JOIV : Int. J. Inform. Visualization, 8(1) - March 2024 141-149

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Predicting Battery Storage of Residential PV Using Long Short-Term Memory

R. Rizky Khaerul M.R. [a], Kurnianingsih [a,*], Amin Suharjono [a], I Ketut Agung Enriko [b], Naoyuki Kubota [c]

[a] *Department of Electrical Engineering, Politeknik Negeri Semarang, Tembalang, Semarang, 50275, Indonesia.*
[b] *Senior Manager IoT Platform, PT Telekomunikasi Indonesia, Mampang Prapatan, Jakarta, 12710, Indonesia.*
[c] *Department of Systems Design, Tokyo Metropolitan University, Hachioji-shi, Tokyo, 192-0397, Japan*
*Corresponding author: *kurnianingsih@polines.ac.id*

Abstract*— Solar power panels, or photovoltaic (PV), have recently grown rapidly as a renewable alternative energy source, especially since the increase in the basic electricity tariff. PV technology can be employed instead of the state electricity company to reduce the electricity used. Indonesia has great potential in producing electricity from PV technology, considering that most of Indonesia's territory gets sunlight for most of the year and has a large land area. Considering the benefits of PV technology, it is necessary to carry out predictive monitoring and analysis of the energy generated by PV technology to maximize energy utilization in the future. The Internet of Things (IoT) and cloud computing system was developed in this research to monitor and collect data in real-time within 27 days and obtained 7831 data for each parameter that affects PV production. These data include the light intensity, temperature, and humidity at the location where the PV system is installed. The feature selection results using Pearson correlation revealed that the light intensity parameter significantly impacted the PV production system. This research used the Long Short-Term Memory (LSTM) method to predict future PV production. By tuning hyperparameters using 3000 epochs, the resulting RMSE value was 171.5720. The results indicated a significant change in the RMSE value compared to 100 epochs of 422.5780. Given the increasing use of electric vehicles in the future, this model can be applied as a forecasting system model at electric vehicle charging stations.

***Keywords*— Forecasting; energy; photovoltaic; LSTM; Internet of Things.**

Manuscript received 25 Jan. 2023; revised 11 Mar. 2023; accepted 3 May. 2023. Date of publication 31 Mar. 2024.

I. Introduction

Electrical needs are expected to increase in the future, considering that electricity is one of the main factors supporting a country's economic growth. In the future, everyone will see the rapid growth and challenges of technological advances in electrical energy production, distribution, and utilization [1]. As electrical energy is predicted to increase rapidly, renewable energy is required to anticipate future energy crises—for instance, photovoltaics (PV), wind and water energy, geothermal, and biomass. PV is a power plant that converts sunlight into electric current. Solar energy is the most promising energy source because it is sustainable and massive. Indonesia has a high potential for solar energy because its climate allows most of its territory to receive sunlight all year. It means that the production of electrical energy from PV in Indonesia is guaranteed, both now and in the future. The average amount of sunlight emitted to the earth is about 1 KW/m2, equivalent to 1000 times the current energy consumption worldwide. In other words, covering only 0.1% of the Earth's surface with solar cells that have an efficiency of 10% can already cover the energy needs of the rest of the world nowadays [2]. Based on the predictions of the International Energy Agency (IEA), by 2050, the contribution of PV utilization in generating electrical energy globally will increase to 16% [3].

To store electrical energy useserated from PV, users on a residential scale commonly use batteries. Hence, unused energy is not wasted but can be used for future purposes. It can be carried out by injecting energy back into the grid when needed [4]. During generating and storing electrical energy, monitoring and analysis are required to find out the performance of the PV system and installed batteries. Therefore, users can find the suitability between the installed PV specification system and the electrical energy generated instead of analyzing the effect of environmental factors such as temperature, humidity, and light intensity.

The use of solar energy cannot be employed directly. However, it is necessary to have a process or series of systems in the form of additional tools called solar panels. In principle, the basic utilization of solar energy is to convert sunlight, which is then received by solar panels and converted into electrical energy [24]. Many researchers have conducted research on PV systems in terms of system performance, natural and environmental factors, efficiency, energy management, and system performance forecasting. In this paper, battery storage predictions using the *Long Short-Term Memory* (LSTM) method will be combined with IoT technology in monitoring and data acquisition in PV systems for household scale.

There are several difficulties and obstacles to manually recording data, i.e., it takes a long time and effort and has a low level of accuracy. Thus, in this research, an IoT and cloud computing-based automation technology was built to monitor and record the installed PV location's light intensity, temperature, and humidity data. IoT-based technology and cloud computing will easily access data within 24 hours, ensuring data security and reliability. The productivity performance of the PV system is affected by the environmental conditions in which the PV is installed. Therefore, this paper will discuss the factors that affect the performance of PV systems. Based on the results of recording and monitoring data for 27 days, the power generated through the battery charging process can be analyzed and predicted using the LSTM algorithm. The author used the LSTM algorithm because it works well on time series data.

II. Materials and Method

A. IoT For PV Performance Control and Monitoring

A PV performance monitoring system using IoT technology can be easily and efficiently implemented and is inexpensive. The system can also be connected directly to cloud computing for data storage and processing. This system can be applied on residential and industrial scales [5]. Not only used for monitoring, but IoT systems deployed on PV systems can also be added to control devices or other systems, e.g., in addition to monitoring PV performance or production, IoT systems can control pumps or other electronic devices automatically [6].

Liang et al. [7] built an IoT and cloud-based system to record weather data through two RS485 RTU modules. The system can be accessed in real-time and has alarm information; thus, system performance can be more effective and efficient. Lopez-Vargas et al. [8] discussed advanced large-scale Solar Home Systems (SHS) monitoring challenges. One of the discussions explained that IoT integration enables us to remotely monitor SHS and visualize data measured by dataloggers using a laptop or smartphone. Using an open-source IoT cloud platform minimizes the cost of monitoring systems.

The Internet of Things (IoT) consists of billions of devices for communication. The strength of this technology is that it can perform big data recording daily with various applications and fields of work. IoT can also be used to control or make decisions as a future technology supporting various sectors, especially business scale-up. In constructing deep learning methods that researchers widely use, the IoT is considered capable of meeting the needs of efficient large-scale data recording [9].

In 2017, Yerpude and Singhal [10] conducted a study on the impact of IoT data on demand forecasting. This research found that the role of IoT in the industrial revolution 4.0 is significant. The number of devices interacting and exchanging data over the network will increase significantly. The data obtained from the recording by IoT technology is useful to be analyzed using various forecasting methods.

B. Correlation Coefficient

Correlation is interconnected or influences each other or variables. In interrelated data, if values change to one of the data, the value of the interrelated data will also change, both by positive and negative correlations. Commonly, the term correlation is applied in the linear context of the continuous relationship between two variables and expressed as Pearson's product-moment correlation [11].

Correlation and regression are different methods. Regression is generally used for predictive purposes (not extrapolating beyond the data used in the analysis). Correlation is used to determine the degree of association, e.g., in situations where the variable x is not fixed or easily selected by the experimenter but is a random covariate to the variable y [12]. The correlation coefficient is a value that indicates how much of a linear relationship between two variables or how strongly the effect of one variable is on another. This correlation value usually ranges from -1 to 1. By a random variable (x, y), Pearson's correlation formula is expressed as follows:

$$r = \frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{[n\sum x^2-(\sum x)^2][n\sum y^2-(\sum y)^2]}} \quad (1)$$

The value is always between -1 and 1:

- +1 = a perfect positive relationship.
- 0 = no relationship.
- -1 = a perfect negative relationship.

TABLE I
INTERPRETATION OF CORRELATION COEFFICIENT VALUE

Correlation coefficient value	Interpretation
0 – 0.25	Very weak
0.25 – 0.50	Weak
0.50 – 0.75	Strong
0.75 – 1	Very strong

Based on Table I, if the correlation coefficient value is between 0 and 0.25, then the relationship between these variables is very weak. If the value is 0.25 to 0.5, the relationship between the variables is weak. Furthermore, if the value is between 0.5 and 0.75, the relationship between the variables is strong. If 0.75 is greater than 1, then the relationship between these variables is firmly strong.

C. Forecasting on PV System Production

Pawar et al. [1] have developed artificial intelligence systems to predict the availability of renewable energy sources accurately. The research proposed several comparisons of predictive forecasting models for energy forecasting. As a result, the PSO-based SVM regression model outperforms several other prediction models regarding performance accuracy. In addition to making predictions using intelligent systems, this research utilizes IoT technology to monitor the system environment.

In 2018, A. Elamim et al. [13] conducted a study by predicting the power generated by photovoltaic solar panels using a *Feed Forward Neural Network* (FFNN) with two input parameters: temperature and solar radiation. The model was tested in sunny, cloudy, and rainy conditions. The test results of this model on the data show that the model works well, with a coefficient of determination between 0.99 and 0.998 for sunny days, 0.961 and 0.965 for cloudy days, and between 0.88 and 0.93 for rainy conditions.

In 2020, Li et al [14] conducted a study on the effect of light intensity on PV performance. The experimental results reveal that the open-circuit voltage, short-circuit current, and maximum output power of solar cells increase with increasing light intensity. Qing and Niu [15] proposed a solar radiation prediction scheme for predicting solar irradiation every hour using weather forecast data. The proposed prediction model employs *Long Short-Term Memory* (LSTM) networks. They compared persistence algorithms, the smallest squares linear regression, and layered feedforward neural networks using backpropagation algorithms (BPNN) for solar radiation prediction. The proposed algorithm results outperform other competitive algorithms.

In 2020, Touati et al. [16] researched PV system monitoring. They analyzed PV system performance using the *Power Prediction Model* conducted in Doha, Qatar. The results state that the collected data sets, using certain ML algorithms such as Artificial Neural Networks (ANN), can successfully provide accurate predictions with a low *Root-Mean-Squared Error* (RMSE) error between the predicted and actual power; hence, this model can support efficient energy planning and management, especially in arid and semi-arid areas.

In 2012, Zeng and Qiao [17] researched short-term solar power prediction using the Support Vector Machine by proposing a model based on the least-square Support Vector Machine (SVM) (LS) for Solar Predictive Power (SPP) in the short term. As a result, the proposed model significantly outperforms the Autoregressive Reference (AR) model and has better prediction accuracy than the Radial Basis Function Neural Network (RBFNN).

D. Data Collection

Data collection was carried out to collect information about environmental conditions at the location of the installed PV system, such as temperature, humidity, and light intensity, as well as the production of power generated by the PV system. The data was obtained by building an IoT system using an ESP32 microcontroller, a DHT11 sensor for temperature and humidity detection, and a BH1750 light intensity sensor to obtain data on the intensity of the sunlight shining on the PV panel. The battery's current, voltage, and charging power data are obtained from the inverter module of the installed PV system. Then, the data collected using these sensors is processed using ESP32 and sent directly to the cloud computing system using an internet connection to be stored and accessed in *real-time* for monitoring purposes. The data recording process was carried out for 27 days, resulting in approximately 7831 timestamps for each parameter. It is necessary to pay attention to the specified timestamp in recording data. In this research, data recording was carried out every five minutes, adjusted to the timestamp obtained from the PV inverter, which produced data on charger power, PV voltage, and PV current. Data obtained from sensors cannot be directly used for forecasting. However, depending on the data model required for forecasting, they must go through data preprocessing stages such as data cleansing, transformation, and integration. The system architecture for data recording in a PV environment using IoT and cloud computing technologies is shown in Fig. 1.

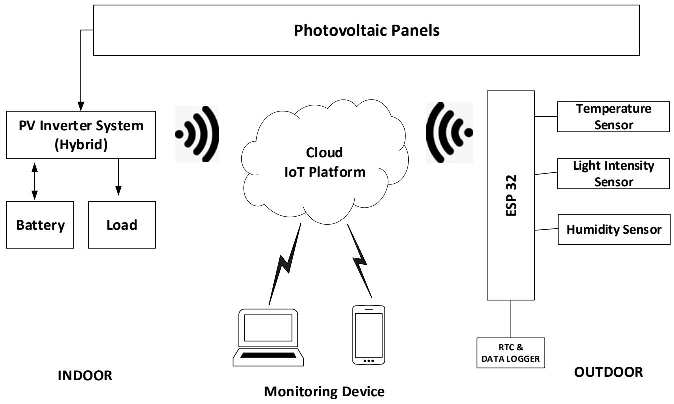


Fig. 1 The system architecture of data collection

E. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture [18]. Neural network LSTM has greatly contributed to solving real problems in various fields [19]. The LSTM algorithm can work better than RNN because it has a more complex cell structure. In this research, the model used data from previous PV production to predict the power generated by the PV system in the future. Hence, LSTM is suitable for studying past data to predict the future. LSTM has a great ability to study long-term dependence as well as remember information over a long period [20].

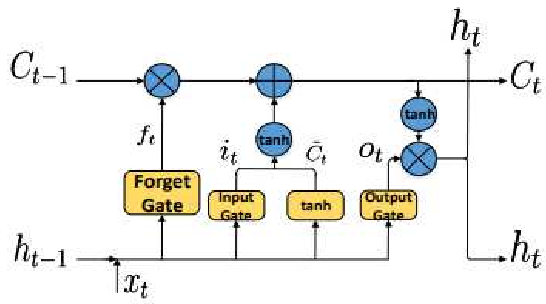


Fig. 2 The Structure of LSTM Cell [19]

The specific transitions in LSTM cells are shown in Fig. 2. Each yellow block is a neural network layer that reads data from h_{t-1} and x_t . A blue block is a calculative unit for vector operations [19]. The first starts via forget gate (f_t). In this step, if information or value is not needed for the case to be processed, then the information or value is omitted; in other words, forgotten using the sigmoid function [21]. x_t is input data (input vector x on timestep t), and h_{t-1} is the hidden state vector on timestep t-1 earlier). In the next step, the information is processed through the input gate (i_t). This process will sort and determine certain information to be updated to the state part of the cell using the sigmoid activation function [21] and construct a new perspective vector using the tan h activation function to be added to the cell state section ($\sim C_t$). Furthermore, update the old cell state value (C_{t-1}) to the new cell state (C_t). Lastly, it occurs to the output gate component by running sigmoid to generate the output value in a hidden state and placing the state of the cell on the *tan h*. After generating the sigmoid output value and the *tan h* output value, the two activation results are multiplied before going to the next step, producing the classification value of all calculations in the LSTM process [22].

Neurons in the structure of the LSTM architecture consist of forgetting gates, input gates, new cell states, and output gates. Each of those parts can regulate the memory of each neuron. Each unit is like a mini-state machine; the gate unit has a different weight learned during training. It allows LSTM to process a large amount of time series data [19].

F. Forecasting using Long Short-Term Memory (LSTM)

In this paper, the LSTM method is employed for forecasting time series data and creating a single prediction; in this case, it is charger power data or battery charging power from the production of PV systems. The flowchart for the forecasting process is visualized in Fig. 3. The first step is to preprocess the raw sensor data. Then, the data that has passed this step is divided into three parts: 80% data for training, 10% for validation, and 10% for testing. Using the LSTM structure, it is then determined how to conduct training and forecasting to get the best model. After getting the best model, the model is tested using pre-prepared test data. In this research, the data training and model simulation process was carried out once. However, some training is carried out with parameters, such as epoch and learning rate, to obtain a model with good data processing efficiency. Therefore, the best model was obtained from several simulations.

G. Performance Evaluation

To evaluate the performance of the LSTM model, it uses MAE (*Mean Absolute Error*). MAE is the average absolute difference between actual and forecast values. RMSE (*Root Mean Squared Error*) was used to measure errors in the forecasting process.

$$MAE = \frac{\sum_{i=1}^{n} |yi - xi|}{n} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (A_t - F_t)^2}{n}} \quad (3)$$

MAE and RMSE are the two most commonly used metrics to measure the accuracy of continuous variables. MAE and RMSE reveal the prediction error of the average model in units of variables of interest. Both metrics can range from 0 to ∞. Both are negatively oriented scores, which means lower scores are better [23]. RMSE gives a relatively high weight for large errors. It means RMSE should be more useful when major errors are highly undesirable.

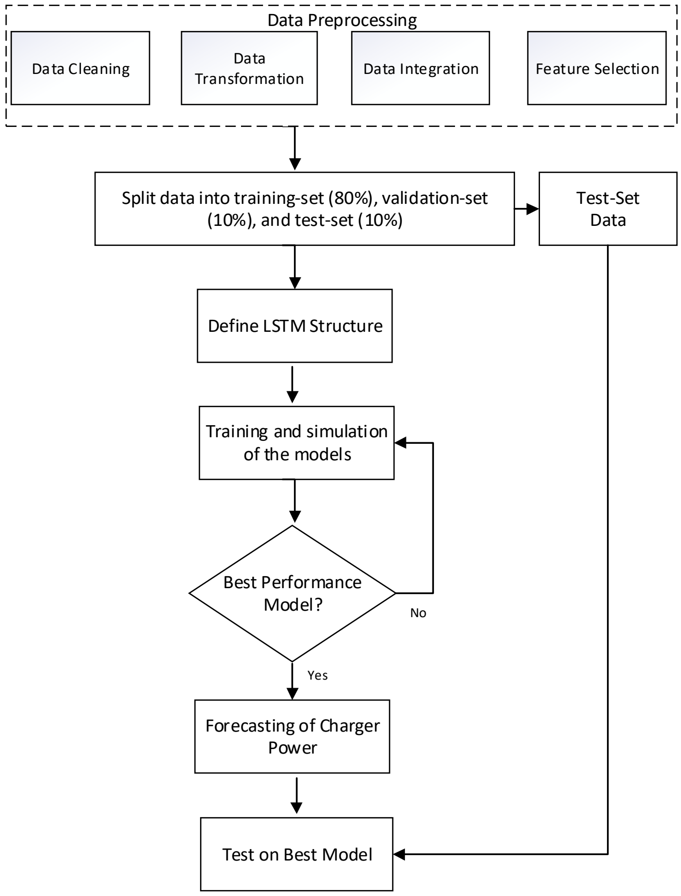


Fig. 3 Forecasting flowchart Charging in PV system.

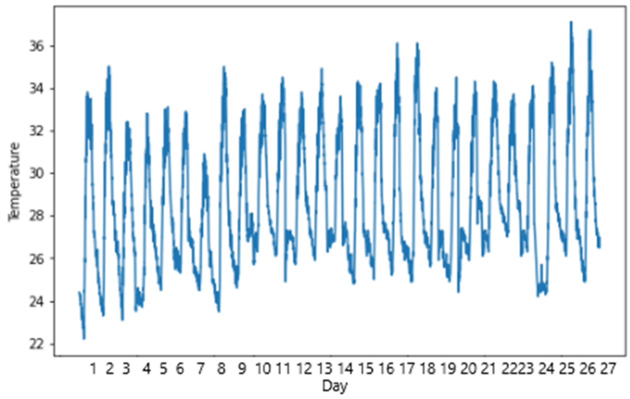
III. Results and Discussion

A. IoT-based Weather Monitoring on PV System

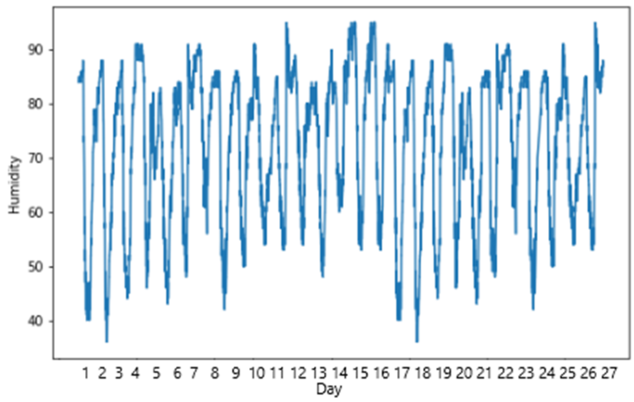
In this research, data monitoring and recording were carried out using IoT systems and cloud computing. Data can be monitored in *real-time* anytime and from anywhere for 24 hours using devices such as laptops or smartphones. The data monitoring process is visualized in Fig. 4. The sensor reading data is directly transferred to cloud computing for storage and can be read anytime. It indicates that IoT systems enable users to manage monitoring and data recording systems [9]. The IoT system developed in this research used low-cost components; thus, if used for recording long-term and large data, it will save more costs. It is necessary to maintain the IoT system and check periodically when there is a disturbance in the system, both from environmental, natural, and system factors.

Fig. 4 IoT-based weather monitoring on PV system

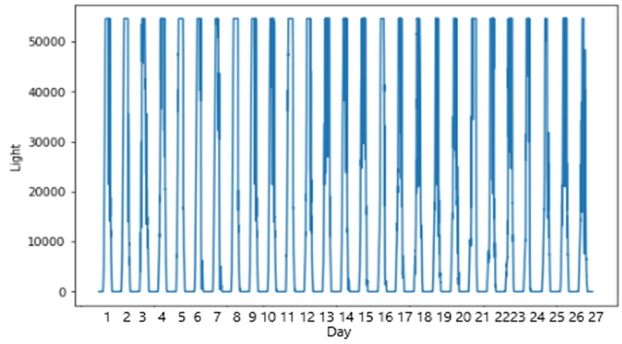
After obtaining the dataset and preprocessing the data, a preliminary investigation was conducted to find patterns and anomalies. It can help with hypothesis testing and provide statistical assumptions based on visually presented data. Fig. 5 (a – f) presents a data visualization for each parameter for 27 days.



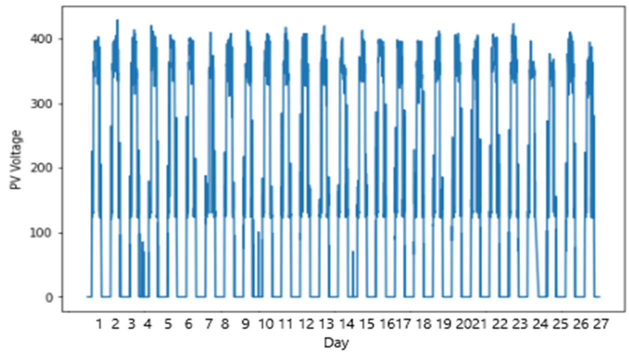
(a) PV system ambient temperature values



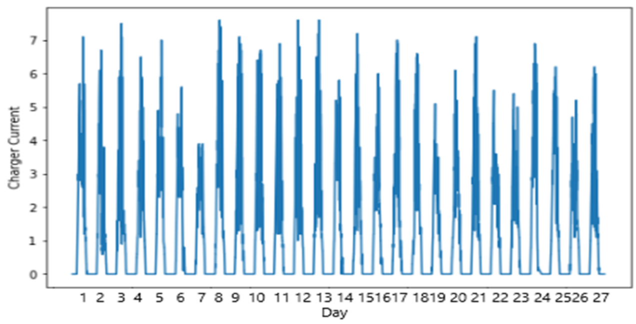
(b) Humidity values in the PV system environment



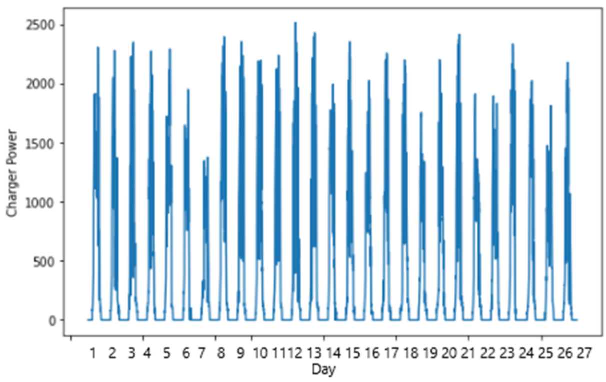
(c) Light intensity values in the PV system environment



(d) PV voltage values



(e) PV current values



(f) PV charger power values

Fig. 5 Parameters that affect solar panels for: a) temperature, b) humidity, c) light intensity, d) PV voltage, e) charger current, f) charger power.

B. Influence of Weather on PV Performance

Based on a basic theory of correlation coefficients described by formula 1 (r) and analyzed by the panda's library in Python programming, it can be known how much each weather parameter affects PV performance. Fig. 6 shows the relationship between weather parameters such as light intensity, temperature, and humidity to PV performance, such as charger power.

Furthermore, Fig. 6 explains that light intensity has a strong positive effect on the production of PV systems, as evidenced by a value of 0.87. Then, temperature also strongly affects the production of PV systems, with a value of 0.73. Humidity has a value of -0.72; humidity has a strong negative effect on the production of PV systems. Based on the analysis, it can be concluded that the higher the value of sunlight intensity and temperature in the environment around the PV system, the better the impact will be on the production performance of the PV system. Conversely, if the humidity value around the PV system environment is higher, it will decrease the production of the PV system.

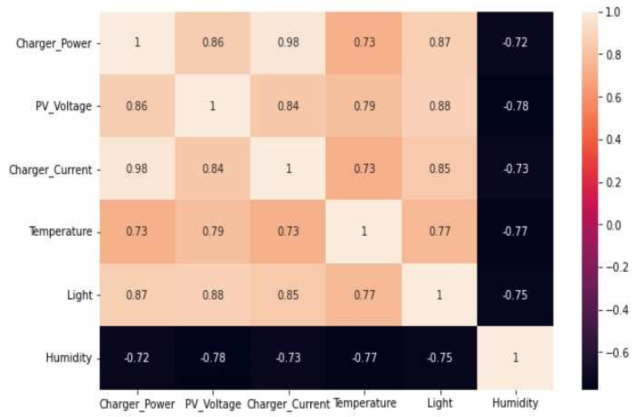


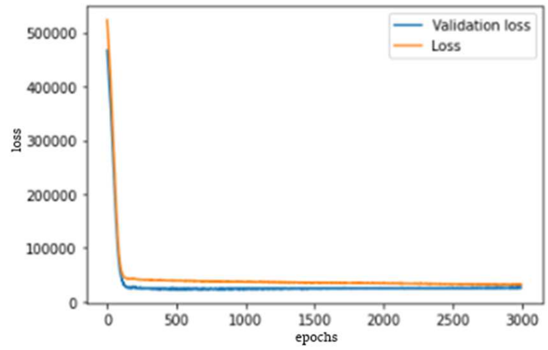
Fig. 6 Influence of Weather on PV Performance

C. Hyperparameters of LSTM

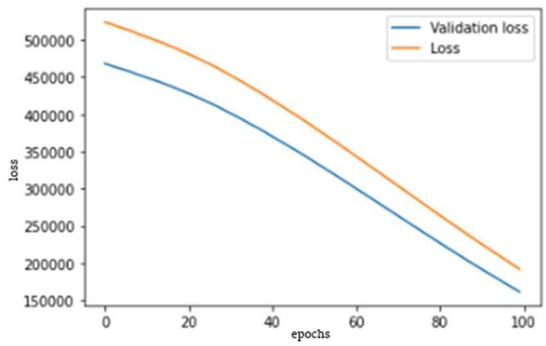
The structure of the LSTM hyperparameters used in this research is presented in Table 2. The LSTM model consists of one (1) hidden layer, 64 neurons, five values to look back at, and relu as an activation function. The number of epochs or iterations used during training was 3000 epochs. Then, the learning rate used was 0.0001. The loss function was the *mean squared error* (MSE), and as an optimizer, it was used to reduce losses using Adam.

TABLE II
HYPERPARAMETERS OF LSTM

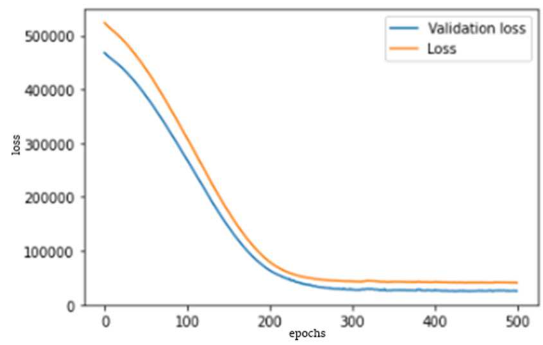
Parameter	Value
Number of hidden layers	1
Number of neurons	64
Look back	5
Activation function	relu
Number of epochs	3000
Learning rate	0.0001
Loss function	MSE
Optimizer	Adam



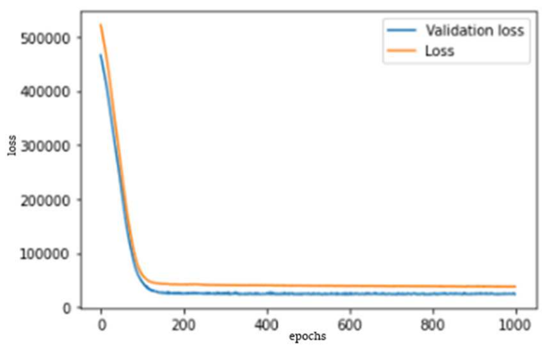
(a) Model loss using 3000 epochs



(b) Model Loss using 100 epochs



(c) Model Loss using 500 epochs



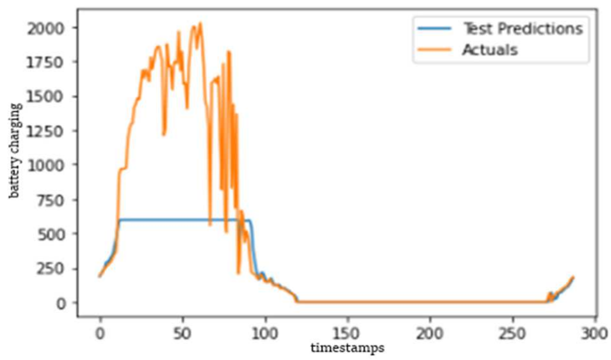
(d) Model Loss using 1000 epochs

Fig. 7 Model Loss: a) 3000 epochs, b) 100 epochs, c) 500 epochs, d) 1000 epochs

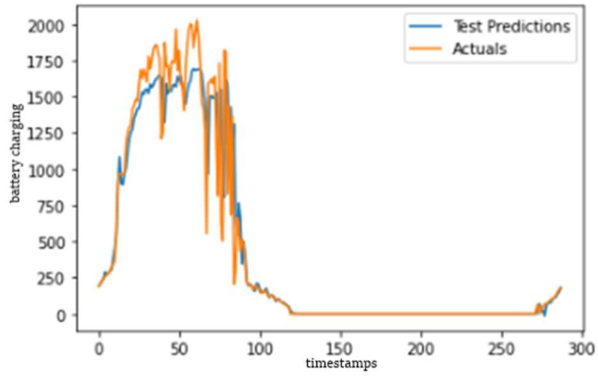
In this study, several training processes were carried out using tuning parameters to produce the best and most efficient model. Using the hyperparameters presented in Table 2, the best version of the model is generated. Fig. 7 (a) presents a loss model during training using 3000 epochs. For several times running the training model process, with different parameters, the parameters that most influence the loss model results are the number of epochs and the value of the learning rate.

From this case, the higher the number of epochs, the lower the loss model value, and the smaller the learning rate, the greater the level of accuracy of the training process, but consequently, the time required to carry out the training process will be longer. Fig. 7 (b) presents the model loss during the training process with an epoch parameter of 100 epochs. From the figure, it can be analyzed that there is still a possibility of decreasing loss for the next training epoch, so the author performs the training process again by changing the epoch parameters. Fig. 7 (c) presents the model loss during the training process with an epoch parameter of 500 epochs. From the picture, it can be analyzed that the decrease in loss has started to show a sloping curve, but according to our assumptions, there is still the possibility of getting a lower loss for the next epoch. Then the retraining process is carried out by increasing the number of training epochs. Fig. 7 (d) presents the model loss during the training process with an epoch parameter of 1000 epochs.

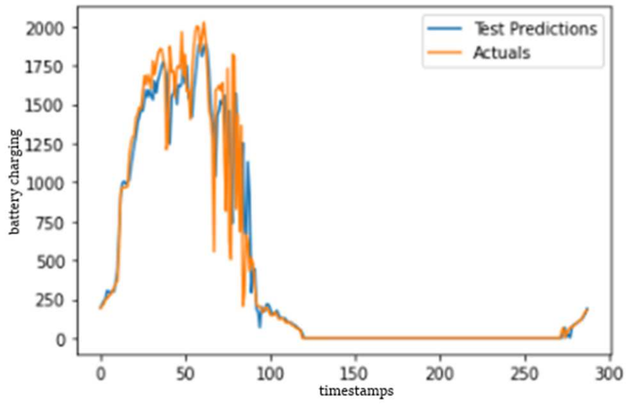
From the figure, it can be analyzed that the decrease in loss has started to show a sloping curve, but according to our assumptions, there is still the possibility of getting a lower loss for the next epoch. Then the retraining process was carried out by increasing the number of training epochs to 3000 epochs. By using 3000 epochs, the result is that the loss curve of the training model begins to slope and does not show a significant loss reduction, as shown in Fig. 7 (b-d). Thus, it can be assumed that the best and most efficient model is to use 3000 epochs.



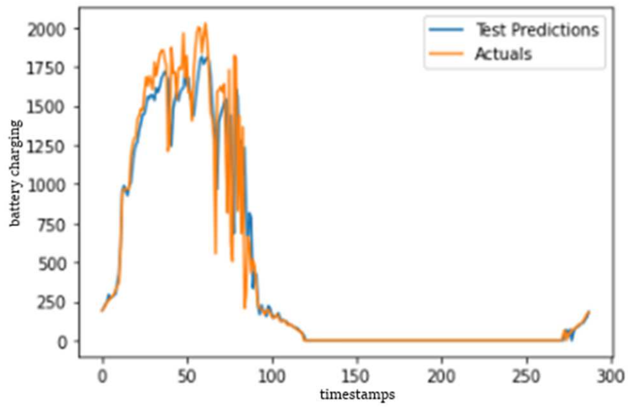
(a) Test predictions vs actuals data using 100 epochs



(b) Test predictions vs actuals data using 500 epochs



(c) Test predictions vs actuals data using 1000 epochs



(d) Test predictions vs actuals data using 3000 epochs

Fig. 8 Test predictions vs actuals data using: a). 100 epochs, b). 500 epochs, c). 1000 epochs, d). 3000 epochs

Fig. 8 (a) visualizes a prediction test with actual data using a training model with a total of 100 epochs. As a result, the accuracy rate is still far from expectations. Many data show discrepancies or inequalities between prediction data and actual data. Hence, it can be assumed, the model is inadequate and worthless by using 100 epochs. Fig. 8 (b) visualizes a prediction test with actual data using a training model of 500 epochs. Fig. 8 (b) visualizes a prediction test with actual data using a training model of 500 epochs. As a result, the accuracy rate is close to expectations. More data shows a match or similarity between the prediction and actual data. It can be assumed that the training process is better than using 100 epochs. However, it has not achieved optimal results. Fig. 8 (c) visualizes a prediction test with actual data using a training model with a total of 1000 epochs. As a result, the accuracy rate is close to expectations. More data shows a match or similarity between the prediction and actual data. It can be assumed that the training process is better than using 100 & 500 epochs. Fig. 8 (d) visualizes a prediction test with actual data using a training model with a total of 3000 epochs. As a result, the accuracy rate is close to expectations. More data shows a match or similarity between the prediction and actual data. The results obtained did not differ markedly between 500 and 1000 epochs.

TABLE III
PV SYSTEM CHARGING TEST RESULTS BASED ON THE NUMBER OF EPOCHS

Number of epochs	RMSE
100	422,5780
500	174,2785
1000	172,3147
3000	171,5720

Table III explains that in this research, four pieces of training were carried out with the number of epochs as high as 100, 500, 1000, and 3000, and obtained values of RMSE of 422.5780 for 100 epochs, 174.2785 for 500 epochs, 172.3147 for 1000 epochs, and 171.5720 for 3000 epochs. From this data, the greater the number of epochs, the smaller the RMSE value. In the case of the research data used, the difference between the RMSE values for epochs 500, 1000, and 3000 did not differ markedly. Hence, the best model for some experiments is 3000 epochs.

Furthermore, Table IV provides examples of some test data predictions with actual data and error percentages. The test used the best model of several experiments using a learning rate of 0.0001 and 3000 epochs. The results showed quite good performance for the LSTM model created. From table 4, the average model has a small error percentage of 5.25% and an MAE of 34.82015.

TABLE IV
EXAMPLES OF TEST DATA PREDICTIONS WITH ACTUAL DATA USING 3000 EPOCHS

Test Predictions	Actuals	Error (%)	MAE
189.0410	191.0000	1.03	1.959
207.9650	206.0000	0.95	1.965
230.4220	218.0000	5.70	12.422
246.3830	248.0000	0.65	1.617
292.1090	255.0000	14.55	37.109
268.9897	270.0000	0.37	1.0103
277.9241	282.0000	1.45	4.0759
286.0820	294.0000	2.69	7.918
297.1844	327.0000	9.12	29.8156
367.9642	346.0000	6.35	21.9642
426.6275	370.0000	15.30	56.6275
604.5535	582.0000	3.88	22.5535
955.2527	934.0000	2.28	21.2527
988.5925	968.0000	2.13	20.5925
962.3010	967.0000	0.49	4.699
926.6661	968.0000	4.27	41.3339
992.9292	974.0000	1.94	18.9292
1.005.4883	1176.0000	14.50	170.5117
1.116.9244	1252.0000	10.79	135.0756
1.206.0286	1291.0000	6.58	84.9714

The results showed that data recording using a cloud computing-based IoT system has several advantages: (1) the data obtained has a fairly high level of accuracy, and it is easy to determine the timestamp of each data retrieval simultaneously for each data parameter needed; (2) by using IoT, users can automate and set the time when data should be recorded and to be stopped. It is one of IoT technology's advantages in data recording [9]; (3) the IoT system proposed in the research is low-cost and easy to implement; (4) the data storage system, by utilizing cloud computing technology, makes the stored data more secure, has high reliability, and can be accessed in real-time anytime and anywhere by users with access rights. The system also supports big data recording.

Based on the results of data recording using IoT technology, which is then processed and analyzed using the correlation coefficient method using Python programming, in this case, it uses the NumPy and Pandas libraries; hence, it can be concluded that the weather conditions around solar panels with a greatest positive effect on the production of PV systems are light intensity parameters. Therefore, it can be stated that the higher the intensity value of sunlight received by solar panels, the better the production of PV systems will be. It is in line with the correlation coefficient theory discussed in the correlation coefficient article [11].

The forecasting process in this research employed the LSTM method to find out the prediction of PV system production. The best forecasting model was obtained by conducting several experiments with various hyperparameter settings. The model results were still far from our expectations when the model training was first carried out, using 100 epochs. The prediction test was still far from the actual data. Based on the loss model, there is still a possibility of reducing loss. Then, the model was retrained by adding epochs with variations of 500, 1000, and 3000 epochs; According to the three models, the results showed a significant decrease in losses. The RMSE value gets smaller as the epoch increases, and the prediction test results get closer to accuracy.

IV. CONCLUSION

This research proposes two contributions: (1) Data monitoring and recording systems for forecasting purposes can be applied by building low-cost IoT systems. It can improve the accuracy of the data obtained and facilitate human tasks compared to manual data recording. The recorded data has high security and reliability because it is stored in the cloud and can be accessed anytime and anywhere by anyone with access rights. (2) An LSTM architecture system is used to forecast PV system production from datasets generated from records using cloud computing-based IoT systems. The data obtained reveals a firmly strong positive correlation between the intensity of sunlight and the production of PV systems. The longer the maximum and consistent exposure to sunlight, the better the PV system production. The LSTM model resulting from this study is good enough to forecast PV system production data. The quality of the forecasting results may not show results with a very high level of accuracy because the datasets used are not much.

Several things need to be improved to improve the quality of future research results, both in data recording and developing forecasting models. A data recording system can be developed into a two-way control system between the user and the system. If there are problems in the field related to the system, it would be better if there was an automatic notification message to the user. Thus, if the data recording system fails, it can be fixed quickly and prevent further data loss. Improving forecasting accuracy can be improved by increasing the number of datasets and tuning hyperparameters of the model architecture. If you want to be more varied, data recording should be carried out in several places with different climatic conditions.

References

[1] P. Pawar, M. TarunKumar, and P. Vittal K., "An IoT based Intelligent Smart Energy Management System with accurate forecasting and load strategy for renewable generation," Measurement, vol. 152, p. 107187, Feb. 2020, doi: 10.1016/j.measurement.2019.107187.

[2] P. M. Fishbane, S. Gasirowicz, and S.T. Thornton: *Physics for scientists and engineers*, 2nd edition, Prentice-hall, New Jersey, 1996.

[3] R. Hasan, S. Mekhilef, M. Seyedmahmoudian, and B. Horan, "Gridconnected isolated PV microinverters : A review," *Renewable and Sustainable Energy Reviews*, Vol. 67, pp.1065-1080, 2017.

[4] A. A. Jamali, N. M. Nor, and T. Ibrahim, "Energy storage systems and their sizing techniques in power system — A review," 2015 IEEE Conference on Energy Conversion (CENCON), Oct. 2015, doi:10.1109/cencon.2015.7409542.

[5] V. Kavitha and V. Malathi, "A Smart Solar PV Monitoring System Using IoT," First International Conference on Secure Reconfigurable Architectures & Intelligent Computing (SRAIC 2019), Nov. 2019, doi: 10.5121/csit.2019.91502.

[6] O. Chieochan, A. Saokaew, and E. Boonchieng, "Internet of things (IOT) for smart solar energy: A case study of the smart farm at Maejo University," 2017 International Conference on Control, Automation and Information Sciences (ICCAIS), Oct. 2017, doi:10.1109/iccais.2017.8217588.

[7] R. Liang, Y. Guo, L. Zhao, and Y. Gao, "Real-time monitoring implementation of PV/T façade system based on IoT," Journal of Building Engineering, vol. 41, p. 102451, Sep. 2021, doi:10.1016/j.jobe.2021.102451.

[8] A. López-Vargas, M. Fuentes, and M. Vivar, "Current challenges for the advanced mass scale monitoring of Solar Home Systems: A review," Renewable Energy, vol. 163, pp. 2098–2114, Jan. 2021, doi:10.1016/j.renene.2020.09.111.

[9] K. Lakshmanna et al., "A Review on Deep Learning Techniques for IoT Data," Electronics, vol. 11, no. 10, p. 1604, May 2022, doi:10.3390/electronics11101604.

[10] S. Yerpude and T. K. Singhal, "Impact of Internet of Things (IoT) Data on Demand Forecasting," Indian Journal of Science and Technology, vol. 10, no. 15, pp. 1–5, Apr. 2017, doi:10.17485/ijst/2017/v10i15/111794.

[11] P. Schober, C. Boer, and L. A. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation," Anesthesia & Analgesia, vol. 126, no. 5, pp. 1763–1768, May 2018, doi:10.1213/ane.0000000000002864.

[12] A. G. Asuero, A. Sayago, and A. G. González, "The Correlation Coefficient: An Overview," Critical Reviews in Analytical Chemistry, vol. 36, no. 1, pp. 41–59, Jan. 2006, doi:10.1080/10408340500526766.

[13] A. Elamim *et al*., "Photovoltaic Output Power Forecast Using Artificial Neural Networks," *Journal of Theoretical and Applied Information Technology,* vol. 96(15): 5116 – 5126, 2018.

[14] Z. Li, J. Yang, and P. A. N. Dezfuli, "Study on the Influence of Light Intensity on the Performance of Solar Cell," International Journal of Photoenergy, vol. 2021, pp. 1–10, Feb. 2021, doi:10.1155/2021/6648739.

[15] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," Energy, vol. 148, pp. 461–468, Apr. 2018, doi:10.1016/j.energy.2018.01.177.

[16] F. Touati, A. Khandakar, M. E.H. Chowdhury, A. Jr. S.P. Gonzales, C. Kim Sorino, and K. Benhmed, "Photo-Voltaic (PV) Monitoring System, Performance Analysis and Power Prediction Models in Doha, Qatar," Renewable Energy - Technologies and Applications, Feb. 2021, doi:10.5772/intechopen.92632.

[17] J. Zeng and W. Qiao, "Short-term solar power prediction using a support vector machine," Renewable Energy, vol. 52, pp. 118–127, Apr. 2013, doi:10.1016/j.renene.2012.10.009.

[18] A. Graves, "Generating Sequences With Recurrent Neural Networks," *arXiv*, :1308.0850, 2014, doi:10.48550/arXiv.1308.0850.

[19] R. Zhang and Q. Zou, "Time Series Prediction and Anomaly Detection of Light Curve Using LSTM Neural Network," Journal of Physics: Conference Series, vol. 1061, p. 012012, Jul. 2018, doi:10.1088/1742-6596/1061/1/012012.

[20] Kurnianingsih, A. Wirasatriya, L. Lazuardi, N. Kubota, and N. Ng, "IOD and ENSO-Related Time Series Variability and Forecasting of Dengue and Malaria Incidence in Indonesia," 2020 International Symposium on Community-centric Systems (CcS), Sep. 2020, doi:10.1109/ccs49175.2020.9231358.

[21] P. W. Hardjita, Nurochman, and R. Hidayat, "Sentiment Analysis of Tweets on Prakerja Card using Convolutional Neural Network and Naive Bayes," IJID (International Journal on Informatics for Development), vol. 10, no. 2, pp. 82–91, Jan. 2022, doi:10.14421/ijid.2021.3007.

[22] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi:10.1162/neco.1997.9.8.1735.

[23] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: an analysis and review," International Journal of Forecasting, vol. 16, no. 4, pp. 437–450, Oct. 2000, doi:10.1016/s0169-2070(00)00065-0.

[24] Suwarti, Wahyono, B. Prasetiyo. "Analisis Pengaruh Intensitas Matahari, Suhu Permukaan & Sudut Pengarah Terhadap Kinerja Panel Surya," *Jurnal Teknik Energi* Vol 14 No. 3, 2018; 78 – 85, Polines, Semarang, Indonesia, 2018.