p–ISSN: 2723 - 6609 e-ISSN: 2745-5254
Vol. 5, No. 7 July 2024 http://jist.publikasiindonesia.id/

Design of Forecasting Electrical Power of Ultra-Short-Term Solar Power Using the Hybrid Model K-Nearest Neighbors LSTM

Tri Wahyu Yulianto[1*], Unit Three Kartini[2], Bambang Suprianto[3]
Universitas Negeri Surabaya, Indonesia
Email: tri22002@mhs.unesa.ac.id[1*], unitthree@unesa.ac.id[2], bambangsuprianto@unesa.ac.id[3]
*Correspondence

ABSTRACT

Keywords: solar PV, solar radiation, PV K-NN panel temperature, LSTM.

For the application of renewable energy at the airport, the use of solar power requires certainty of the electricity produced. The certainty of electricity generated from solar power can be predicted using machine learning methods. Predictions made on PV electrical power output are based on historical data from direct measurements of solar PV parameters, including solar radiation and PV panel temperature. Various types of machine learning methods for predicting PV output power have been used in previous studies with different evaluation values of prediction results. In this study, the author conducted a hybrid K-NN method with LSTM to predict the PV electrical power of solar PV output with solar radiation parameters and PV panel temperature. After making predictions using this method, excellent RSME results were obtained with a value of 0.015424830635781967. The results of the PV output power value graph in this prediction are also very good, where the predicted value is close to the value of the testing data or actual data.

Introduction

For the application of new and renewable energy at the airport, the use of solar power is felt to be the most appropriate because, in addition to being easy to install, it can be installed on the rooftop of the building or the ground, as well as the ease of availability of solar power plants on the market (PT Angkasa Pura 2, 2022a).

Solar PV is a type of power plant with a working system that converts solar energy into electrical energy with several variables that affect the production of electrical energy or optimal PV power output (Rifa'i, Ananda, & Fadhli, 2018). The first variable that affects PV power output is PV module temperature and ambient temperature, two types of temperature that affect PV power output. The temperature of the PV module has a stronger influence of about 20~30% of the ambient temperature for the output of the PV power produced (Institute of Electrical and Electronics Engineers, n.d.). The second variable that influences PV power output is solar irradiance. In the tests that have been carried out, the distribution of solar radiation (solar irradiance) and the temperature of the PV module influence the power output performance produced by the PV module (Nurlindah, Mustami, & Musdalifah, 2020).

Various studies in the form of testing and prediction experiments to simulate the performance of PV modules based on historical data have been carried out with the help of mathematical calculations, this is because 2 factors that affect the performance of PV modules to produce power output have different conditions at different times and conditions (Liu et al., 2021). One of the measurements to determine/assess the prediction results of a method is by using RSME (Root Square Mean Error). Prediction experiments have been carried out using Times series and ANN modelling techniques with the KNN prediction model producing a lower RSME value if in the process of predicting the "k" value of the K-NN method is getting larger (Sivakumar et al., 2022). Experiments by combining several methods (hybrid) have also been carried out, namely based on correlation data between PV module power output, solar irradiance and PV module temperature using ARIMA, ANFIS, ANN, and SVM modelling techniques in the first stage of modelling, where the results in the first stage are then combined with the prediction method using GA (Genetic Algorithm) which produces better prediction data (Ozbek, Yildirim, & Bilgili, 2022).

In this study, the author uses a combination of methods to predict the power output of PV modules with the K-Nearest Network (KNN) modelling method and the KNN prediction data is used as new data to be modelled and re-predicted using the Long short-term memory (LSTM) method (Wu, Chen, & Abdul Rahman, 2014). To obtain better prediction results, where later the prediction results will be measured/reviewed by the RSME method, the author selects the prediction period on the daily data in the morning and afternoon when the PV module produces power output only (Purwantoro, Kartini, Suprianto, & Agung, 2022).

Solar irradiance and PV module temperature are very dynamic variables and their changes are affected by the surrounding environmental conditions for each period, so the value is not a time series (Asfah & Kartini, 2020). The RSME value does influence determining the accuracy value of a prediction method used to predict the module's PV power output every period. However, the selection of the prediction method and the period to be carried out also have a role and influence on the predictions produced (Agam & Kartini, 2020).

The novelty of previous research/Novelty of the research conducted concerning previous research is to make a very short-term forecast design at solar power plants located at Soekarno-Hatta airport based on meteorological data, namely temperature and solar irradiance using the K-Nearest Neighbors hybrid method and the Long Short Term Memory (KNN-LSTM) method.

Research Methods

The object of this research is the Soekarno-Hatta Airport Solar Power Plant installed in Terminal 2E. Solar Power Plant in Terminal 2E can capture solar energy with a capacity of 636.65kWp and the minimum amount of electrical energy that can be generated is 1887MWh. This solar power plant has 1,168 PV modules and the capacity of each PV module is 545Wp.

Data collection

The data collection time in this study is September 27 and 29, 2023. The data used in this study are the following data:

Time

This data is the time when other data were taken. This data shows the time in hours and minutes. The period in 1 day of research data was taken from 05.30 to 17.30 with a time range of 5 minutes for each data.

PV Output

This data is the value of electrical power generated from solar PV, where this data value is obtained according to the time in the "Time" data. The unit amount of electrical power generated from solar power is Wh (Watt hour).

Solar Irradiance

This data is the solar radiation value contained regarding the PV module of the solar PV, where this data value is obtained according to the time in the "Time" data. The unit size of solar radiation that hits the PV module of the solar PV is W/m^2.

PV Temp

This data is the temperature value on the PV panel of the solar module, where this data value is obtained according to the time in the "Time" data. The unit size of PV panel modules from solar PV is °C.

The data used in this study were collected using direct observation from the web display to monitor and measure each parameter of the solar power plant.

Exploration Data Analysis (EDA)

Normalization is one of the most frequently used data preparation techniques. In machine learning and data mining [6], this process helps to change the numeric column values in the dataset to use a common scale. One of the challenges that exists in databases is the existence of attributes with different units, ranges, and scales.

Applying data mining or machine learning algorithms to data with drastic ranges can provide less accurate results. Data normalization is a basic element of data mining to ensure that records in the dataset remain consistent.

In the normalization process, it is necessary to transform the data or convert the original data into a format that allows efficient data processing. The main goal of data normalization is to eliminate data redundancy (repetition) and standardize information for better data workflows.

Data normalization is used to scale the data of an attribute so that it is within a smaller range, such as -1 to 1 or 0 to 1. It is generally useful for classification algorithms. The min-max normalization method converts a dataset into a scale ranging from 0 (min) to 1 (max). The original data underwent linear modifications in this data normalization procedure. The minimum and maximum values of the data are retrieved, and each value is changed using the below formula:

Xnorm = (x^'-min(x))/(max(x)-min(x)) (Newman(x) – Newman(x)) + Newman(x)

x = date attribute

min(x) = Minimum absolute value of x

max(x) = Maximum absolute value of x

x' = the old value of each entry in the data

Newsmax(x) = Max value of x

Newman(x) = min value of x

Prediksi K-Nearest Neighbor

The KNN algorithm is a classification technique that determines the class of new data by taking several K of the nearest data as the basis for comparison. This algorithm works by comparing similarities or similarities between data (Ismail, 2018). The working principle of K-Nearest Neighbor (KNN) is to find the closest distance between the data to be evaluated and the nearest K-Nearest Neighbor in ththe e training data. The following is the formula for finding distance using the Euclidian fms (Agusta, 200).

$$hal\sqrt{\sum_{i=1}^{p}(x_{2i}-x_{1i})^2}$$

Information:

p = Data dimension

i = Variable Data

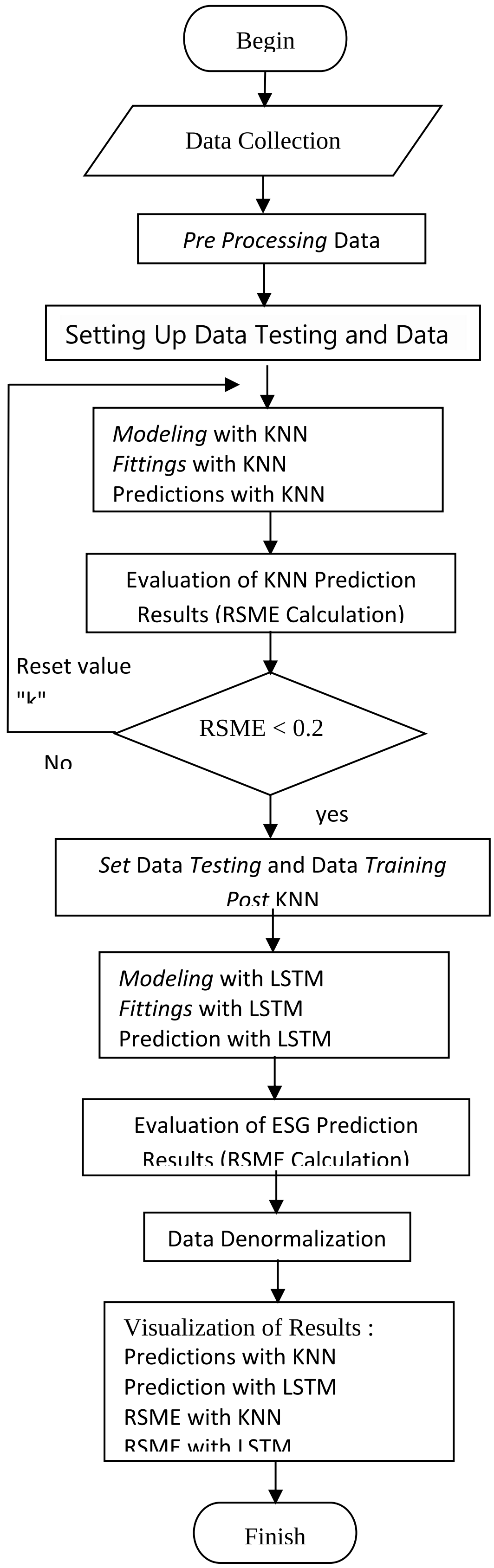
x1 = Sample data

x2 = Test data or testing data

d = Distance

K-Nn Lstm Hybrid Method Prediction

The hybrid methods of KNN and LSTM combine the power of both algorithms to improve prediction performance in some cases. KNN (K-Nearest Neighbors) is an instance-based learning algorithm used primarily for classification and regression problems, while LSTM (Long Short-Term Memory) is a type of recursive neural network (RNN) architecture that is effective in modelling sequences and patterns in Time series data. The process of the research implementation flow is illustrated in the flowchart in Figure 1.


Begin
Data Collection
Pre Processing Data
Setting Up Data Testing and Data
Modeling with KNN
Fittings with KNN
Predictions with KNN
Evaluation of KNN Prediction Results (RSME Calculation)
Reset value "k"
RSME < 0.2
No
yes
Set Data Testing and Data Training Post KNN
Modeling with LSTM
Fittings with LSTM
Prediction with LSTM
Evaluation of ESG Prediction Results (RSME Calculation)
Data Denormalization
Visualization of Results :
Predictions with KNN
Prediction with LSTM
RSME with KNN
RSME with LSTM
Finish

Results and Discussion

In the flow process of research, the first step is to collect research data. The data used in this study were collected using direct observation from the web display to monitor and measure each parameter in the solar power plant. The following are the research data that have been collected and preprocessed data.

```
                    Time  PV(Wh)  IRRADIANCE (W/m²)  PV Temp (°C)
0    27/09/2023 05:30:00   0.000                0.7          23.2
1    27/09/2023 05:35:00   0.307                1.8          23.2
2    27/09/2023 05:40:00   1.309                3.5          23.2
3    27/09/2023 05:45:00   2.455                6.0          23.3
4    27/09/2023 05:50:00   4.583                9.7          23.4
..                   ...     ...                ...           ...
140  27/09/2023 17:10:00  24.280               58.4          33.1
141  27/09/2023 17:15:00  19.909               47.4          32.6
142  27/09/2023 17:20:00  15.702               34.9          32.1
143  27/09/2023 17:25:00  12.117               27.0          31.6
144  27/09/2023 17:30:00   5.978               20.4          31.3

[145 rows x 4 columns]
```

Figure 2 Solar PV data on September 27, 2023,
Source: Personal, 2023

```
Data columns (total 4 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Time               145 non-null    object
 1   PV(Wh)             145 non-null    float64
 2   IRRADIANCE (W/m²)  145 non-null    float64
 3   PV Temp (°C)       145 non-null    float64
dtypes: float64(3), object(1)
memory usage: 4.7+ KB
```

Figure 3
Solar PV Data Type Info on September 27, 2023,
Source: Personal, 2024

	PV(Wh)	IRRADIANCE (W/m²)	PV Temp (°C)
count	145.000000	145.000000	145.000000
mean	243.959207	495.113793	46.176552
std	142.054050	304.499873	11.059605
min	0.000000	0.700000	23.200000
25%	113.270000	214.700000	37.800000
50%	275.267000	538.000000	49.900000
75%	385.432000	812.300000	55.200000
max	416.529000	891.300000	60.800000

Figure 4
Description of Solar PV Data on September 27, 2023,
Source: Personal, 2024

```
Data setelah normalisasi kolom ke- ['PV(Wh)', 'IRRADIANCE (W/m²)', 'PV Temp (°C)']
                    Time    PV(Wh)  IRRADIANCE (W/m²)  PV Temp (°C)
0    27/09/2023 05:30:00  0.000000           0.000000      0.000000
1    27/09/2023 05:35:00  0.000737           0.001235      0.000000
2    27/09/2023 05:40:00  0.003143           0.003144      0.000000
3    27/09/2023 05:45:00  0.005894           0.005951      0.002660
4    27/09/2023 05:50:00  0.011003           0.010106      0.005319
..                   ...       ...                ...           ...
140  27/09/2023 17:10:00  0.058291           0.064788      0.263298
141  27/09/2023 17:15:00  0.047797           0.052437      0.250000
142  27/09/2023 17:20:00  0.037697           0.038401      0.236702
143  27/09/2023 17:25:00  0.029090           0.029531      0.223404
144  27/09/2023 17:30:00  0.014352           0.022120      0.215426

[145 rows x 4 columns]
```

Figure 5 Preprocessing of Solar PV Data on September 27, 2023,
Source: Personal, 2024

In the data preprocessing process, it is known that the number of data is 145 data with the number of data columns as many as 4. After the data preprocessing process, a training data set with a percentage of 0.7 of the total data and a testing data set with a percentage of 0.3 of the total data were carried out. The results of the training and testing data sets are shown in Figure 6.

```
Data Train:
     IRRADIANCE (W/m²)  PV Temp (°C)
0             0.000000      0.000000
1             0.001235      0.000000
2             0.003144      0.000000
3             0.005951      0.002660
4             0.010106      0.005319
..                 ...           ...
96            0.913991      0.853723
97            0.911296      0.840426
98            0.905120      0.816489
99            0.861329      0.829787
100           0.881653      0.829787

[101 rows x 2 columns]

Data Test:
     IRRADIANCE (W/m²)  PV Temp (°C)
101           0.859421      0.808511
102           0.839996      0.827128
103           0.818774      0.824468
104           0.790254      0.779255
105           0.793847      0.752660
106           0.773187      0.776596
107           0.763530      0.773936
108           0.718280      0.771277
109           0.695262      0.718085
110           0.675051      0.718085
111           0.674489      0.704787
112           0.654839      0.696809
113           0.629688      0.678191
114           0.614866      0.667553
115           0.590950      0.648936
116           0.577139      0.632979
```

Figure 6.
Testing Data Set and Training Data on Solar Power Plant Data on September 27, 2023,
Source: Personal, 2024

The next process is modelling, fitting and prediction using K-NN. This process is carried out on several "k" values in K-NN with values from k=2 to k=10. This is done to obtain the prediction value with the best evaluation results. After the prediction results are obtained and the prediction results are evaluated, the prediction results for each "k" are obtained as follows in the bar diagram Figure 7.

Figure 7. RSME Prediction Score Using K-NN,
Source: Personal, 2024

According to Figure 10, the K-NN prediction with the best RSME value is in the K-NN prediction with a value of "k" is 10 with a value of 0.11344051656173618. Therefore, the results of the K-NN prediction with a value of k=10 are then used as initial data to be predicted again using the LSTM method. The prediction results with LSTM obtained better RSME results of 0.015424830635781967 or closer to "0" when compared to the predicted RSME value which only used the K-NN k=10 method.

Meanwhile, the results of the prediction of the output PV power value based on solar radiation parameters (Solar Irradiance) and PV panel temperature (PV Temp.) using the two predictions above can be seen in Figure 8 and Figure 9.

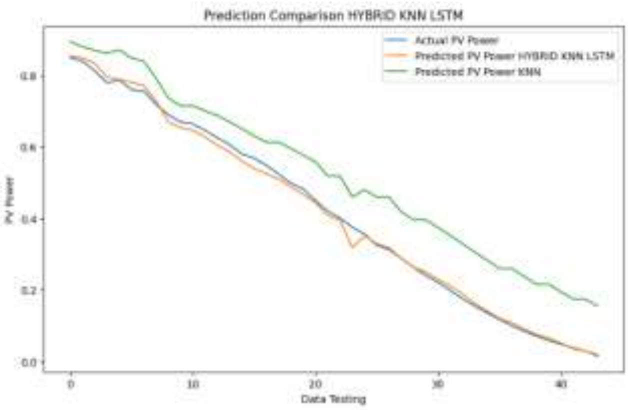


Figure 8
Graph comparison of the values of each PV output power prediction method on the testing data,
Source: Personal, 2024

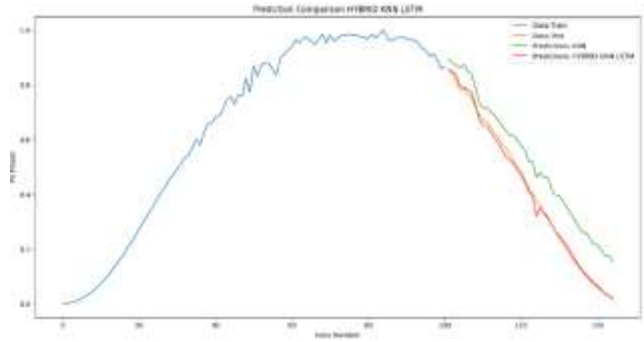


Figure 9. Graph of PV Power Output Power Value Training Data and Values of Each PV Output Power Prediction Method on Data Testing,

Source: Personal, 2024

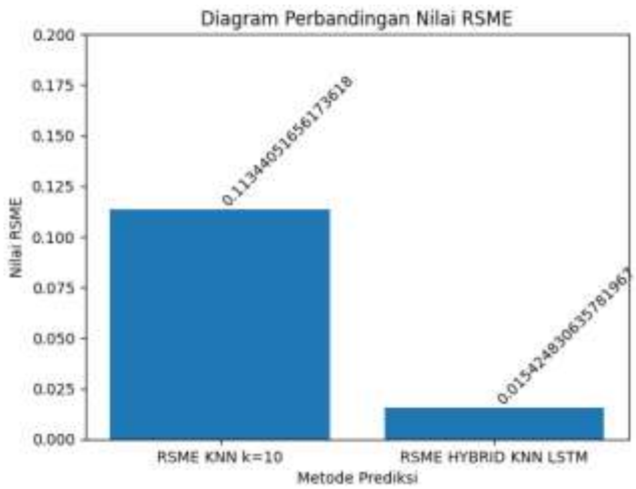


Figure 10

RSME Prediction Score Using K-NN k=10 and K-NN Hybrid Prediction k=10 LSTM,

Source: Personal, 2024

Conclusion

Modelling and prediction of PV power output of solar PV at Soekarno-Hatta Airport Terminal 2E have been carried out using solar radiation parameter data and PV panel module temperature. The prediction method used is the K-NN method with a "k" value of 10 and the RSME value of this prediction method is 0.11344051656173618. The results of this prediction are used as preliminary data to re-predict the PV Output power using the LSTM method because although the previous method has a good RSME value, the PV Output power value has not approached the testing data as shown in Figure 11 and Figure 12. After making predictions using the LSTM method, better RSME results were obtained with a value of 0.015424830635781967. The results of the PV Output power value graph display in this prediction are also better, where the predicted value is close to the value of the testing data or actual data. So the first conclusion can be obtained that the prediction of PV electrical power Output with solar radiation parameters and PV panel temperature using the KNN prediction method with a value of k=10 obtained good prediction results. The following conclusion is that better prediction results can be obtained when the prediction in the initial method is carried out hybrid method using the LSTM method.

Bibliography

Agam, Masviki, & Kartini, Unit Three. (2020). Peramalan Daya Listrik PLTS On Grid pada Rumah Tinggal Menggunakan Metode K-Nearest Neighbor Decomposition Feed Forward Neural Network Berdasarkan Data Meteorologi. *Jurnal Teknik Elektro*, *9*(2).

Agusta, Yudi. (2007). K-means–penerapan, permasalahan dan metode terkait. *Jurnal Sistem Dan Informatika*, *3*(1), 47–60.

Asfah, Rani Fajriyah Islamiyati, & Kartini, Unit Three. (2020). Peramalan Radiasi Global Matahari Jangka Pendek Menggunakan Modeltriple Exponential Smoothing-Feed Forward Neural Network. *Jurnal Teknik Elektro*, *9*(3), 677–684.

Ismail, Asep Maulana. (2018). Cara Kerja Algoritma k-Nearest Neighbor (k-NN). *Dipetik*, *9*(17), 2019.

Liu, Zhi Feng, Luo, Shi Fan, Tseng, Ming Lang, Liu, Han Min, Li, Lingling, & Mashud, Abu Hashan Md. (2021). Short-term photovoltaic power prediction on modal reconstruction: A novel hybrid model approach. *Sustainable Energy Technologies and Assessments*, *45*, 101048.

Nurlindah, Nurlindah, Mustami, Muh Khalifah, & Musdalifah, Musdalifah. (2020). Manajemen Pendidik Dan Tenaga Kependidikan Dalam Meningkatkan Mutu Pendidikan. *Idaarah*, *4*(1), 40–51. https://doi.org/10.24252/idaarah.v4i1.13893

Ozbek, Arif, Yildirim, Alper, & Bilgili, Mehmet. (2022). Deep learning approach for one-hour ahead forecasting of energy production in a solar-PV plant. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, *44*(4), 10465–10480.

Purwantoro, Kukuh Eko, Kartini, Unit Three, Suprianto, Bambang Suprianto, & Agung, Achmad Imam. (2022). Prediksi Daya Listrik Jangka Sangat Pendek Pembangkit Photovoltaic Berbasis Internet of Things Menggunakan Feed Forward Neural Network. *JURNAL TEKNIK ELEKTRO*, *11*(3), 386–396.

Rifa'i, Muhammad, Ananda, Rusydi, & Fadhli, Muhammad. (2018). *Manajemen peserta didik (Pengelolaan peserta didik untuk efektivitas pembelajaran)*.

Sivakumar, S., Neeraja, B., Jamuna Rani, M., Anandaram, Harishchander, Ramya, S., Padhan, Girish, & Gurusamy, Saravanakumar. (2022). Machine Learning Approach on Time Series for PV-Solar Energy. *Advances in Materials Science and Engineering*, *2022*(1), 6458377. https://doi.org/10.1155/2022/6458377

Wu, Yuan Kang, Chen, Chao Rong, & Abdul Rahman, Hasimah. (2014). A novel hybrid model for short-term forecasting in PV power generation. *International Journal of Photoenergy*, *2014*(1), 569249.