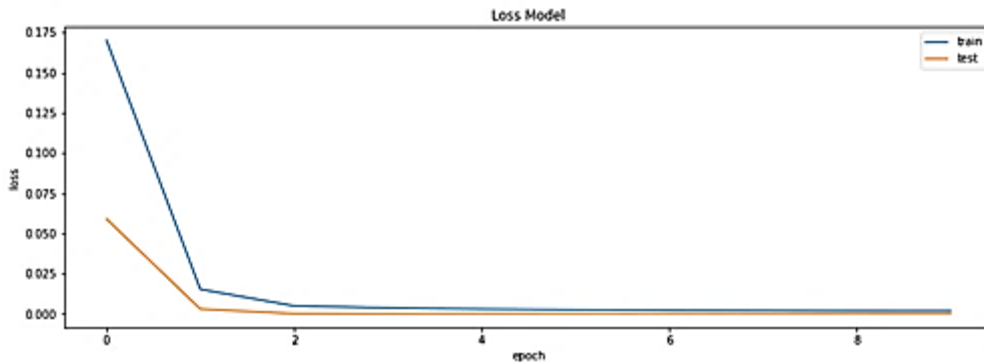

Figure 6. Testing of 10 epoch Values

Figure 6 shows a graph of the epoch against the loss value. The model tries to study the data by looking at the train line that follows the validation line to a value of 0 to epoch 10. The train model is parallel to the validation line and converges to 0. It can also be seen that the test data loss does not exceed the training data which shows that the model is fit with the training data.
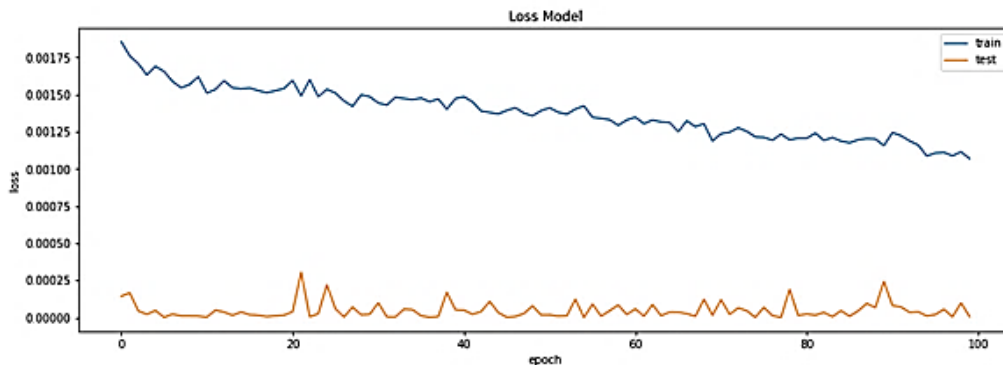

Figure 7. Testing of 100 epoch Values

Figure 7 shows a graph of the epoch against the loss value. The model tries to study the data by looking at the train line that follows the validation line towards a value of 0. However, starting from epoch 20 the loss on the train line seems to decrease closer to the value 0 while the test line fluctuates in the loss interval from 0 – 0.00025. In this model, it is also seen that the loss of the test data does not exceed the train data so it shows that the model is not overfit.
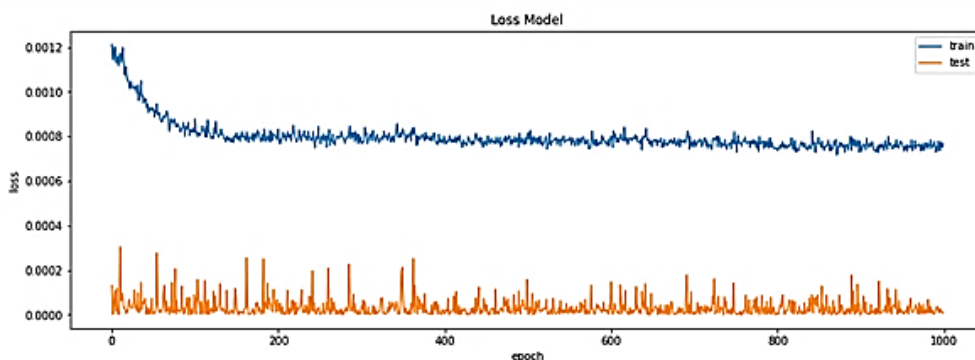

Figure 8. Testing of 1000 epoch Values

Figure 8 is a graph of the epoch against the loss value. The model tries to study the data by looking at the train line that follows the validation line to a value of 0. On the train line the loss value begins to decrease at epoch 150 and fluctuations occur at the loss value 0.0008 up to epoch 1000. The loss line of the test data does not exceed the train data so it shows that the model does not overfit.

3.3 Forecasting results

Based on the ANN model that has been trained, the ANN model was chosen with the best model evaluation results, namely epoch 1000 with a total of 40 hidden layer 1 neurons, 80 hidden layer 2 neurons and 150 of 150 hidden layer 3 neurons. This ANN model was selected to predict photovoltaic power because it has a lower MSE value.
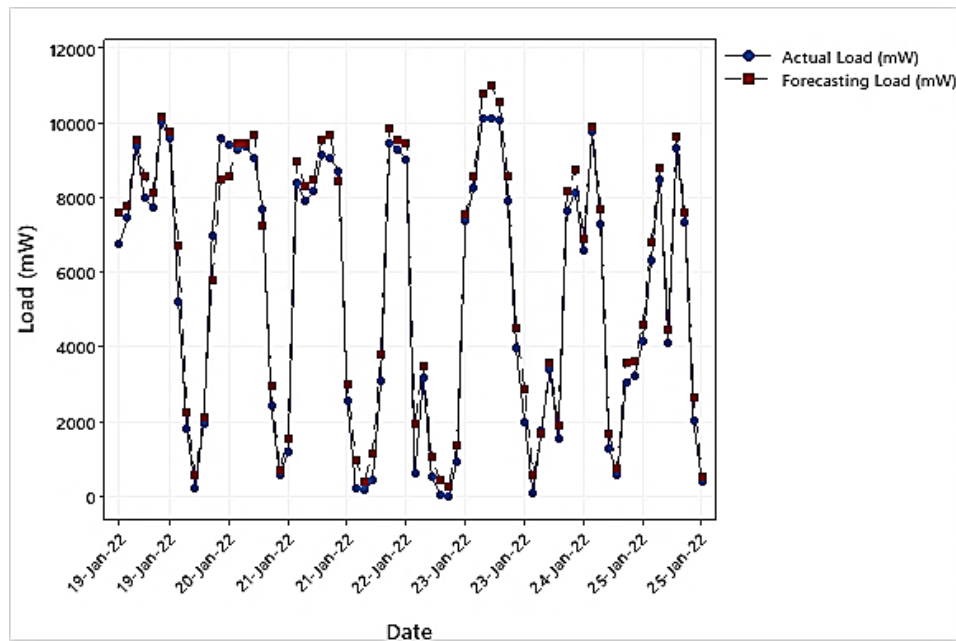

Figure 9. Prediction of the Photovoltaic Output Power

Figure 9 shows how the photovoltaic output power is predicted in the next 4 days. The x-axis is the amount of data and the y-axis is the power in mW units. Max power is 10000mW. The prediction results are close to the true value. Based on the resulting error, namely MSE of 0.086, MAE of 0.078, and RMSE of 0.065, this prediction can be said to be quite good. This ANN method uses the Epoch 1000 forecasting model because it has the lowest error value compared to other models. The results of forecasting the output of photovoltaic power can be used as a reference for building a solar power generation system that uses solar panels. This can be a future reference in analyzing the need for using solar cells using historical data on photovoltaic power output.

4. Conclusion

Based on the results of testing and research that have been done that is a good optimal model for predicting photovoltaic output power using the Artificial Neural Network method in the next 4 days with an Epoch value of 1000 because it has an error value of MSE 0.0012, MAE 0.0220, RMSE 0, 0344, the lowest error value compared to other models. In this study, five environmental parameters were used, namely wind speed, humidity, dust, humidity, and temperature. Of these 5 parameters, the most influential with photovoltaic panels is temperature. A temperature increase of 1 degree Celsius can cause a power increase of 0.5 to 1 W.

References

[1] S. S. Hegedus and A. Luque, "Status, Trends, Challenges and the Bright Future of Solar Electricity from Photovoltaics," in *Handbook of Photovoltaic Science and Engineering*, 2005.

[2] M. R. S. Shaikh, "A Review Paper on Electricity Generation from Solar Energy," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 5, no. 9, pp. 1884–1889, 2017, doi: 10.22214/ijraset.2017.9272.

[3] Y. K. Semero, J. Zhang, and D. Zheng, "PV power forecasting using an integrated GA-PSO-ANFIS approach and Gaussian process regression based feature selection strategy," *CSEE J. Power Energy Syst.*, vol. 4, no. 2, pp. 210–218, 2018, doi: 10.17775/CSEEJPES.2016.01920.

[4] H. Suyono, R. N. Hasanah, R. A. Setyawan, P. Mudjirahardjo, A. Wijoyo, and I. Musirin, "Comparison of solar radiation intensity forecasting using ANFIS and multiple linear regression methods," *Bull. Electr. Eng. Informatics*, vol. 7, no. 2, 2018, doi: 10.11591/eei.v7i2.1178.

[5] T. Hidayat, "Wind Power in Indonesia: Potential, Challenges, and Current Technology Overview," in *Indonesia Post-Pandemic Outlook: Strategy towards Net-Zero Emissions by 2060 from the Renewables and Carbon-Neutral Energy Perspectives*, 2022.

[6] H. D. Puspitarini, "Beyond 443 GW: Indonesia's infinite renewable energy potentials," 2021.

[7] A. Kumar Tripathi, M. Aruna, and C. S. N. Murthy, "Performance Degradation of PV Module Due to Different Types of Dust Pollutants," 2018.

[8] M. Cleef, P. Lippens, and J. Call, "Superior Energy Yields of UNI-SOLAR Triple Junction Thin Film Silicon Solar Cells compared to Crystalline Silicon Solar Cells under Real Outdoor Conditions …," pp. 22–26, 2001.

[9] J. Amajama and D. Oku, "Effect of Relative humidity on Photovoltaic panels' Output and Solar Illuminance/Intensity," vol. 2016, pp. 126–130, Aug. 2016.

[10] E. P. C. Edi Ismanto, "Jaringan Syaraf Tiruan Algoritma Backpropagation Dalam Memprediksi Ketersediaan Komoditi Pangan Provinsi Riau," *Rabit J. Teknol. dan Sist. Inf. Univrab*,

vol. 2, no. 2, 2017, doi: 10.36341/rabit.v2i2.152.

[11] R. Sabzehgar, D. Z. Amirhosseini, and M. Rasouli, "Solar power forecast for a residential smart microgrid based on numerical weather predictions using artificial intelligence methods," *J. Build. Eng.*, vol. 32, p. 101629, 2020, doi: https://doi.org/10.1016/j.jobe.2020.101629.

[12] K. B. Adam, D. K. Silalahi, B. S. Aprillia, and H. A. Adhari, "Support Vector Regression Method for Predicting Off-Grid Photovoltaic Output Power in the Short Term ," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 4 SE-Information Systems Engineering Articles, Aug. 2022, doi: 10.29207/resti.v6i4.4134.

[13] M. Haenlein and A. Kaplan, "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence," *Calif. Manage. Rev.*, vol. 61, no. 4, pp. 5–14, 2019, doi: 10.1177/0008125619864925.

[14] N. Djebbri and M. Rouainia, "Artificial neural networks based air pollution monitoring in industrial sites," in *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017*, 2018, vol. 2018-January, doi: 10.1109/ICEngTechnol.2017.8308151.

[15] R. Dastres and M. Soori, "Artificial Neural Network Systems," *International Journal of Imaging and Robotics (IJIR)*, 2021.

[16] T. Gonen, "Electric Power Distribution Engineering - Third Edition," *CRC Press Taylor and Francis Group*, vol. 1999, no. December. 2006.

[17] T. I. Hermanto, A. Idrus, L. Sugiyanta, D. Nasution, and I. Gunawan, "Neural Network Back-Propagation Method as Forecasting Technique," in *Journal of Physics: Conference Series*, 2022, vol. 2394, no. 1, doi: 10.1088/1742-6596/2394/1/012002.

[18] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geoscientific Model Development*, vol. 15, no. 14. 2022, doi 10.5194/gmd-15-5481-2022.

[19] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim. Res.*, vol. 30, no. 1, pp. 79–82, 2005, doi: 10.3354/cr030079.

[20] S. M. Robeson and C. J. Willmott, "Decomposition of the mean absolute error (MAE) into systematic and unsystematic components," *PLoS One*, vol. 18, no. 2 February 2023, doi: 10.1371/journal.pone.0279774.

[21] E. K. Jackson, W. Roberts, B. Nelsen, G. P. Williams, E. J. Nelson, and D. P. Ames, "Introductory overview: Error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption," *Environmental Modelling and Software*, vol. 119. 2019, doi: 10.1016/j.envsoft.2019.05.001.

[22] D. F. McCaffrey, "Volume 14: Quantitative Research and Educational Measurement," in *International Encyclopedia of Education(Fourth Edition)*, 2023.

[23] P. T. Von Hippel, "Mean, Median, and Skew : Correcting a Textbook Rule," *J. Stat. Educ.*, 2010.

[24] R. Shanmugam and R. Chattamvelli, "Skewness and Kurtosis," in *Statistics for Scientists and Engineers*, 2016.

[25] S. M. Gandhi and B. C. Sarkar, "Conventional and Statistical Resource/Reserve Estimation," in *Essentials of Mineral Exploration and Evaluation*, 2016.

[26] H.-Y. Kim, "Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis," *Restor. Dent. Endod.*, vol. 38, no. 1, 2013, doi: 10.5395/rde.2013.38.1.52.

[27] O. Calin, "Heat-flow transformation of a probability distribution," *Rev. Roum. Math. Pures Appl.*, vol. 65, no. 4, 2020.

[28] P. H. Westfall, "Kurtosis as Peakedness," *Am. Stat.*, vol. 68, no. 3, 2015.

[29] R. J. Janse *et al.*, "Conducting correlation analysis: important limitations and pitfalls," *Clin. Kidney J.*, vol. 14, no. 11, pp. 2332–2337, 2021, doi: 10.1093/ckj/sfab085.

[30] A. G. Dufera, T. Liu, and J. Xu, "Regression models of Pearson correlation coefficient," *Stat. Theory Relat. Fields*, vol. 0, no. 0, pp. 1–10, 2023, doi: 10.1080/24754269.2023.2164970.

[31] O. Esezi Isaac and A. Eric Chikweru, "Test for Significance of Pearson's Correlation Coefficient (r)," *Int. J. Innov. Math. Stat. Energy Policies*, vol. 1, no. 1, 2018.