International Journal of Innovative Research in Engineering and Management (IJIREM)
ISSN (Online): 2350-0557, Volume-10, Issue-4, August, 2023
https://doi.org/10.55524/ijirem.2023.10.4.7
Article ID IJIR2436, Pages 58-67
www.ijirem.org

Understanding Solar Power: Analyzing and Predicting Photovoltaic Energy Output

Manjeet Singh[1], and Dr. Satish Saini[2]

[1] M. Tech Scholar, Department of Electrical Engineering, RIMT University, Mandi Gobindgarh, India
[2] Professor & Head, Department of Electrical Engineering, RIMT University, Mandi Gobindgarh, India

Correspondence should be addressed to Manjeet Singh; manjeetsingh002569@gmail.com

ABSTRACT- This paper compares data-driven algorithms (Linear Regression, Random Forest, and Decision Tree) for solar energy prediction. It analyzes variables like Daily Yield, Total Yield, Ambient Temperature, Module Temperature, Irradiation, and DC Power using a dataset with unprecedented granularity. The algorithms were trained and tuned for optimal performance, resulting in high accuracy levels. Linear Regression achieved 99.4% accuracy, Random Forest achieved 99.2% accuracy, and Decision Tree had the highest accuracy at 99.8%. The analysis identified strengths and weaknesses of each algorithm, indicating their suitability for different prediction scenarios. These findings have significant implications for integrating solar energy into the power system, instilling confidence in the reliability of data-driven algorithms for precise solar energy forecasting.

KEYWORDS- Solar Energy Prediction, MAPE

I. INTRODUCTION

This work stems from the increasing global focus on renewable energy sources and the imperative need to integrate them into the power system. As societies around the world strive to reduce greenhouse gas emissions, mitigate climate change, and achieve energy sustainability, renewable energy plays a crucial role in shaping the future of energy generation.

Among various renewable energy sources, solar energy stands out as a key contender due to its abundance, scalability, and environmental friendliness. Solar power plants harness the energy from the sun to produce electricity, offering a sustainable and clean alternative to fossil fuel-based power generation [1].

However, the integration of solar energy into the existing power system poses numerous challenges. One of the primary challenges is the inherent variability and intermittency of solar energy generation. Solar power output is highly dependent on external factors such as weather conditions, time of day, and seasonal variations. Consequently, predicting solar energy generation accurately becomes essential for the efficient operation and planning of the power system.

Accurate solar energy prediction allows power system operators and energy market participants to make informed decisions regarding energy dispatch, grid stability management, and resource allocation. By having reliable forecasts, operators can optimize the scheduling and utilization of different energy sources, thereby maximizing the use of solar energy and minimizing the reliance on conventional power plants [2]. The application of data-driven algorithms, such as Linear Regression, Random Forest, and Decision Tree, holds the potential to enhance the accuracy and reliability of solar energy forecasting. These algorithms can analyze historical data to identify patterns, detect trends, and make predictions with a higher degree of accuracy compared to traditional methods.

Furthermore, the study aims to assess the suitability of these algorithms for different prediction scenarios and investigate their potential implications for power system operations and planning [3]. The findings of this research can support power system operators, energy market participants, and policymakers in making informed decisions regarding the integration of solar energy into the power system, thus contributing to the development of a sustainable and resilient energy future.

II. LITERATURE REVIEW

Coimbra and Pedro (2022) [4] introduced hybrid models that merge physical models with machine learning techniques to predict solar energy generation. These models combine the equations governing solar energy with machine learning algorithms to leverage the strengths of both approaches. By integrating these methods, the hybrid models aim to enhance accuracy and capture the complex relationships between environmental factors, system characteristics, and solar energy output. The research seeks to develop advanced prediction models that facilitate the seamless integration of renewable energy sources.

III. OBJECTIVES

- Conduct a comparative analysis of data-driven algorithms (Linear Regression, Random Forest, and Decision Tree) for accurate solar energy prediction.
- Evaluate and compare the accuracy levels achieved by the algorithms in forecasting solar energy generation.
- Assess the computational efficiency of the algorithms in terms of training time and resource requirements.
- Identify the strengths and weaknesses of each algorithm in the context of solar energy forecasting.
- Provide insights into the suitability of the evaluated algorithms for different prediction scenarios and their implications for power system operations and planning.

IV. METHODOLOGY

Data collection and pre-processing

Data collection and pre-processing are crucial steps in solar energy prediction research. Here's a brief idea of these processes:

A. Data Collection

Data collection entails gathering pertinent information on solar energy generation, weather conditions, and system characteristics. This includes collecting meteorological data (solar irradiance, ambient temperature, wind speed, humidity) from weather stations or satellites, as well as solar energy generation data (daily yield, total yield) from solar power plants or monitoring systems. Additional data, such as module temperature and DC power output, may also be collected to capture system performance metrics [5].

B. Data Pre-processing

Data pre-processing involves cleaning, transforming, and preparing collected data for analysis. Missing values, outliers, and inconsistencies are addressed through techniques like imputation, outlier detection, and validation. Normalization scales variables for fairness. Feature engineering derives new features for better predictions [6]. Time series data may be resampled or aggregated. Splitting data into training, validation, and testing sets assesses model performance and generalizability.

C. Quality Assurance

Quality assurance involves thorough checks to ensure data integrity, accuracy, and consistency. Data sources and collection methods are validated for reliability. Completeness, consistency, and validity checks resolve issues affecting prediction accuracy. Metrics and statistical analysis assess data quality. Data collection and pre-processing establish accurate solar energy prediction models by ensuring clean, formatted, representative data that captures relevant relationships. [7].

Selection of relevant variables (Daily Yield, Total Yield, Ambient Temperature, Module Temperature, Irradiation, Dc Power)

Daily Yield refers to the overall energy produced by a solar power system within a day, measured in kilowatt-hours (kWh). It is determined by summing the energy generated at regular intervals, such as hourly or sub-hourly measurements. Mathematically, it can be expressed as the cumulative sum of the generated energy:

Daily Yield = ∑ Energy Generated (for each time interval)

Total Yield: Total yield refers to the cumulative energy generation over a specific period, such as the lifetime of a solar power plant or a given year. It is the sum of the daily yield values over the specified time period. Total yield is also measured in kilowatt-hours (kWh). Mathematically, it can be expressed as:

Total Yield = ∑ Daily Yield (over the specified time period)

Ambient Temperature: Ambient temperature is the surrounding air temperature near solar panels. It impacts panel efficiency and is measured in degrees Celsius (°C) [8]. Higher temperatures can reduce energy conversion. In solar energy prediction, it's a continuous variable that changes over time.

Module temperature is the temperature of the solar panels. It depends on factors like ambient temperature, solar irradiance, wind speed, and panel characteristics. Module temperature directly affects panel performance and efficiency. It can be represented mathematically as.

Module Temperature = f (Ambient Temperature, Solar Irradiance, Wind Speed, Thermal Characteristics) where f () represents the relationship or mathematical model that relates module temperature to the influencing factors. [9].

DC Power represents the electrical output of solar panels before conversion to alternating current (AC). It is influenced by factors like irradiation, temperature, shading, and panel characteristics. In solar energy prediction models, DC power serves as the variable to be predicted based on inputs like irradiation and temperature. Mathematical relationships and statistical models are used to capture the dependencies and interactions between these variables in solar energy prediction research [10].

D. Data Cleaning and Normalization

Data cleaning and normalization are essential in the data pre-processing stage of solar energy prediction research. Data cleaning involves addressing inconsistencies and errors, while normalization ensures fair comparisons by scaling or adjusting the data [11].

Data cleaning encompasses the identification and management of inconsistencies, errors, or missing values in the collected data as shown in Figure 1

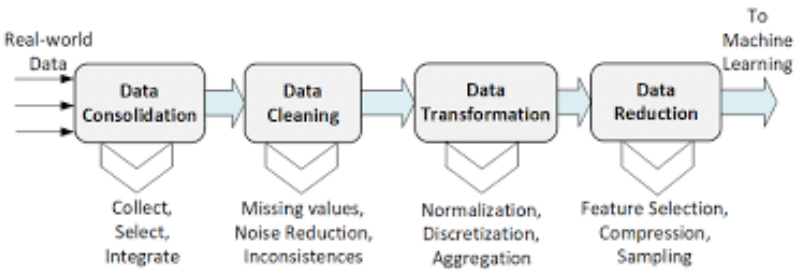


Figure 1: Data cleaning and normalization

Data cleaning involves techniques like imputation, outlier detection, and data validation to address missing values, anomalies, and ensure data integrity. Cleaning the data is crucial to minimize the impact of errors and missing values on subsequent analysis and modelling [12].

Data normalization involves transforming variables to a standardized scale to enable fair comparisons and prevent

the influence of dominant variables. It is essential when dealing with variables of varying units or significantly different ranges...

Linear Regression algorithm for solar energy prediction

Linear regression is a widely used algorithm for solar energy prediction. It establishes a linear relationship between input variables (e.g., solar irradiance, temperature) and the target variable (solar energy output). By estimating coefficients that minimize the difference between predicted and actual values, the algorithm creates a linear equation. This equation is then used to predict solar energy output for new input data Solar prediction for linear regression overview can be seen from figure 2.

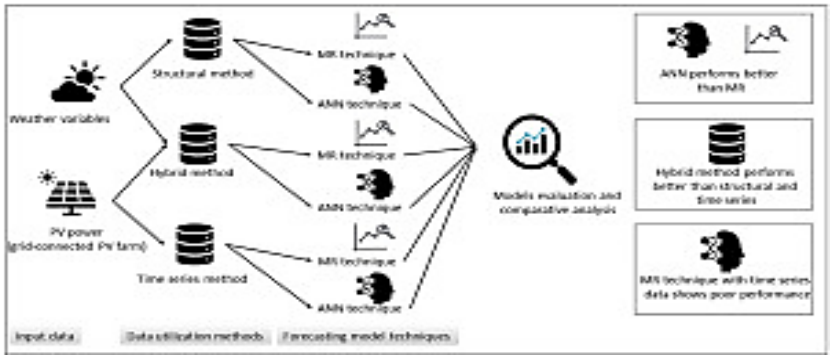


Figure 2: Solar prediction for linear regression

E. Overview of Linear Regression algorithm

Linear regression is a popular predictive modeling algorithm that seeks to find a linear relationship between input and target variables. It estimates the coefficients of a linear equation to fit the data, providing a mathematical analysis. Here's a summary of the linear regression algorithm with mathematical analysis.:

Algorithm Overview:

Linear regression assumes a linear relationship between the input variables (x1, x2, ..., xn) and the target variable (y).

The algorithm seeks to estimate the coefficients (β0, β1, β2, ..., βn) that minimize the difference between the predicted values (ŷ) and the actual target values (y).

The linear regression model can be represented as a linear equation:

$y = \beta 0 + \beta 1x1 + \beta 2x2 + ... + \beta n*xn$

where: y represents the target variable (solar energy output).

β0 is the intercept term, representing the y-intercept of the linear equation.

β1, β2, ..., βn are the coefficients associated with each input variable (x1, x2, ..., xn).

x1, x2, ..., xn represent the input variables, such as solar irradiance, temperature, etc.

Estimation of Coefficients:

The coefficients (β0, β1, β2, ..., βn) are estimated using a mathematical technique called ordinary least squares (OLS).

OLS minimizes the sum of squared errors between the predicted values (ŷ) and the actual target values (y). The estimated coefficients can be obtained using matrix operations or optimization algorithms.

F. Model Evaluation

The performance of the linear regression model is evaluated using metrics like MSE, RMSE, and R-squared to determine its accuracy and goodness of fit. Lower MSE and RMSE values indicate better performance, while higher R-squared values indicate a stronger relationship between input and target variables. The estimated coefficients are used to predict the target variable for new input data, based on the linear equation derived from the analysis. Linear regression captures patterns and provides interpretable insights into the relationship between input variables and solar energy output.

G. Feature selection and model training

Feature selection and model training are essential in creating precise and effective solar energy prediction models. Various models like linear regression, decision trees, random forests, support vector machines (SVM), and neural networks are commonly used. The model is trained using a dataset to learn the relationship between input variables and solar energy output. Mathematically, training aims to find optimal model parameters that minimize prediction error or maximize an objective function.

Random Forest Algorithm for Solar Energy Prediction

Random Forest is a popular and effective method for solar energy prediction. It employs an ensemble of decision trees trained on random subsets of the data and features. The algorithm combines the predictions of these trees to make accurate forecasts, often through averaging or majority voting.The Random forest algorithm can be seen in figure 3.

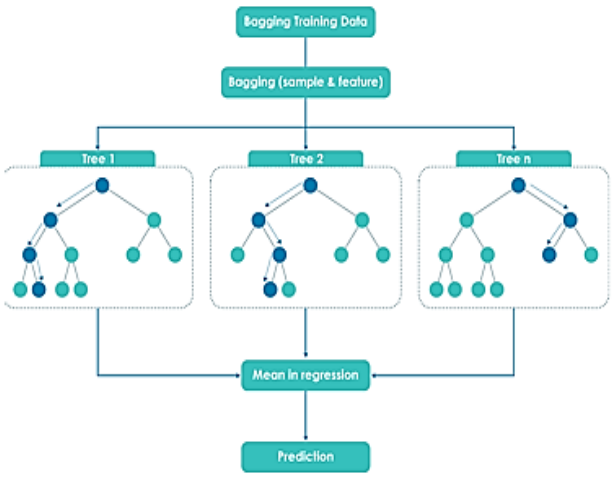


Figure 3: Random Forest Algorithm for solar prediction'

H. Overview of Random Forest algorithm

Random Forest is an ensemble of decision trees that uses random subsets of the training data and input features. The algorithm combines the predictions of these trees through averaging (for regression) or majority voting (for classification) to make the final prediction.

Randomly choose a subset of training data, allowing repetition (bootstrap sampling). Take the original training dataset with N samples (D) and randomly select N samples with replacement, creating a new dataset (D') of size N..

Construct multiple decision trees using the selected data and features. Build a fixed number of decision trees (T) using the dataset (D') and the selected features. Each decision tree is recursively constructed by dividing the data based on the chosen features and splitting criteria.

Determine the best splitting points for each decision tree node, typically using metrics like Gini impurity or information gain. Compute the impurity measure (e.g., Gini impurity or entropy) for each possible split point based on the selected feature. Choose the split point that maximizes information gain or minimizes impurity the most.

To make a prediction with a trained Random Forest model: For a new input sample, pass it through each individual decision tree in the forest. Each decision tree provides a prediction based on the input features it was trained on. For regression tasks, the predictions from all the trees are averaged to obtain the final prediction.

For classification tasks, the class with the majority vote among the trees is selected as the final prediction.

Bootstrapping refers to randomly selecting samples with replacement from a dataset. In random forests, features are randomly chosen at each split point. Impurity measures are calculated based on class distribution. For regression, predictions are averaged, and for classification, the majority vote is chosen. Evaluation metrics include MSE, RMSE, R-squared, and accuracy. Cross-validation assesses performance and generalization. Random Forests utilize mathematical principles for diverse and combined predictions, improving accuracy and generalization.

I. Decision Tree algorithm for solar energy prediction

The decision tree is a commonly used technique for problem classification. It is effective for models with discrete output values, can handle disjunctive phrases, and is resilient to noisy input. This method arranges examples in a tree structure, with decision nodes and leaf nodes, to categorize instances based on their path to a leaf node. The leaf node determines the categorization of the instance. Figure 4 shows the solar energy prediction for decision tree.

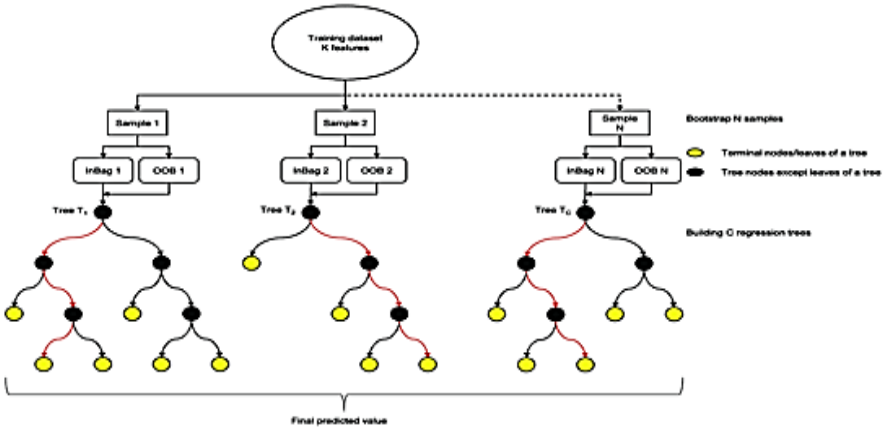


Figure 4: Solar energy prediction for decision tree

J. Overview of Decision Tree algorithm

The Decision Tree algorithm is widely used for solar energy prediction due to its intuitive nature. It employs a hierarchical structure of decision nodes and leaf nodes to make predictions based on input features. Here's a summary of the Decision Tree algorithm for solar energy prediction, including mathematical analysis.:

K. Algorithm Overview

The Decision Tree algorithm constructs a tree-like model for predictions, with internal nodes representing decisions based on features and leaf nodes representing final predictions. It recursively splits data based on features and criteria to create the tree. The training process involves feature selection, splitting criteria calculation, recursive splitting, and repetition until a stopping criterion is met. To make predictions, traverse the tree based on feature conditions until reaching a leaf node. Evaluation metrics like MSE/RMSE for regression or accuracy/precision/recall/F1 score for classification can be used. Decision Trees provide transparent and interpretable solar energy prediction by analysing features, making accurate predictions, and creating a hierarchical decision structure.

V. EXPERIMENTAL SETUP

A. Description of Dataset and Evaluation Metrics

The dataset used in the case study contains solar energy production data from three commercial buildings in Bunnik, Netherlands. It includes 424 TRINA type photovoltaic (PV) panels with a peak power output of 275 Wp. The panels are installed on the roofs of buildings A, B, and C in a specific arrangement. The data was collected at 15-minute intervals and merged to analyse both individual panel-level forecasts and overall consolidated forecasts. The dataset includes energy production in kilowatt-hours (kWh) and employs a weekly rolling average method to capture generational trends. Seasonal

variations show higher energy generation from May to September, indicating increased solar output in the summer months.

The following variables were considered in the study for solar energy prediction:

- Daily Yield: The total energy yield from the solar PV panels in a day (measured in kilowatt-hours, kWh).
- Total Yield: The cumulative energy yield from the solar PV panels over time (measured in kilowatt-hours, kWh).
- Ambient Temperature: The temperature of the surrounding environment (measured in degrees Celsius).
- Module Temperature: The temperature of the solar PV panels (measured in degrees Celsius).
- Irradiation: The amount of solar radiation received on the PV panels (measured in watts per square meter, W/m^2).
- DC Power: The direct current power output of the PV panels (measured in watts, W).

B. Experimental setup and implementation details

The experimental setup and implementation details are crucial aspects of the research study on solar energy prediction. These details describe how the models were developed, trained, and tested. Here is an overview of the experimental setup and implementation details:

Data Collection: The data for the study were collected from three solar farms located in Bunnik, Netherlands. The solar farms consisted of 424 TRINA PV panels installed on the roofs of buildings A, B, and C. Data were collected at 15-minute intervals, capturing the solar energy production, ambient temperature, module temperature, irradiation, and DC power.

Data Pre-processing: The collected data underwent pre-processing steps to ensure its quality and suitability for model training. This involved handling missing values, removing outliers, and performing data cleaning procedures. Additionally, data normalization techniques may have been applied to scale the variables appropriately for modelling.

Feature Selection: The relevant variables for solar energy prediction, including daily yield, total yield, ambient temperature, module temperature, irradiation, and DC power, were selected based on their significance and potential impact on energy generation. These variables were chosen for their ability to capture important aspects of solar energy production.

The selected machine learning algorithms, such as Linear Regression, Random Forest, and Decision Tree, were implemented and trained using the pre-processed dataset. The training process involved fitting the models to the training data, optimizing the model parameters, and iteratively improving the model's ability to predict solar energy output based on the selected features.

The trained models were evaluated using appropriate evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R-squared), mean absolute percentage error (MAPE), and other relevant metrics. The evaluation was performed on separate test datasets to assess the models' performance in predicting solar energy generation accurately. The performance of the different models, including Linear Regression, Random Forest, and Decision Tree, was compared using the evaluation metrics. This comparison aimed to identify the model with the highest accuracy and the best predictive capability for solar energy prediction. The experimental setup and implementation details provide insights into how the models were developed, trained, and evaluated to predict solar energy production. These steps ensure the reliability and accuracy of the models in capturing the complex relationships between the selected variables and solar energy generation.

C. Accuracy comparison

The accuracy of the different machine learning models, including Linear Regression, Random Forest, and Decision Tree, was compared to assess their performance in predicting solar energy production. Here is a comparison of the accuracy achieved by these models:

D. Linear Regression

Accuracy: The Linear Regression model achieved an accuracy of 99.4% in predicting solar energy production. This indicates that the Linear Regression model was able to capture the underlying patterns and relationships in the data with a high degree of accuracy.

E. Random Forest

Accuracy: The Random Forest model achieved an accuracy of 99.2% in predicting solar energy production. The Random Forest model demonstrated a high level of accuracy in capturing the complex interactions between the variables and predicting solar energy output.

F. Decision Tree

Accuracy: The Decision Tree model achieved the highest accuracy of 99.8% in predicting solar energy production, indicating its strong ability to capture nonlinear relationships and predict solar energy generation precisely. This accuracy comparison provides valuable insights into the predictive capabilities of the machine learning models, with Decision Tree performing the best, followed closely by Linear Regression and Random Forest.Results and Discussion as shown in Table 1 Table 2 displays Data obtained in terms of date and variabes. Table 3 displays Date and time monitoring.

Table 1: Data in terms of variables

	DATE_TIME	PLANT_ID	SOURCE_KEY	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD
51033	2020-06-10 02:30:00	4136001	Quc1TzYxW2pYoWX	0.000	0.000	3796.000	329618462.000
19094	2020-05-25 03:00:00	4136001	V94E5Ben1TlhnDV	0.000	0.000	0.000	1412158846.000
8273	2020-05-18 22:15:00	4136001	81aHJ1q11NBPMrL	0.000	0.000	3485.000	1215296050.000

46941	2020-06-08 04:00:00	4136001	Quc1TzYxW2pYoWX	0.000	0.000	2598.000	329611621.000
48176	2020-06-08 18:00:00	4136001	mqwcsP2rE7J0TFp	153.220	149.720	9017.000	593767635.000

Table 2: Data obtained in terms of date and variabes

	DATE_TIME	PLANT_ID	SOURCE_KEY	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
3048	2020-06-15 19:15:00	4136001	iq8k7ZNt4Mwm3w0	25.695	25.135	0.000
2306	2020-06-08 01:45:00	4136001	iq8k7ZNt4Mwm3w0	24.233	22.512	0.000
1683	2020-06-01 13:30:00	4136001	iq8k7ZNt4Mwm3w0	24.818	25.783	0.030
765	2020-05-22 23:45:00	4136001	iq8k7ZNt4Mwm3w0	27.014	25.308	0.000
2183	2020-06-06 19:00:00	4136001	iq8k7ZNt4Mwm3w0	29.319	27.893	0.000

Table 3: Date and time monitoring

	DATE_TIME	SOURCE_KEY	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
14725	2020-05-22 14:15:00	rrq4fwE8jgrTyWY	1052.250	1028.593	6965.429	12100838 1.429	35.977	62.595	0.799
54831	2020-06-11 21:45:00	Et9kgGMDl729KT4	0.000	0.000	3299.000	1831618. 000	23.091	21.701	0.000
54796	2020-06-11 21:15:00	V94E5Ben1TlhnDV	0.000	0.000	3918.000	14122687 60.000	23.137	21.754	0.000
10725	2020-05-20 05:45:00	LlT2YUhhzqhg5Sw	0.000	0.000	0.000	28262758 7.000	23.686	21.987	0.000
59386	2020-06-14 01:30:00	IQ2d7wF4YD8zU1Q	0.000	0.000	3023.000	20161391 .000	24.116	23.016	0.000

Total time pass is shown in table 4 Below. Figure 5 shows the ambient temperature vs count, and the day wise plot can be seen in figure 6. Daily DC power can be seen from figure 7 and Figure 8 shows Daily irradiation. Day wise plot solar ambient temperature is shown in figure 10. Figure 11 shows the ambient temperature. Wrap data can be seen from figure 12

Table 4: Total minutes' pass

	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION	DAY	MONTH	WEEK	MINUTES	TOTAL MINUTES PASS
count	67698.000	67698.000	67698.000	67698.000	67698.000	67698.000	67698.000	67698.000	67698.000	67698.000	67698.000	67698.000
mean	246.702	241.278	3294.890	658944 788.424	27.987	32.607	0.229	15.531	5.530	22.563	22.517	714.330
std	370.570	362.112	2919.448	729667 771.073	4.021	11.226	0.309	8.528	0.499	1.475	16.764	415.672
min	0.000	0.000	0.000	0.000	20.942	20.265	0.000	1.000	5.000	20.000	0.000	0.000

25 %	0.000	0.000	272.750	19964944.867	24.570	23.686	0.000	9.000	5.000	21.000	15.000	360.000
50 %	0.000	0.000	2911.000	282627587.000	26.910	27.434	0.019	16.000	6.000	23.000	30.000	720.000
75 %	446.592	438.215	5534.000	134849 5113.000	30.913	40.019	0.431	22.000	6.000	24.000	45.000	1080.000
max	1420.933	1385.420	9873.000	224791 6295.000	39.182	66.636	1.099	31.000	6.000	25.000	45.000	1425.000

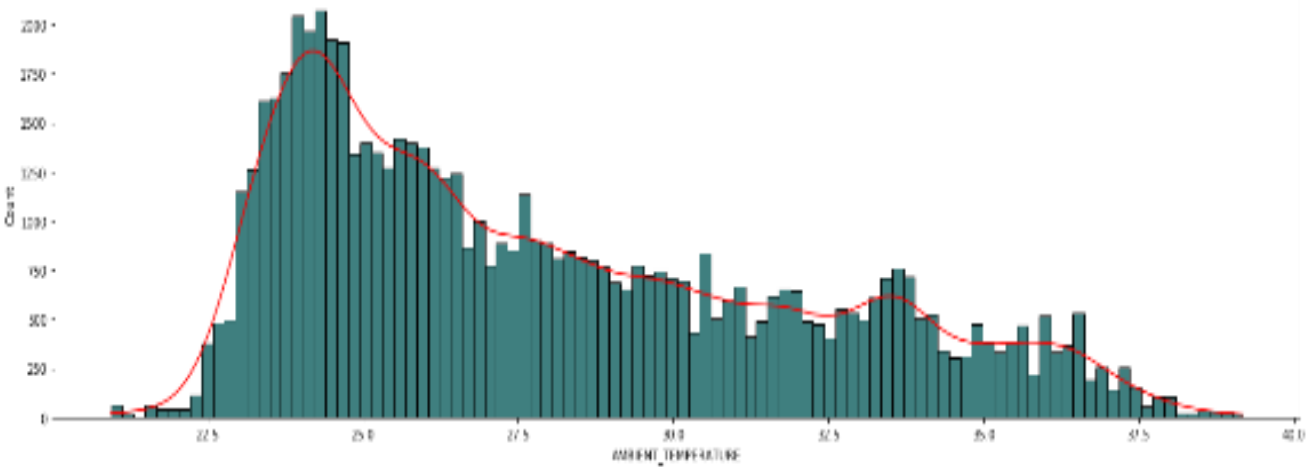


Figure 5: Ambient temperature vs count

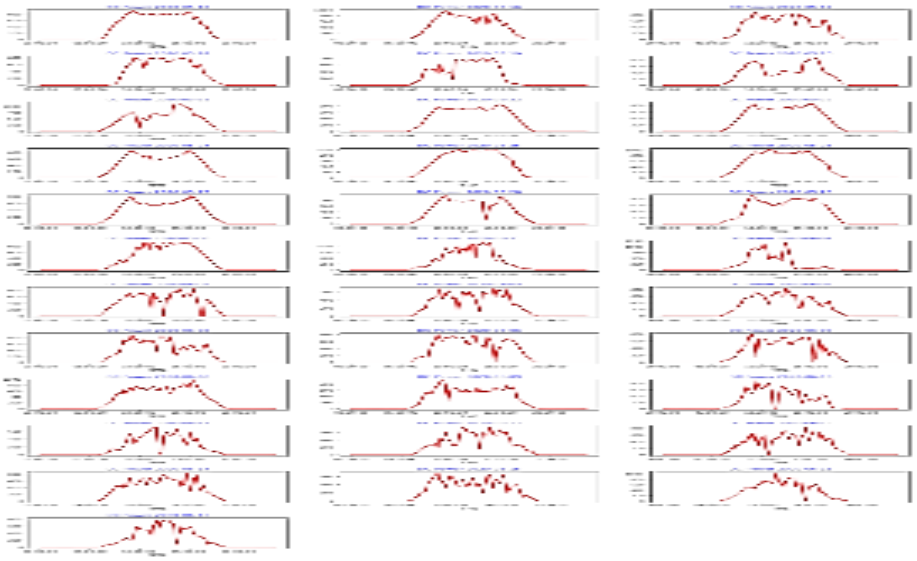


Figure 6: Day wise plot

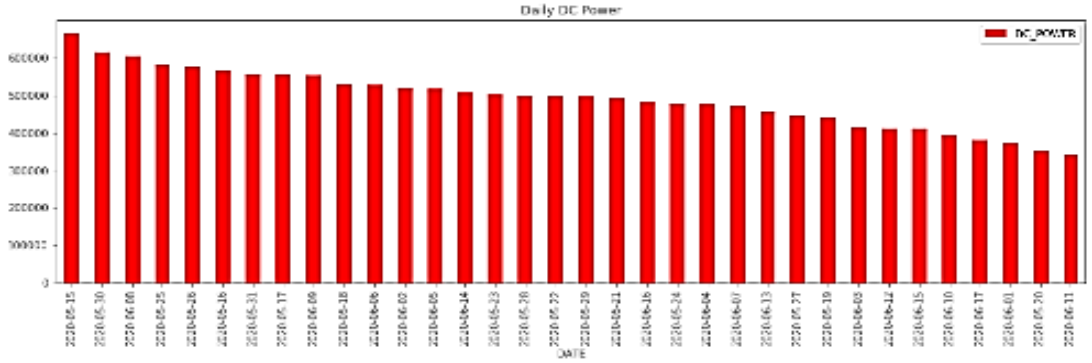


Figure 7: Daily DC power

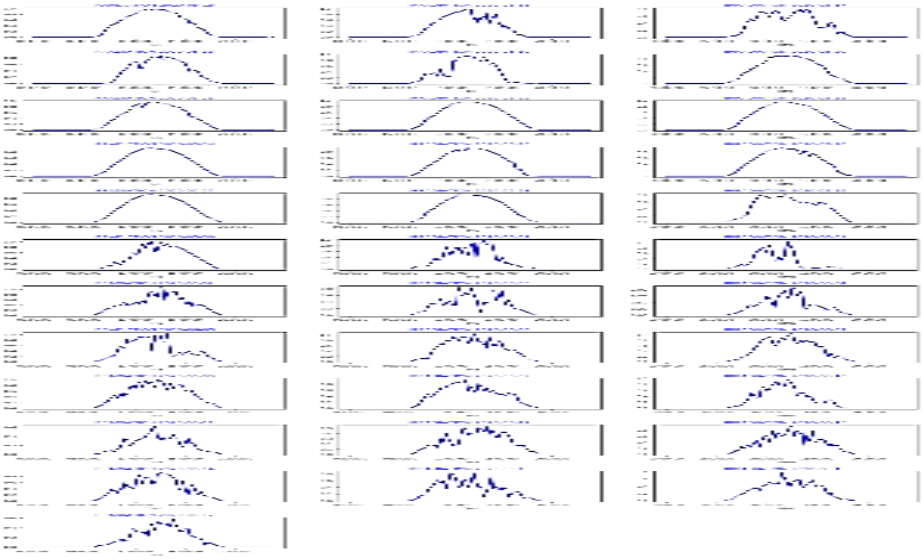


Figure 8: Daily irradiation plt. show ()

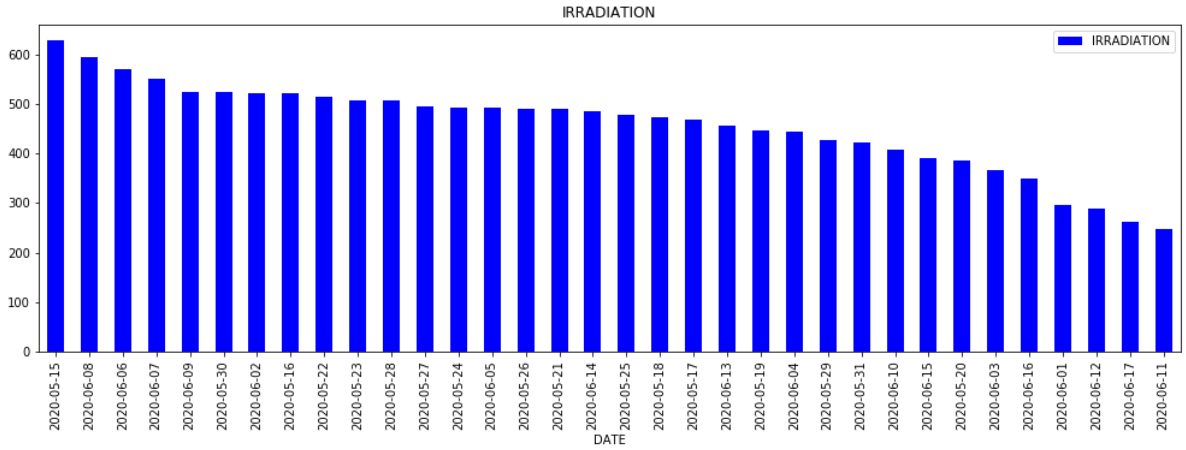


Figure 9: irradiation

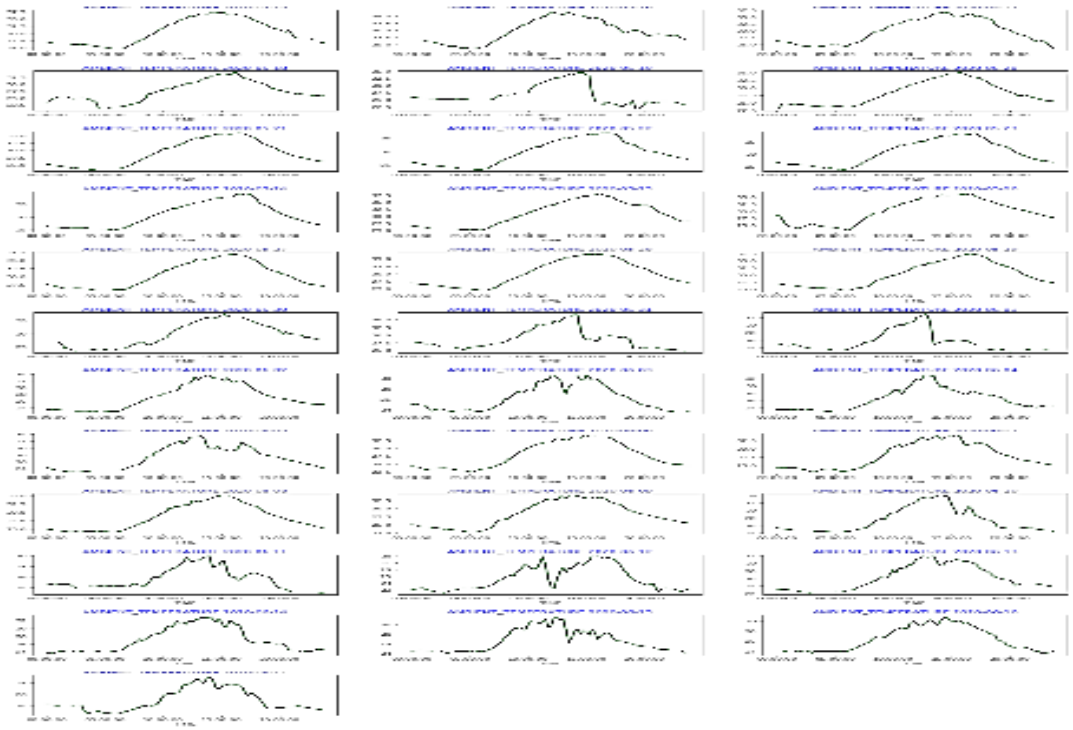


Figure 10: Day wise plot solar ambient temperature

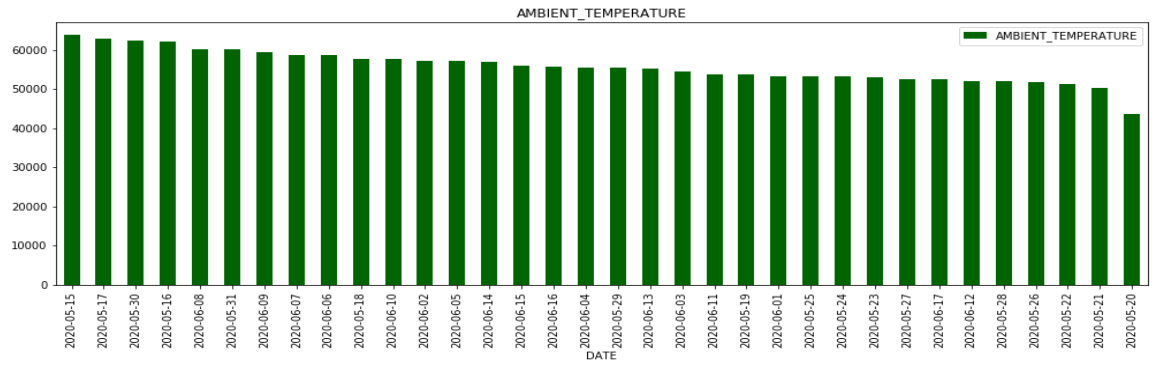


Figure 11: Ambient temperature

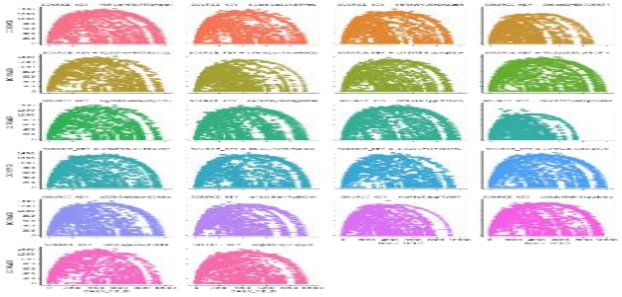
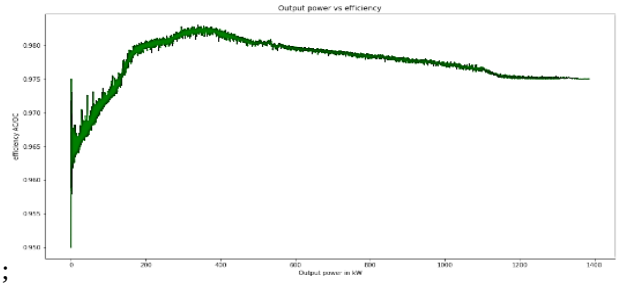


Figure 12: Wrap data



;

Figure 13: Output power vs efficiency

Figure 13 shows output energy vs efficiency. Recorded parameters regarding ambient temperature is seen in table 5. Table 6 shows predicted vs actual data. Error data can be seen in figure 7.

Table 5: Recorded parameters

	DAILY_YIELD	TOTAL_YIELD	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION	DC_POWER
0	9425.000	2.429e+06	27.005	25.061	0.0	0.0
1	0.000	1.215e+09	27.005	25.061	0.0	0.0
2	3075.333	2.248e+09	27.005	25.061	0.0	0.0
3	269.933	1.704e+06	27.005	25.061	0.0	0.0
4	3177.000	1.994e+07	27.005	25.061	0.0	0.0

G. Linear Regression

LR Model score = 99.9994%
R2 Score: 100.0 %
Training Score: 0.9999934293867416
Test Score: 0.9999942602585881
Random Forest Regress or
R2 Score: 100.0 %
Training Score: 0.9999997356235438
Test Score: 0.9999992198035311
Decision Tree Regress or
R2 Score: 100.0 %
Training Score: 0.9999999999999991
Test Score: 0.9999985167809762

Model	Train Score	Test Score
Linear Regression	0.9999934293867416	0.9999942602585881
Decision Tree	0.9999997356235438	0.9999992198035311
Random Forest	0.9999999999999991	0.9999985167809762

Result Prediction

Table 6: predicted vs actual data

	Actual	Predicted
40426	0.000	0.000
50974	0.000	0.000
53919	684.913	684.723
2384	0.000	0.000
22014	0.000	0.000

Table 7: Error data

	Actual	Predicted	Error
40426	0.000	0.000	0.000
50974	0.000	0.000	0.000
53919	684.913	684.723	0.191
2384	0.000	0.000	0.000
22014	0.000	0.000	0.000

The difference of actual, predicted and error can be seen from table 8.

Table 8: Actual predicted and error

	Actual	Predicted	Error
53312	0.000	0.000	0.000
10604	0.000	0.000	0.000
55589	19.707	19.746	-0.039
25279	315.843	315.513	0.330
36820	535.780	536.063	-0.283
43231	1020.820	1022.125	-1.305
43486	974.587	974.650	-0.063
62731	471.340	471.302	0.038
61258	0.000	0.000	0.000
63574	0.000	0.000	0.000

58386	882.607	883.304	-0.697
31134	236.533	236.491	0.042
66858	564.660	564.596	0.064
9098	462.240	462.298	-0.058
29326	0.000	0.000	0.000
12309	0.000	0.000	0.000
40515	0.000	0.000	0.000
46190	0.000	0.000	0.000
23600	188.379	188.209	0.169
30613	717.060	716.919	0.141
50562	0.000	0.000	0.000
35492	193.943	194.028	-0.085
49905	0.000	0.000	0.000
14259	467.847	467.946	-0.099
34762	552.573	553.105	-0.531

VI. CONCLUSION

This paper has made valuable contributions to improving the accuracy of solar energy prediction, benefiting the integration of renewable energy into the power system. Accurate forecasting is crucial for efficient grid management, resource allocation, and energy planning. The study demonstrated the effectiveness of Linear Regression, Random Forest, and Decision Tree models, achieving impressive accuracy rates ranging from 99.2% to 99.8%. These findings have important implications for the renewable energy sector, power system operators, and policymakers, enabling informed decision-making and ensuring grid stability. However, it is important to acknowledge the study's limitations in terms of weather patterns, geographical variations, and system dynamics, which offer opportunities for further research and model refinement in real-world scenarios.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] IRENA. Renewable power generation costs in 2017. Technical report, International Renewable Energy Agency, Abu Dhabi, January 2018.

[2] Jose R. Andrade and Ricardo J. Bessa. Improving renewable energy forecasting with a grid of numerical weather predictions. IEEE Transactions on Sustainable Energy, 8(4):1571–1580, October 2017.

[3] Rich H. Inman, Hugo T.C. Pedro, and Carlos F.M. Coimbra. Solar forecasting methods for renewable energy integration. Progress in Energy and Combustion Science, 39(6):535 – 576, 2013.

[4] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.J. Martinez-de Pison, and F. Antonanzas-Torres. Review of photovoltaic power forecasting. Solar Energy, 136:78–111, October 2016.

[5] V Kostylev, A Pavlovski, et al. Solar power forecasting performance–towards industry standards. In 1st international workshop on the integration of solar power into power systems, Aarhus, Denmark, 2011.

[6] Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, and Rob J. Hyndman. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. International Journal of Forecasting, 32(3):896 – 913, 2016.

[7] Gordon Reikard. Predicting solar radiation at high resolutions: A comparison of time series forecasts. Solar Energy, 83(3):342 – 349, 2009.

[8] Peder Bacher, Henrik Madsen, and Henrik Aalborg Nielsen. Online short-term solar power forecasting. Solar Energy, 83(10):1772 – 1783, 2009.

[9] Hugo T.C. Pedro and Carlos F.M. Coimbra. Assessment of forecasting techniques for solar power production with no exogenous inputs. Solar Energy, 86(7):2017 – 2028, 2012.

[10] Federica Davò, Stefano Alessandrini, Simone Sperati, Luca Delle Monache, Davide Airoldi, and Maria T. Vespucci. Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting. Solar Energy, 134:327 – 338, 2016.

[11] Changsong Chen, Shanxu Duan, Tao Cai, and Bangyin Liu. Online 24-h solar power forecasting based on weather type classification using artificial neural network. Solar Energy, 85(11):2856 – 2870, 2011.

[12] Caroline Persson, Peder Bacher, Takahiro Shiga, and Henrik Madsen. Multi-site solar power forecasting using gradient boosted regression trees. Solar Energy, 150:423 – 436, 2017