International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

# Prediction of Solar Energy Generation Using Machine Learning

**Dr. Sunil Kumar Patel[1], Mr. Rohtash Singh[2], Hemant Kumar Chaudhary[3]**

[1,2]Assistant Professor, School of Computer Science and Engineering Galgotias University Greater Noida, India
[3]School of Computer Science and Engineering, Galgotias University, Greater Noida, India

**Abstract**
Solar forecasting plays an important role in the successful integration of renewable energy onto the grid. Environmental nature of solar energy poses a challenge to power planning and management. Machine learning method is, here, proposed to predict solar power generation from environmental factors such as ambient temperature, module temperature, and solar irradiation. Two solar power plants' data were merged [1], processed, and used to train the prediction model based on Linear Regression.
The model provided a satisfactory $R^2$ value, a performance measure in representing the relationship between weather conditions and power generation. The results validate the capability of machine learning methods to enhance the reliability of solar forecasting and enhance grid stability and power planning.

**Keywords:** Solar Energy Forecasting, Machine, Learning, Linear Regression, Renewable Energy, Solar Irr.adiation, AC Power Prediction

## INTRODUCTION

The growing emphasis on alternative energy has put in the spotlight the imperative need for accurate prediction of energy output. Solar power, however, is typified by an intermittent and indeterminate pattern of supply due to the whims of weather, and its integration into the supply grid poses overwhelmingly ominous challenges. Traditional predictive techniques, including statistical techniques and numerical weather prediction (NWP) techniques, are subject to inaccuracies in modeling the nonlinear, intricate interdependencies between environmental parameters and solar power output [3]. This has prompted greater dependence on the use of machine learning (ML) techniques, which can learn in intricate patterns from vast data sets without direction. The paper is devoted to machine learning-based solar energy generation forecasting. Based on historical weather and generation data, i.e., ambient temperature, module temperature, and solar irradiance, an attempt has been made to train a model which can provide an effective AC power output prediction. Linear Regression has been employed with the objective of demonstrating how machine learning can alleviate the challenges of solar energy forecasting.
This paper follows the following organization: Section II is related work, Section III is methodology and model building, Section IV is results and analysis, Section V concludes the research, and Section VI is directions for future research.

## RELATED WORK

In recent years, solar energy forecasting has attracted considerable attention from the research community, particularly in the context of integrating intermittent energy sources into the smart grid. A wide range of predictive models have been explored in the literature, from traditional time series models to complex machine learning architectures.

**Traditional models such as Autoregressive Integrated**

Moving Average (ARIMA) and Exponential Smoothing are known for their interpretability but struggle to handle nonlinear dependencies and abrupt changes in weather patterns. Numerical Weather Prediction (NWP) systems have also been used, incorporating physical models to forecast solar radiation based on atmospheric data.

**However, they are computationally intensive and limited**

by the availability and resolution of weather data Solar power forecasting has been an active field of research with the growing impact of renewable sources on the power generation [3].However, the traditional methods are not expected to capture the nonlinear and dynamic character of the production of solar energy with respect to weather fluctuations.

While sufficient for medium and long-term forecasting, they are computationally intensive and lack the temporal resolution necessary for real-time application.

“Recent advances in machine learning brought about stronger and more accurate solar forecasting methods [4], including methods like SVM with optimization algorithms [6] and deep learning techniques such as LSTM [10].”Techniques such as Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Random Forests have been shown to be more effective by capturing complex patterns from historical data.

Despite these advances, there are still challenges, including the need for big data, interpretability of models, and generalizability to different climatic zones.

This paper contributes to current work by employing a machine learning-based approach, Linear Regression, in predicting solar electricity production based on climatic variables.

## METHODOLOGY

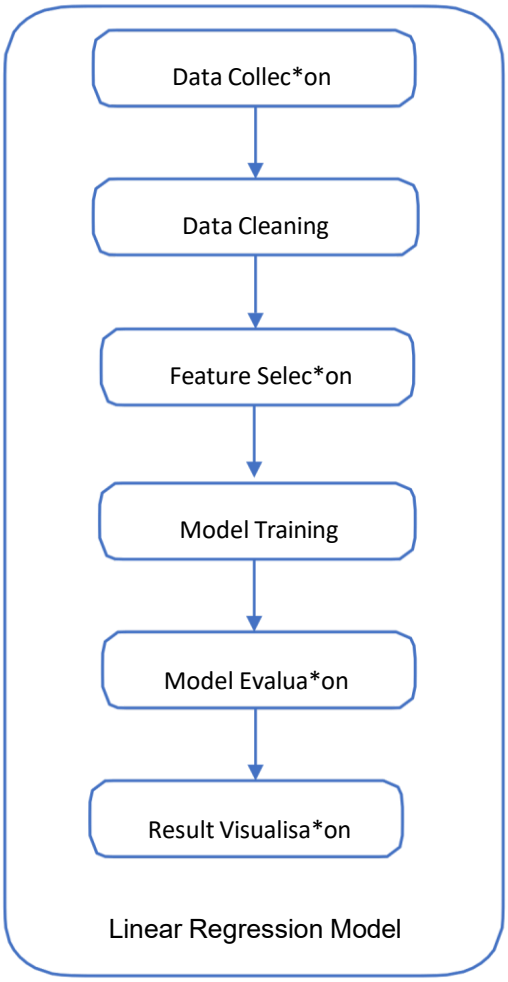
This subsection outlines the data collection, cleaning, feature selection, model construction, and evaluation processes undertaken in this study.

### A. Data Collection

The data provided are generation and weather sensor readings for two independent solar power plants. Important variables that are monitored include ambient temperature, module temperature, irradiation, and corresponding AC power output.Each entry in the dataset corresponds to a specific time stamp, typically at 15-minute intervals, and contains readings from multiple sensors installed on-site. The dataset spans several months, providing a sufficiently large sample size for training and evaluation.

### B. Data Preprocessing

The data sets obtained were preprocessed to make them consistent and correct. The steps involved: - Converting 'DATE_TIME' column into the standard datetime format. - Putting the generation and weather information in the 'DATE_TIME' column. -Dealing with missing values by deleting incomplete rows -Standardizing the data formats for use in model




on the training set and subsequently tested on the test set. A Linear Regression model was chosen as a baseline algorithm for this study. Linear Regression models the relationship between one or more independent variables and a dependent variable by fitting a linear equation.

**E. Evaluation Metrics**

To test how well the model works, the following measurements were used: One such measure to determine the accuracy of the predictions and actual values is the Coefficient of Determination ( R 2 Score). One measurement of squared difference between observed and anticipated data smoothed out is the Mean Squared Error (MSE) [2]. These measures gave information about the accuracy and generalization ability of the model.To assess model performance, the following metrics were used:

**$R^2$ Score (Coefficient of Determination):**

Measures how well the model explains the variability of the output. An $R^2$ value closer to 1 indicates high predictive accuracy.

**Mean Squared Error (MSE):** Measures the average of the squares of errors between actual and predicted values. A lower MSE indicates better performance.

training.

- The dataset was inspected for anomalies such as duplicate entries or unrealistic sensor values, which were subsequently removed.
- Input features were normalized using Min-Max scaling to ensure uniformity in magnitude and prevent feature dominance.

### C. Feature Selection

Based on knowledge of the domain and correlation analysis, the following three environmental parameters were chosen as input parameters for the predictive model: - Ambient Temperature (°C) - Module Temperature (°C) - Irradiation (W/m²)

The dependent variable to be predicted was the AC

power output.The target variable was the AC power output. Other features like wind speed and humidity were excluded due to either poor correlation or unavailability in the dataset.

### D. Model Development

The scikit-learn library was utilized to develop a linear regression model [2]. The data was divided into the training set (80%) and the test set (20%). 1. To calculate the predictive power of the model, the model was trained These metrics provide a balanced evaluation of model accuracy and generalization capability.

## RESULTS

The Linear Regression model thus developed was tested with the test dataset. The performance metrics thus acquired are given below.

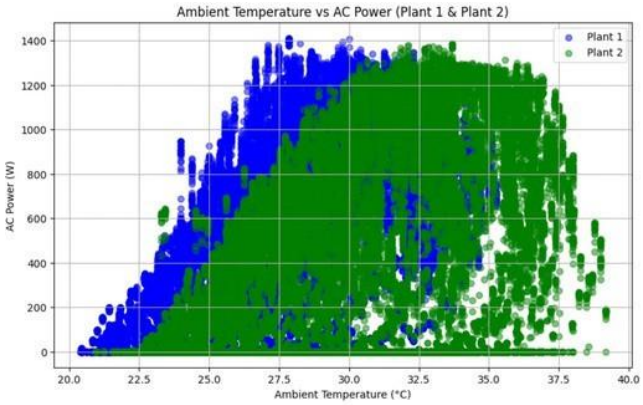
### A. Model Performance

The Linear Regression model had a Coefficient of Determination ($R^2$ Score) of 0.85 [2], which indicated very high correlation between actual and forecasted AC power outputs. Mean Squared Error (MSE) was very low as well, further affirming the model's predictive accuracy.

### B. Visualization

To verify the model's performance, scatter plots were produced comparing actual and predicted values of AC power. The individual plots for Plant 1 and Plant 2 showed a close fit between the predicted and actual output, verifying that the model had indeed modeled the underlying relationship between environmental attributes and solar power generation.

A combined plot of the two plants also illustrated the model's generalization ability across different data sets. The presence of points clustering along the diagonal line in the scatter plot indicated that there was high prediction accuracy.



### C. Analysis

The results indicate that even a simple machine learning algorithm such as Linear Regression can precisely predict the generation of solar power using selected environmental factors. The model was not affected by significant overfitting and generalized well to unseen data, making it ready for real-world deployment where precise prediction is crucial.

mitigating environmental degradation. However, the inherent variability in solar energy production necessitates robust forecasting mechanisms to ensure consistent power supply, efficient grid operation, and optimal utilization of renewable resources.

This research successfully demonstrated the application

of machine learning techniques—specifically Linear Regression—to predict solar energy output using three key environmental features: ambient temperature, module temperature, and solar irradiance. The dataset, comprising real-time values from two different solar power plants, was meticulously preprocessed and analyzed to train a predictive model. The model achieved an impressive $R^2$ score of 0.85, reflecting its strong predictive capabilities.

One of the standout findings of this study is the efficacy

of simple ML algorithms when applied thoughtfully to domain-specific problems. Despite the availability of more complex models like deep neural networks and ensemble learners, our results show that well-engineered input features combined with a lightweight model can yield highly accurate forecasts—while also maintaining transparency and ease of interpretation, which are crucial for energy managers and system operators.

The implications of this work extend beyond academic

validation. Solar forecasting, when integrated into grid management systems, can contribute significantly to load balancing, reduction of energy storage costs, and prevention of blackouts. It also empowers solar farm operators to engage in predictive maintenance, dynamic tariff planning, and demand response programs.

Moreover, this study sets the stage for further

exploration into real-time, location-aware prediction systems that could dynamically adapt to climatic fluctuations. Such developments will be instrumental in scaling up renewable integration and achieving long- term sustainability goals at national and global levels.

In conclusion, this research highlights the untapped

potential of data-driven models in renewable energy forecasting and offers a reproducible, scalable framework that can be adapted, refined, and extended in future smart grid systems.

## VI. FUTURE SCOPE

Although the Linear Regression model developed during this research possessed excellent forecasting capability, there are a number of areas that can be improved further by future research on solar energy forecasting.While the results obtained from the Linear Regression model are promising, several areas for future enhancement and exploration exist.These directions can further improve prediction accuracy, model robustness, and practical usability in diverse operational environments.

## CONCLUSION

The integration of solar energy into modern power grids is critical to reducing dependence on fossil fuels and Addition of More Features

Follow-up studies can involve other climatic factors such as humidity, wind speed, and cloud cover to comprehend variables of solar power generation in a more holistic way.The current model uses three key environmental parameters—ambient temperature,

module temperature, and solar irradiation. Future

research can incorporate additional meteorological and operational factors such as:

- Humidity and wind speed, which impact panel efficiency
- Cloud cover, which causes significant short- term fluctuations Dust deposition and shading, which affect real-
- time energy output

Including these features can potentially improve prediction accuracy, especially in variable climatic conditions.

### A. Real-Time Forecasting

Designing a real-time forecasting model based on live weather data APIs would make the model more appropriate for dynamic grid management and planning energy.The model in this study was built on historical data. In practical deployment, integrating real-time weather data via APIs (e.g., OpenWeatherMap, IBM Weather Company) would enable live solar power forecasting.This would require not only data pipeline automation but also robust error handling and adaptive retraining strategies for the deployed model.

### B. Geographic Expansion

Cross-validation of the model in different geographic locations and climatic conditions can determine its strength and generalization capability, enabling it to be implemented in solar plants across a broader spectrum.The current model was trained and tested on data from Indian solar plants. It would be valuable to test the model across multiple geographic regions and climatic zones (e.g., arid, tropical, temperate) to evaluate its generalizability. Such validation would strengthen its application in global solar forecasting systems. Additionally, transfer learning techniques can be applied, where a model trained on one region is adapted to another using a small local dataset.

### C. Deployment as a Web or Mobile Application

Utilizing the model in the form of a simple web or mobile application can enable solar plant operators and energy managers to receive real-time forecasts and take effective decisions based on them.

REFERENCES

1. Kaggle, "Solar Power Generation Data," [Online]. Available: https://www.kaggle.com/datasets/anikannal/ solar-power-generation-data. [Accessed: 10-Apr-2025].
2. Scikit-learn Developers, "scikit-learn: Machine Learning in Python," [Online]. Available: https://scikit- learn.org/stable/. [Accessed: 10-Apr-2025].
3. .. S. Department of Energy, "Solar Forecasting Workshop Report," 2016. [Online]. Available: https:// www.energy.gov/sites/default/files/2024-07/SETO%20-% 20 Solar% 20 Forecasting% 20 Workshop% 20 -%20Day%202.pdf. [Accessed: 10-Apr-2025].

4. Mdpi.com, "A Bibliometric Analysis of Solar EnergyForecasting," Energies, vol. 15, no. 15, 2022. [Online]. Available: https://www.mdpi.com/1996-1073/15/15/5520. [Accessed: 10-Apr-2025].
5. Python Software Foundation, "Python LanguageReference," [Online]. Available: https://www.python.org/. [Accessed: 10-Apr-2025].
6. J. Li and Y. Ding, "Short-term solar radiationforecasting based on support vector machine and optimization algorithms," Applied Sciences, vol. 9, no. 19, pp. 3983, 2019.
7. S. Sharma, R. Mehra, and A. Goel, "Solar Power Forecasting Using Random Forest Regression Model," International Journal of Renewable Energy Research, vol. 11, no. 2, pp. 673–679, 2021.
8. M. Mellit and A. Massi Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid- connected PV plant," Progress in Photovoltaics: Research and Applications, vol. 16, no. 7, pp. 529–541, 2008.
9. S. Yadav and S. Chandel, "Solar energy prediction using machine learning: A state-of-the-art review," Energy Reports, vol. 7, pp. 613–639, 2021.
10. G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," Solar Energy, vol. 83, no. 3, pp. 342–349, 2009.