

Article

An Adaptive Decision Tree Regression Modeling for the Output Power of Large-Scale Solar (LSS) Farm Forecasting

Nabilah Mat Kassim ¹, Sathiswary Santhiran ¹, Ammar Ahmed Alkahtani ¹, Mohammad Aminul Islam ^{2,*},
Sieh Kiong Tiong ¹, Mohd Yusrizal Mohd Yusof ³ and Nowshad Amin ^{1,*}

¹ Institute of Sustainable Energy (ISE), Universiti Tenaga Nasional, Kajang 43000, Selangor, Malaysia

² Department of Electrical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Kuala Lumpur, Malaysia

³ TNB Renewables Sdn. Bhd., No. 16A, Persiaran Barat, Petaling Jaya 46050, Selangor, Malaysia

* Correspondence: aminul.islam@um.edu.my (M.A.I.); nowshad@uniten.edu.my (N.A.)

Abstract: The installation of large-scale solar (LSS) photovoltaic (PV) power plants continues to rise globally as well as in Malaysia. The data provided by LSS PV consist of five weather stations with seven parameters, a 22-unit inverter, and 1-unit PQM Meter Grid as a big dataset. These big data are rapidly changing every minute, they lack data quality when missing data, and need to be analyzed for a longer duration to leverage their benefits to prevent misleading information. This paper proposed the forecasting power LSS PV using decision tree regression from three types of input data. Case 1 used all 35 parameters from five weather stations. For Case 2, only seven parameters were used by calculating the mean of five weather stations. While Case 3 was chosen from an index correlation of more than 0.8. The analysis of the historical data was carried out from June 2019 until December 2020. Moreover, the mean absolute error (MAE) was also calculated. A reliability test using the Pearson correlation coefficient (r) and coefficient of determination (R^2) was done upon comparing with actual historical data. As a result, Case 2 was proposed to be the best input dataset for the forecasting algorithm.

Keywords: large-scale solar PV; decision tree regression; forecast; PV plant output; global irradiance; energy



Citation: Kassim, N.M.; Santhiran, S.; Alkahtani, A.A.; Islam, M.A.; Tiong, S.K.; Mohd Yusof, M.Y.; Amin, N. An Adaptive Decision Tree Regression Modeling for the Output Power of Large-Scale Solar (LSS) Farm Forecasting. *Sustainability* **2023**, *15*, 13521. <https://doi.org/10.3390/su151813521>

Academic Editor: Mohamed A. Mohamed

Received: 26 December 2022

Revised: 14 February 2023

Accepted: 6 March 2023

Published: 10 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Malaysia has a potential for solar power plant development due to the high average daily solar radiation, with a range of 4–8 h daily [1]. Based on this, Malaysia's government has set a target of 20% renewable energy by 2025 [2]. PV systems range from small systems to large-scale PV plants, such as Pico PV systems, off-grid domestic systems, off-grid non-domestic systems, hybrid systems, grid-connected distributed systems, and grid-connected centralized systems [3]. The most popular types of hybrid power generation are grid-connected photovoltaic (GCPV) systems. The LSS PV system needs to be installed near the grid electricity network offered by the local power provider to be linked to as GCPV systems. Thus, if the GCPV is frequently unstable, the main advantage of having a GCPV system becomes unachievable [4].

Solar radiation, cloud covering, temperature, humidity, atmospheric pressure, wind speed, and other elements all have an impact on solar power generation [5]. Because of the chaotic nature of the Earth's weather system, these environmental parameters can change substantially at any time, making dependable and precise forecasting of solar power difficult [6]. So far, various studies on solar power prediction have been completed, which may be separated into physical models, statistical approaches, regression methods, and their hybrid methods [7]. Among these, the artificial intelligence (AI) algorithm serves as the foundation for the present solar power forecast systems. AI is a field of computer science and developing technology that uses computer simulation to study human logical

thinking, reasoning, and group behaviors [8]. Machine learning (ML), expert systems, fuzzy logic, and heuristic optimization are some of the most extensively used AI methods [9].

Accurate LSS PV power forecasting is important when connected to the grid for an effective plant [10]. Hussein et al. [10] publish a time series forecasting of solar power generation for large-scale photovoltaic plants to predict PV power output by using statistical methods based on artificial intelligence. They mentioned the time series forecasting for PV power plants is only reliable for one hour ahead prediction. Planning and procedures must adjust according to these changes in the forecast LSS PV condition [11]. Several writers have explored the forecasting of PV output power based on the classifications of meteorological factors [12–16]. By using a support vector machine for weather classification, Shi et al. [12] established a unique prediction model to estimate the power output of a PV plant with a 30 MW capacity for a 24 h time horizon. Mellit et al. [14] used two artificial neural networks to predict the power produced by a 50 MW PV plant using more than one year of data. Dahlan et al. [17] forecasting power generation model of an LSS PV farm using Artificial Neural Network (ANN) and Particle Swarm Optimization (PSO) technique of 50 MW as a case study. Nguyen et al. [18] propose a new model for short-term forecasting power generation capacity of large-scale solar power plant (SPP) in Vietnam considering the fluctuations of weather factors when applying the Long Short-Term Memory networks (LSTM) algorithm. They add new features to improve the selected LSTM model precision and cope with the problem of error due to weather forecast data. The meteorological data and historical power output can map the relationship using a correlation coefficient to predict the output energy in the future [19].

According to [20], the highly correlated variables with PV power are solar irradiance, air temperature, and dew point, while relative humidity and cloud type are somewhat negatively correlated with it. However, the fast fluctuation of PV power with a high time resolution makes it difficult to forecast the output energy. Thus, this situation is necessary to explore the reasons behind the power changes [21].

This paper aims to present a prediction of total kWh energy output from a GCPV system using an Adaptive Decision Tree Regression modeling. The location of the operating system is the grid-connected PV systems at Pahang, Malaysia. Among the climatic parameters being monitored were the Total Global Horizontal Irradiance (W/m^2), Total Horizontal Irradiance (Wh/m^2), Global Irradiance on the Module Plane (W/m^2), Total Slope Irradiance (Wh/m^2), Ambient Temperature ($^{\circ}\text{C}$), PV Module Temperature ($^{\circ}\text{C}$), and Wind Speed (m/s). These datasets were trained by three cases. Case 1 uses all data from five weather stations. Case 2 uses mean data of five weather stations, while Case 3 uses an index correlation of more than 0.8. All these three datasets will be trained using the TDR model to predict output power (KW) of LLS PV. The best case is identified by evaluating the performance metric of the model.

This paper presents the comparison of various numbers of input weather stations to predict total kWh energy LSS PV. Section 2 illustrates the LSS PV configuration, flowchart, and Decision Tree Regression model. The performance of forecasting will also be evaluated according to a mean absolute error (MAE) and the coefficient of the determinant (R^2). Section 3 comprises a comparison and discussion of the three cases tested utilizing the Decision Tree Regression model to identify the best input variable for forecasting total kWh LSS PV. Section 4 summarizes the paper's conclusion.

2. Methodology

2.1. LSS Information

The UiTM 50 MW LSS PV in Pahang has been used as a case study. This location was selected because the capacity is more than 30 MWac and it is called LSS PV. Figure 1a presents the LSS PV power plant configuration from the PV panel to TNB-Grid Meter. With a 50 MWac capacity and 290 acres of land, the LSS PV can supply power to up to 22,000 dwellings [12,13]. The area of the plant is 290 acres, with a total of 184,850 Polycrystalline PV modules installed. The plant is facing South, the tilt angle is 15° , and the

shading angle due to the presence of parallel rows of PV modules is 20° . In addition, this solar power plant has five weather stations.

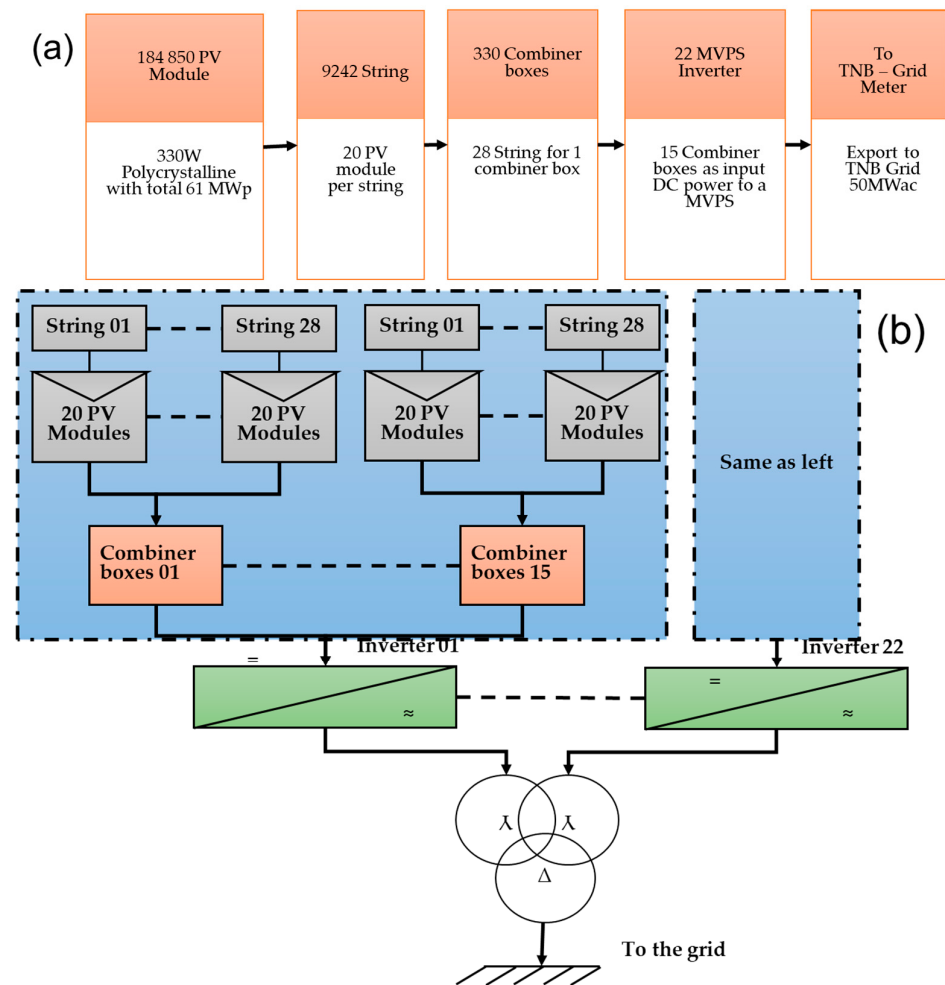


Figure 1. (a) Large-scale Solar photovoltaic configuration, and (b) schematic block diagram of the considered PV plant (50 MWac).

This location has a tropical rainforest climate, and sunshine will start around 7:15 A.M. until sunset around 7:20 P.M. The ambient temperature ranges from 27 to 35 °C, and the air temperature is 26.4 °C throughout the year. The high amount of cloud cover dominates the atmosphere, with the occurrence of clear skies rarely going beyond the next day. According to the Global Solar Atlas, the direct normal irradiation per year is 998.2 kWh/m², Global horizontal irradiation per year is 1715.9 kWh/m² and diffuse horizontal irradiation per year is 955.5 kWh/m² [22].

Figure 1b shows the block diagram of the considered PV plant with a capacity 50 MWac. The strings are made of 20 series-connected PV modules, while groups of 28 strings are parallel-connected into a DC combiner box where fuses prevent overcurrents in the strings. Fifteen units of DC combiner box are then connected to a DC side inverter that converts the direct current produced by the PV strings into an alternating current compatible and synchronized with the grid. Finally, the AC sides of the 22-unit inverters are connected to a double primary transformer that converts the low voltage output of the inverters (33 kV) up to 132 kV, corresponding to the nominal voltage of the electricity grid.

2.2. Historical Dataset

The dataset in this paper includes Total Global Horizontal Irradiance (W/m²), Total Horizontal Irradiance (Wh/m²), Global Irradiance on the Module Plane (W/m²), Total

Slope Irradiance (Wh/m^2), Ambient Temperature ($^{\circ}\text{C}$), PV Module Temperature ($^{\circ}\text{C}$), and Wind Speed (m/s) by station EM04, EM11, EM12, EM17, and EM21 then an output Power (kW) at the TNB-Grid Meter.

The dataset was created as a CSV file. A perfect date series with the following format was created by combining the Year, Month, Day, Hour, and Minute columns: YYYY-MM-DD HH: MM. The dataset chosen for training was June–December 2019 and the predicted data was for January–December 2020. In the LSS PV plant, a full set of pyranometers was installed at least one (1) for every 10 MWac of plant size at approximate locations within the site [23,24]. In addition, according to Malaysia Standard, at least one (1) set of full weather stations shall be installed for every 10 MWac of plant size [25].

Figure 2 shows a flowchart of the best data selection from weather stations using the DTR model. Eighteen months of historical data were used in this study. The first part involved data preprocessing corresponding to extraction and preprocessing operation. Then, identifying the correlation between input and output variables data, and then checking any hypothesis. Next, calculating the mean data from the 5-weather stations for mean data. Three cases were developed to test the suitable input data for the machine learning algorithm. Case 1 used all data provided by LSS PV, while Case 2 used mean data from five weather stations, and lastly, Case 3 the weather data selection according to a correlation index of more than 0.6. Therefore, the weather data was used are total global horizontal irradiance(tghi), global irradiance on the module plane (gimp) and PV module temperature (pvmt). All three cases implemented the DTR model to predict the power output LLS PV. Lastly, we evaluated using evaluation metrics and presented the best input data suitable for implementation with other machine learning algorithms.

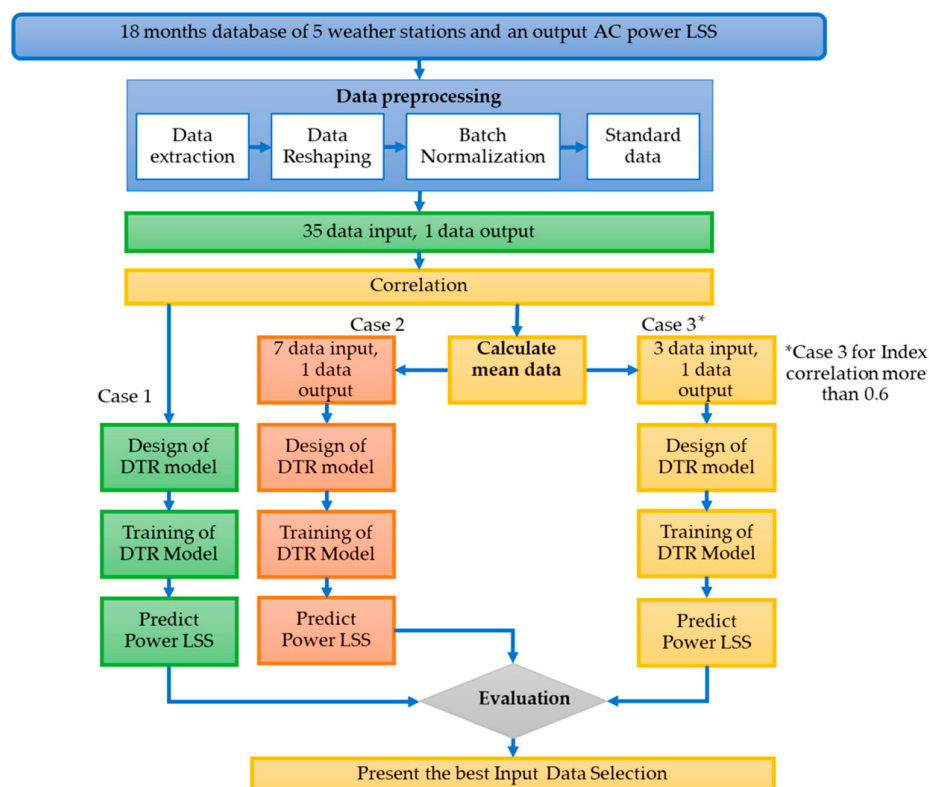


Figure 2. Flowchart of the best data selection of weather station using the DTR model.

2.3. Training the Model Decision Tree Regression

A non-parametric supervised learning model called Decision Trees (DT) is employed for both classification and regression analysis. This tree-structured classifier contains three different types of nodes. The original node that represents the complete sample, known as the root node, may be divided further into other nodes. Interior nodes represent the

properties of a dataset, while branching represents its decision-making procedures. The leaf nodes display the outcome in the end. This approach is quite beneficial for dealing with problems concerning decisions. The model was trained using the DTR function from the sklearn library [26–28]. Figure 3 shows the decision tree structure.

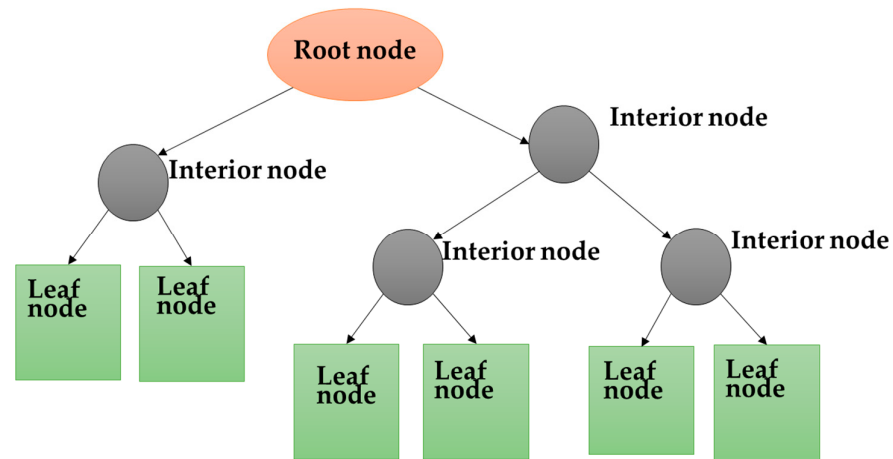


Figure 3. Decision Tree Nodes, Root, Interior, and Leaf.

Purity is a statistic used to confirm the choice to divide at each node. A node is 100% impure when it is split evenly in half, and the tree does not have a clear decision for the output. When every piece of input information is paired with a single decision value, the node is 100% pure. The accuracy will increase when the purity is highest. When the splitting is finished, the tree is trimmed to make it less complex by using the average as the final value for the target. In order to forecast the output, pruning looks for and eliminates any tree branches that are redundant or non-critical.

$$\Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (1)$$

where s corresponds to a potential split at any node t , which is divided by s into a left (t_L) and right (t_R) child nodes in proportion to P_L and P_R , respectively. Here, $i(t)$ is a pre-defined impurity standard for splitting, and $\Delta i(s, t)$ is the final measure of decreasing impurity from split s .

The three impurity measurements that are most often employed are gain-ratio, Gini index, and Chi-square. In this study, the algorithmic evolution of decision tree regression inside the *Scikitlearn* package uses the Gini index, which might favor bigger partitions and moderate computing. The Gini impurity index (I_G) varies between 0 and 1, and it can be formulated as

$$I_G(t_{X(x_i)}) = 1 - \sum_{k=1}^m f(t_{X(x_i)}, k)^2 \quad (2)$$

where $f(t_{X(x_i)}, k)$ represents x_i probabilities of each sample in the leave k at node t . The splitting criterion on DT is derived from the determination of the feature with the lowest I_G .

2.4. Accuracy Assessment Methods

The mean absolute error (MAE), which calculates the mean absolute distance for a model's fitness in percentage terms, was used to validate the model. Using values between 0 and 1, the coefficient of the determinant (R^2) measures the difference between the true and predicted values [29]. Values that are closer to 1 show a collaborative link, whereas values that are closer to 0 show a weaker correlation [30]. These markers are often employed to

verify machine-learning algorithms. The formulas for determining the R^2 and MAE are shown in Equations (1) and (2), respectively.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - X_i| \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^n (X_i - X_m)(Y_i - Y_m)}{\sqrt{\left(\sum_{i=1}^n (X_i - X_m)^2\right) \left(\sum_{i=1}^n (Y_i - Y_m)^2\right)}} \quad (4)$$

where n is the total number of data points or instances, X_i and Y_i are the actual and predicted values, respectively, X_m and Y_m are the mean of the actual and predicted values, respectively.

3. Results and Discussion

Figure 4a–h depicts the historical data of weather station EM04 and output power from LSS PV. The dataset is presented in six-month duration for training DTR Model. From the graph, we can see the appearance of the minimum and maximum values of the dataset. It can be seen that the range of total global irradiance (tghi) is between 6 and 1408 W/m^2 , total horizontal irradiance (thi) is between 6 and 6892 W/m^2 , global irradiance module plane (gimp) from 9 to 1416 W/m^2 , total slope irradiance (tsi) is in the range 3 to 6923 W/m^2 , while ambient temperature (at) is between 22 and 36 $^{\circ}C$, PV module temperature (pvmt) is in the range 21 to 64 $^{\circ}C$, and wind speed (ws) in the range 0.04 to 6 m/s, as shown in Table 1.

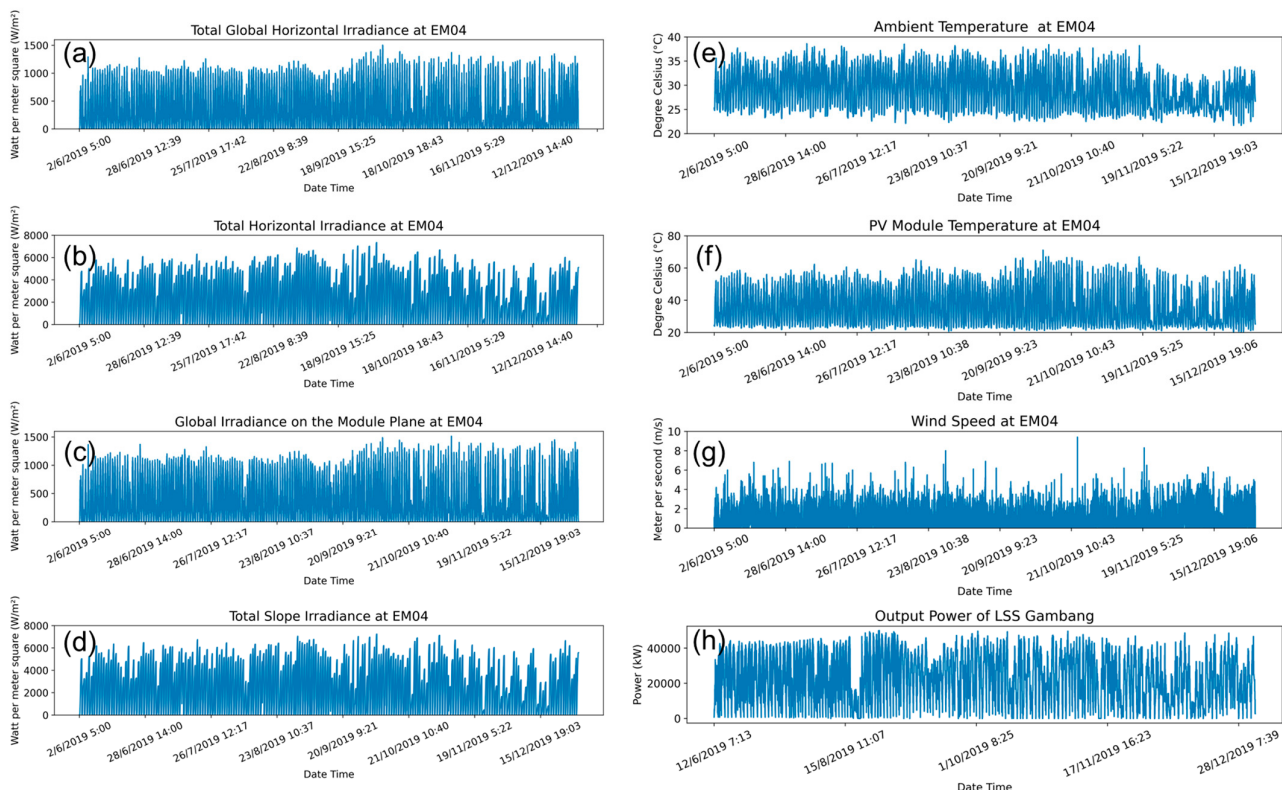


Figure 4. The dataset of (a) total global irradiance (tghi), (b) total horizontal irradiance (thi), (c) global irradiance module plane (gimp), (d) total slope irradiance (tsi), (e) ambient temperature (at), (f) module temperature (pvmt), (g) wind speed (ws), and (h) Output Power (kW) for weather station EM04.

Table 1. The Statistical Data Analysis of June–December 2019 (training) and January–December 2020 (test).

	Data	tghiMean	thiMean	gimpMean	tsiMean	atMean	pvmtMean	wsMean
count	Train	82,528	82,528	82,528	82,528	82,528	82,528	82,528
	Test	134,940	134,940	134,940	134,940	134,940	134,940	134,940
mean	Train	462.97	2778.03	478.08	2874.77	30.14	40.18	1.23
	Test	487.53	2877.03	492.91	2918.61	30.44	39.87	1.50
std	Train	297.25	1782.31	305.52	1835.16	2.50	8.67	0.56
	Test	297.98	1805.50	299.41	1823.98	2.39	7.71	0.58
min	Train	6.40	6.11	9.40	3.22	22.58	21.92	0.04
	Test	5.00	4.77	6.60	2.55	22.84	21.78	0.14
max	Train	1408.40	6892.11	1416.40	6923.39	36.24	63.92	5.38
	Test	1416.80	7329.28	1441.40	7249.056	37.28	62.70	5.20

Table 1 shows the statistical data analysis of the datasets for June–December 2019 (Train) and January–December 2020 (Test). Train data have 82,528 for six months, while test data have 134,940 for year duration. This dataset has a good standard deviation value because it is a positive value. From the statistic maximum test data, we can see the maximum yearly irradiance of the LSS PV. The maximum value recorded for total global irradiance (tghi) is 1416.80 W/m², for maximum total horizontal irradiance (thi), it is 7329.28 W/m², while for maximum global irradiance module plane (gimp), it was recorded at 1441.40 W/m², and the maximum total slope irradiance (tsi) is 7249.056 W/m². However, as high temperature is not for a good PV module, it reduces the efficiency of the PV module. The high temperature recorded at LLS PV is 37.28 °C.

The bar graph in Figure 5 shows a correlation between the historical 35 weather datasets and the power (kW) for the June–December 2019 dataset. This paper uses Pearson's correlation coefficient (r) to test the statistical relationship or correlation between the weather dataset and output power. The graph shows that data for the total global irradiance (tghi), the global irradiance module plane (gimp), and the PV module temperature (pvmt) indicate a very strong positive relationship (0.80 to 1.00) for all weather stations. However, the ambient temperature (at) and the wind speed (ws) indicate a variety of relationships as moderate positive (0.40 to 0.59) and weak positive (0.20 to 0.39). However, total horizontal irradiance (thi) and total slope irradiance (tsi), with values ranging from 0.00 to 0.19, show only a very weakly positive connection with power (kW). It is found that the output power of the LSS PV is greatly affected by factor of total global irradiance (tghi), the global irradiance module plane (gimp), and the PV module temperature (pvmt). Figure 6a–h shows a day plot graph from 5:00 A.M. to 8:00 P.M. of the weather station in LSS PV. Data preprocessing was applied to the dataset weather stations to improve the efficiency of model building. In this step, abnormal data were removed to avoid interference during training and model testing. Then, missing data were replaced or corrected by using different methods so that the resulting dataset is of the size needed to serve the training. However, this may lead to inaccurate data because the missing data range is large and the characteristic of data in this range is not close to linearity.

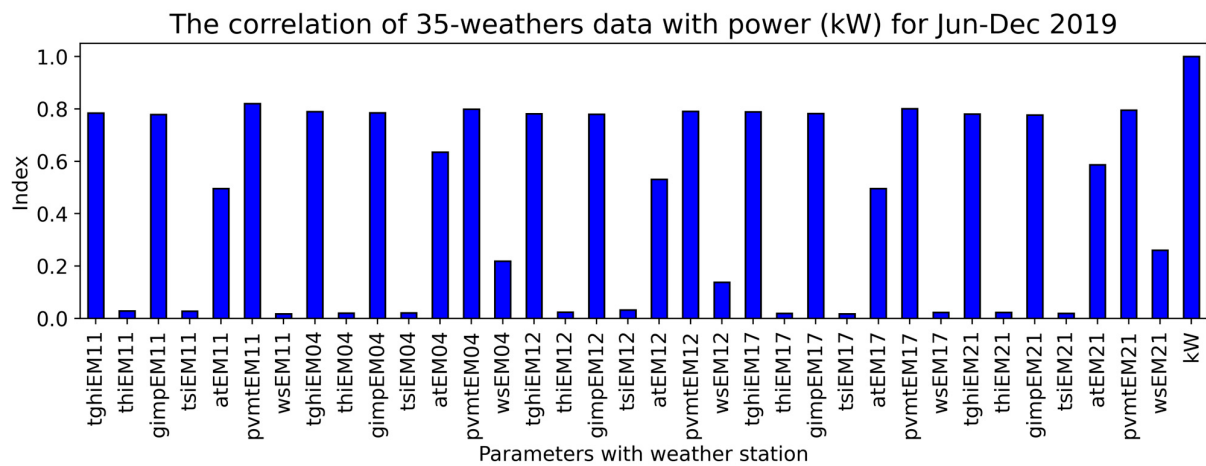


Figure 5. Correlation bar graph for June–December 2019.

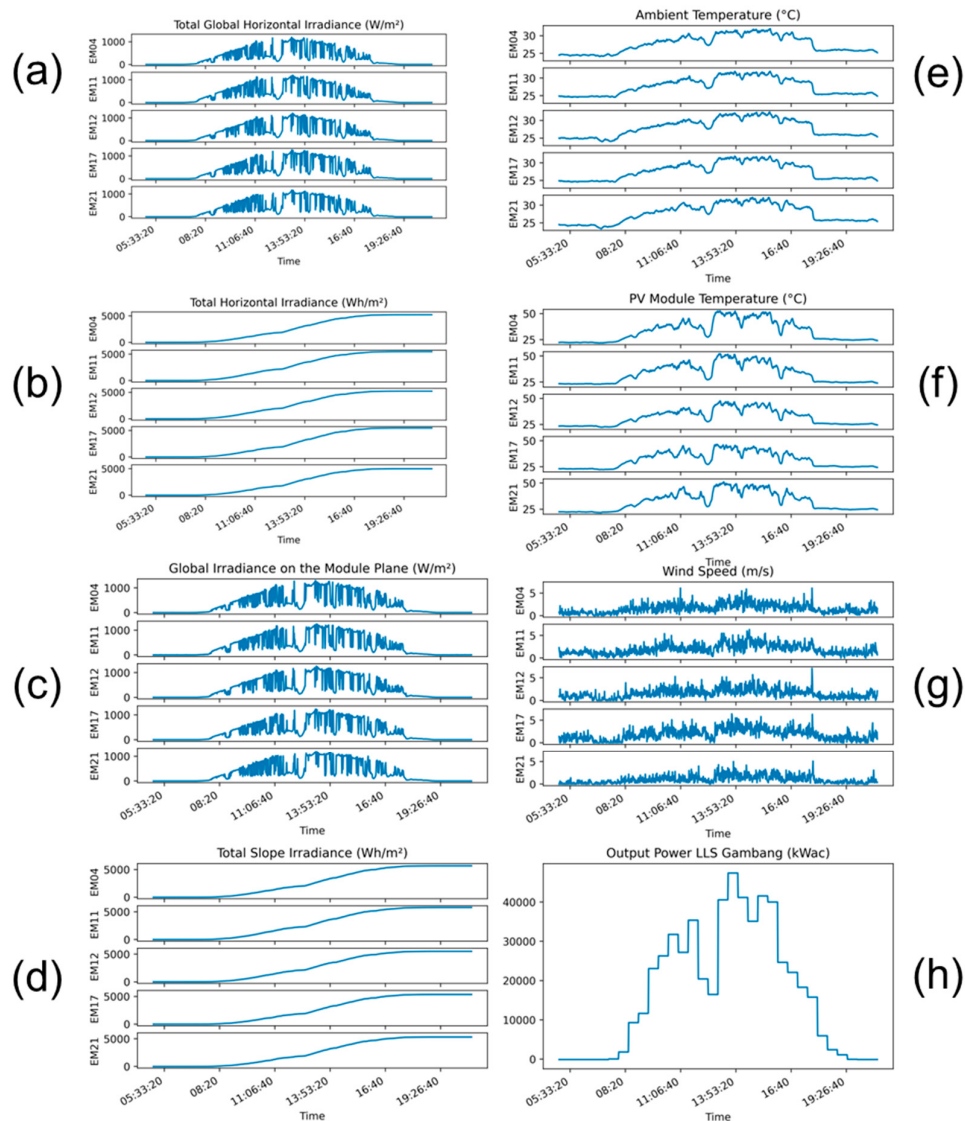


Figure 6. A day plot graph data with (a) total global irradiance, (b) total horizontal irradiance, (c) global irradiance module plane, (d) total slope irradiance, (e) ambient temperature, (f) PV module temperature, (g) wind speed, and (h) output power for station EM04 in LSS.

After data preprocessing, the dataset was split into two parts which are data training and data test. The data training is used to train the model, while the data test is considered as unknown observations to measure the accuracy of the model. The length of the data train is 6 months, while the length of data test is 365 days to reflect full year context. The input data are Meteorological data (tghi, thi, gimp, tsi, at, pvmt, and ws), while the output data are the output power (kW).

Next, a Decision Tree Regression was used to predict the LSS PV output. This dataset contains 82,528 data train, and 134,940 data test was recorded with the features of total global irradiance (tghi), total horizontal irradiance (thi), global irradiance module plane (gimp), total slope irradiance (tsi), ambient temperature (at), PV module temperature (pvmt), wind speed (ws), and Power (kW). After training the models on the training dataset, the test dataset, which included the real historical weather data from January to December 2020 collected from the LSS PV is used to check the quality of the models. For the machine learning model, we use a python library that is *DecisionTreeRegressor* of the *sklearn* package. Here, we train our Decision Tree model as a regressor using the library in python and then fitting the model with training data of features and labels. After that, we can predict test values using test data of features, as shown in Figure 7. We have three case studies to predict output power at LLS PV. The list of case studies it as follows:

Case 1: 35 input parameters of 5 weather station datasets with 1 output power data

Case 2: 7 input parameters of mean weather station dataset with 1 output power data

Case 3: 3 input parameters with a correlation index of more than 0.8 and with 1 output power data

```
In [75]: from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_absolute_error
# DECISION TREE REGRESSION MODEL
model = DecisionTreeRegressor(random_state = 0)
model.fit(input_df2019, target_df2019)

Out[75]: DecisionTreeRegressor(random_state=0)

In [76]: predicted_data2019 = model.predict(input_df2019)
print("Predicted 2019 Power (All Data):\n",predicted_data2019)

Predicted 2019 Power (All Data):
[ 769.71    856.54    950.16    ... 2740.978516 2740.978516
2740.978516]
```

Figure 7. Using the sklearn library DTR Model to train dataset.

To verify the best input data suitable for machine learning to reduce complexity and memory space, the performance comparison was tested using an accuracy score of MAE, r , and R^2 for Case 1, Case 2, and Case 3 as shown in Table 2. Although the DTR model performs very well with Case 1 and Case 2, when using Case 3, which is a dataset with a correlation index of more than 0.8, the error increases significantly, with MAE being 1.4932 percent. It is obvious that there is a significant increase in error when using only three parameters compared with seven parameters of weather station data due to insufficient datasets for training DTR model. The results show that the MAE of Case 1 and Case 2 are better compared with Case 3.

Table 2. Evaluation metrics of decision tree for Case 1, Case 2, and Case 3.

	Case 1	Case 2	Case 3
Mean Absolute Error (MAE)	0.0414	0.0414	1.4932
The correlation coefficient (r)	0.7169	0.7332	0.6894
The coefficient of determination (R^2)	0.4197	0.4413	0.3613

However, the objective of this study is to propose a feasible solution that helps to apply the best input data into the machine learning model for PV output power with good results. Referring to the correlation coefficient (r), Case 2 is better than Case 1 and Case 3, while a low coefficient of determination (R^2) can be an indicator of imprecision predictions. However, R^2 does not tell us directly whether the predictions are sufficiently precise for our requirements. Still, Case 2 is found to be superior to Case 1 and 3.

Figure 8 presents a graph of the actual and predicted power for Case 1, Case 2, and Case 3 using the DTR model for the year 2020, while Figure 9 shows the actual and predicted power for a day. In comparison to the results from Table 2, Figures 8 and 9, it can be concluded that Case 2 is found to be the best input data for machine learning in all respects.

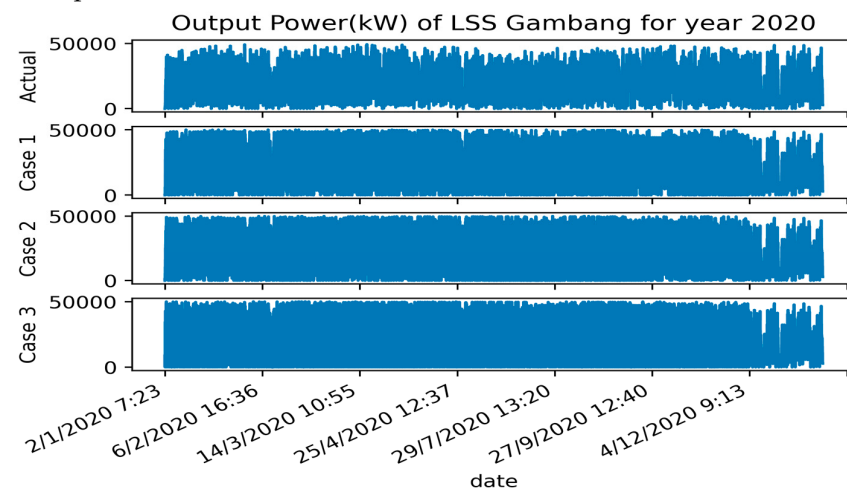


Figure 8. Actual and predicted power result of the year 2020.

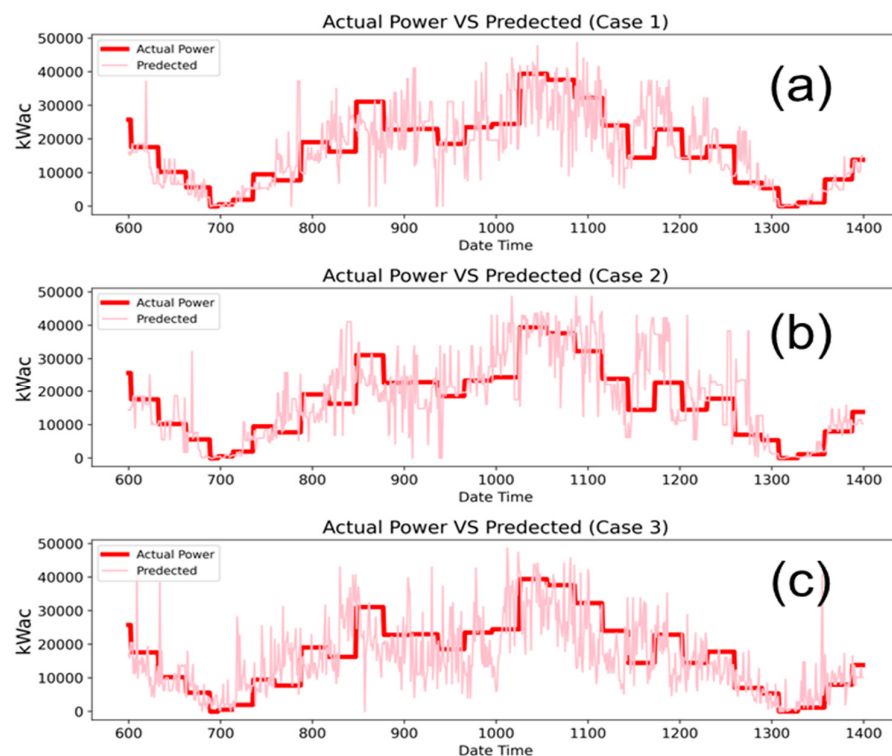


Figure 9. The actual power compared with Predicted Power for a day (a) Case 1, (b) Case 2, and (c) Case 3.

4. Conclusions

The paper presents a comparison of various types of input data applying DTR algorithm to forecast a LSS PV at Pahang. The meteorological variables considered in this study are total global irradiance (tghi), total horizontal irradiance (thi), global irradiance module plane (gimp), total slope irradiance (tsi), ambient temperature (at), PV module temperature (pvmt), wind speed (ws) and Power (kW). The authors analyzed and evaluated 36 datasets from five weather stations with seven parameters for each station, a 22-unit inverter, and 1-unit PQM Meter Grid as a big dataset were acquired from UiTM LSS PV. The exploratory data analysis was employed to identify the pattern of the dataset and the correlation between input parameters with power (kW). Three cases were tested to find the best input dataset using the machine learning algorithm. The results show that Case 1 and Case 2 have higher MAEs than Case 3. However, in terms of the correlation coefficient (r), Case 2 outperforms Cases 1 and 3, while a low coefficient of determination (R^2) can be found for Case 3, which gives an indication of imprecision predictions. Particularly, R^2 does not tell us directly whether the predictions are sufficiently precise for our requirements or not. Thus, Case 1 and Case 2 are better compared with Case 3. Conversely, Case 2 can be implemented in another machine learning algorithm. In future work, further tests will be done to verify if a longer historical dataset can further reduce prediction errors. Moreover, the model performance will be verified on other solar plant test cases.

Author Contributions: Conceptualization, N.A.; methodology, N.M.K.; software, N.M.K.; validation, N.A. and M.A.I.; formal analysis, M.A.I.; investigation, N.M.K.; resources, A.A.A.; data curation, S.S.; writing—original draft preparation, N.M.K.; writing—review and editing, N.A. and M.A.I.; visualization, M.A.I.; supervision, M.A.I. and N.A.; project administration M.Y.M.Y.; funding acquisition, S.K.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Universiti Tenaga Nasional (@UNITEN, The Energy University) through TNB Seeding Grant with project code of U-TE-RD-20-16. Authors also deeply appreciate the publication support from the iRMC of UNITEN with the code “BOLDREFRESH2025—CENTRE OF EXCELLENCE”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are provided by TNB Renewables Sdn. Bhd located at one of LSS PV in Pahang. Data will be available from the corresponding authors upon request

Acknowledgments: We thank the TNB Renewables Sdn. Bhd for accepting us for the site visit and providing data on the LSS PV.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Soonmin, H.; Lomi, A.; Okoroigwe, E.C.; Urrego, L.R. Investigation of Solar Energy: The Case Study in Malaysia, Indonesia, Colombia and Nigeria. *Int. J. Renew. Energy Res.* **2019**, *9*, 86–95.
2. Abdullah, W.S.W.; Osman, M.; Kadir, M.Z.A.A.; Verayiah, R. The potential and status of renewable energy development in Malaysia. *Energies* **2019**, *12*, 2437. [\[CrossRef\]](#)
3. Cheong, K.H.; Tai, V.C.; Tan, Y.C.; Rahman, N.F.A.; Chiong, K.S.; Saw, L.H. An outlook on large-scale solar power production in Peninsular Malaysia for scenario year 2030. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *463*, 012154. [\[CrossRef\]](#)
4. Hussain, T.N.; Sulaiman, S.I.; Musirin, I.; Shaari, S.; Zainuddin, H. A hybrid artificial neural network for grid-connected photovoltaic system output prediction. In Proceedings of the 2013 IEEE Symposium on Computers & Informatics (ISCI), Langkawi, Malaysia, 7–9 April 2013; pp. 108–111. [\[CrossRef\]](#)
5. Gao, M.; Li, J.; Hong, F.; Long, D. Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM. *Energy* **2019**, *187*, 115838. [\[CrossRef\]](#)
6. Sun, M.; Feng, C.; Zhang, J. Probabilistic solar power forecasting based on weather scenario generation. *Appl. Energy* **2020**, *266*, 114823. [\[CrossRef\]](#)
7. Sharma, A.; Kakkar, A. Forecasting daily global solar irradiance generation using machine learning. *Renew. Sustain. Energy Rev.* **2018**, *82*, 2254–2269. [\[CrossRef\]](#)

8. Ciulla, G.; D'Amico, A.; Lo Brano, V.; Traverso, M. Application of optimized artificial intelligence algorithm to evaluate the heating energy demand of non-residential buildings at European level. *Energy* **2019**, *176*, 380–391. [\[CrossRef\]](#)
9. Yang, L.B. Application of Artificial Intelligence in Electrical Automation Control. *Procedia Comput. Sci.* **2020**, *166*, 292–295. [\[CrossRef\]](#)
10. Sharadga, H.; Hajimirza, S.; Balog, R.S. Time series forecasting of solar power generation for large-scale photovoltaic plants. *Renew. Energy* **2020**, *150*, 797–807. [\[CrossRef\]](#)
11. Stuber, M. A Differentiable Model for Optimizing Hybridization of Industrial Process Heat Systems with Concentrating Solar Thermal Power. *Processes* **2018**, *6*, 76. [\[CrossRef\]](#)
12. Shi, J.; Lee, W.-J.; Liu, Y.; Yang, Y.; Wang, P. Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines. *IEEE Trans. Ind. Appl.* **2012**, *48*, 1064–1069. [\[CrossRef\]](#)
13. Jiang, H.; Hong, L. Application of BP Neural Network to Short-Term-Ahead Generating Power Forecasting for PV System. *Adv. Mater. Res.* **2012**, *608–609*, 128–131. [\[CrossRef\]](#)
14. Mellit, A.; Sağlam, S.; Kalogirou, S.A. Artificial neural network-based model for estimating the produced power of a photovoltaic module. *Renew. Energy* **2013**, *60*, 71–78. [\[CrossRef\]](#)
15. Chen, C.; Duan, S.; Cai, T.; Liu, B. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Sol. Energy* **2011**, *85*, 2856–2870. [\[CrossRef\]](#)
16. Mellit, A.; Pavan, A.M.; Oglari, E.; Leva, S.; Lughi, V. Advanced methods for photovoltaic output power forecasting: A review. *Appl. Sci.* **2020**, *10*, 487. [\[CrossRef\]](#)
17. Dahlan, N.Y.; Zamri, M.S.M.; Zaidi, M.I.A.; Azmi, A.M.; Zailani, R. Forecasting Generation of 50MW Gambang Large Scale Solar Photovoltaic Plant Using Artificial Neural Network-Particle Swarm Optimization. *Int. J. Renew. Energy Res.* **2022**, *12*, 10–18.
18. Nguyen, N.Q.; Bui, L.D.; Van Doan, B.; Sanseverino, E.R.; Cara, D.; Nguyen, Q.D. A new method for forecasting energy output of a large-scale solar power plant based on long short-term memory networks a case study in Vietnam. *Electr. Power Syst. Res.* **2021**, *199*, 107427. [\[CrossRef\]](#)
19. Malvoni, M.; Hatziargyriou, N. One-day ahead PV power forecasts using 3D Wavelet Decomposition. In Proceedings of the 2019 International Conference on Smart Energy Systems and Technologies (SEST), Porto, Portugal, 9–11 September 2019; pp. 1–6.
20. Ahmed, R.; Sreeram, V.; Mishra, Y.; Arif, M.D. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renew. Sustain. Energy Rev.* **2020**, *124*, 109792. [\[CrossRef\]](#)
21. Wang, F.; Xuan, Z.; Zhen, Z.; Li, K.; Wang, T.; Shi, M. A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. *Energy Convers. Manag.* **2020**, *212*, 112766. [\[CrossRef\]](#)
22. Global Solar Atlas. Site Info Gambang, Pahang. 2023. Available online: <https://globalsolaratlas.info/map?c=3.707243,103.096502,11&s=3.707243,103.096502&m=site> (accessed on 13 February 2023).
23. Ibrahim, N.A.; Wan Alwi, S.R.; Manan, Z.A.; Mustaffa, A.A.; Kidam, K. Impact of Extreme Temperature on Solar Power Plant in Malaysia. *Chem. Eng. Trans.* **2022**, *94*, 343–348. [\[CrossRef\]](#)
24. Suruhanjaya Tenaga (Energy Commission). *Guidelines on Large Scale Solar Photovoltaic Plant for Connection to Electricity Networks (Act A1501), 3rd revision*; Suruhanjaya Tenaga (Energy Commission): Putrajaya, Malaysia, 2020; pp. 32–36.
25. Singh, N.; Jena, S.; Panigrahi, C.K. A novel application of Decision Tree classifier in solar irradiance prediction. *Mater. Today Proc.* **2022**, *58*, 316–323. [\[CrossRef\]](#)
26. Singh Kushwah, J.; Kumar, A.; Patel, S.; Soni, R.; Gawande, A.; Gupta, S. Comparative study of regressor and classifier with decision tree using modern tools. *Mater. Today Proc.* **2022**, *56*, 3571–3576. [\[CrossRef\]](#)
27. Zulkifly, Z.; Baharin, K.A.; Gan, C.K. Improved Machine Learning Model Selection Techniques for Solar Energy Forecasting Applications. *Int. J. Renew. Energy Res.* **2021**, *11*, 308–319.
28. Wu, W.; Tang, X.; Lv, J.; Yang, C.; Liu, H. Potential of Bayesian additive regression trees for predicting daily global and diffuse solar radiation in arid and humid areas. *Renew. Energy* **2021**, *177*, 148–163. [\[CrossRef\]](#)
29. Rahman, M.M.; Shakeri, M.; Tiong, S.K.; Khatun, F.; Amin, N.; Pasupuleti, J.; Hasan, M.K. Prospective methodologies in hybrid renewable energy systems for energy prediction using artificial neural networks. *Sustainability* **2021**, *13*, 2393. [\[CrossRef\]](#)
30. Haider, S.A.; Sajid, M.; Sajid, H.; Uddin, E.; Ayaz, Y. Deep learning and statistical methods for short- and long-term solar irradiance forecasting for Islamabad. *Renew. Energy* **2022**, *198*, 51–60. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.