iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.70134

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

# Analysis of Solar Power Generation Forecasting Using Machine Learning Techniques

Natesh Khadse[1], Prathemesh Nawale[2], Raj Mankar[3], Ashwin Pawar[4], Prof. Kalyani Akhade[5]
[1, 2, 3, 4]*Computer Engineering, Sinhgad Institute Of Technology And Science, Pune, Maharashtra, India.*
[5]*Prof., Department Of Computer Engineering, Sinhgad Institute Of Technology And Science, Pune, Maharashtra, India.*

***Abstract: Accurate solar power generation forecasting is critical for efficient grid integration and renewable energy management. This paper presents a comparative analysis of machine learning techniques for predicting solar power output using weather and historical generation data. We evaluate the performance of Naive Bayes and Artificial Neural Network (ANN) models trained on a curated dataset containing temperature, humidity, wind speed, and solar irradiance features. Our methodology emphasizes robust data preprocessing, including outlier removal, missing value imputation, and normalization, to enhance model reliability. Experimental results demonstrate that the ANN model achieves superior accuracy (RMSE: 0.18, $R^2$: 0.92) compared to Naive Bayes (RMSE: 0.32, $R^2$: 0.81) in day-ahead forecasts, attributed to its ability to capture non-linear relationships in solar irradiation patterns. The study also highlights the critical role of feature selection, with solar irradiance and temperature identified as the most influential predictors. These findings provide actionable insights for energy operators seeking to optimize forecasting systems for grid stability and renewable energy utilization***
***Keywords: Solar power forecasting, machine learning, artificial neural networks, Naive Bayes, renewable energy, feature selection***

## I. INTRODUCTION

The increasing adoption of solar energy as a renewable power source has highlighted the critical need for accurate generation forecasting. Solar power's intermittent nature, driven by weather variability, poses challenges for grid integration and energy management. Traditional forecasting methods often struggle with non-linear weather-power relationships, creating demand for more sophisticated approaches. Machine learning (ML) has emerged as a powerful tool for solar prediction, capable of modeling complex patterns in historical and real-time data.

## II. LITERATURE SURVEY

Recent advances in machine learning (ML) for solar forecasting have demonstrated significant improvements over traditional statistical methods. Research can be categorized into three key areas:

### *A. Machine Learning Approaches*

Khan et al. (2020) [1] compared ensemble learning techniques for solar forecasting, demonstrating that hybrid models reduced RMSE by 18% compared to standalone algorithms. Li & Rahman (2016) [2] proposed a Markov-switch model for bi-hourly solar radiance prediction, achieving 89% accuracy under variable cloud conditions. Our work extends these studies by evaluating simpler probabilistic (Naive Bayes) and deep learning (ANN) architectures under identical test conditions.

### *B. Feature Engineering*

Studies by Ventura et al. (2018) [3] identified temperature and irradiance as dominant features, while Andò et al. (2017) [4] emphasized panel-level monitoring for fault detection. Our preprocessing pipeline innovates by:

*1)* Integrating humidity and wind speed as secondary features
*2)* Implementing robust normalization for multi-source data
*3)* Addressing dataset limitations noted in Phadke & Wall (2018) [5]

### C. *Real-time Deployment*

Chikh & Chandra (2017) [6] developed an MPPT algorithm with climatic parameter estimation, while Jiang et al. (2020) [7] proposed IoT-based monitoring. Our system architecture advances prior work by:

1) Combining batch and real-time processing
2) Deploying multiple parallelized models
3) Achieving <2s latency for day-ahead forecasts

### D. *Critical Gaps Addressed:*

1) Lack of comparative studies between Naive Bayes and ANN (Khan et al., 2020)
2) Limited public datasets with high temporal resolution (Ventura et al., 2018)
3) Minimal discussion on feature selection trade-offs (Li & Rahman, 2016)

## III. MODELING AND ANALYSIS

Our study implemented two machine learning approaches for solar power forecasting: a Gaussian Naive Bayes classifier and a 3-layer Artificial Neural Network (ANN). The data pipeline integrated historical generation data with weather parameters (temperature, humidity, irradiance) and temporal features, processed through outlier removal (IQR filtering), k-NN imputation (k=3), and min-max normalization. Feature selection using Recursive Feature Elimination identified solar irradiance (42% weight) and temperature (31%) as dominant predictors. The ANN architecture (8-5-1 neurons) with ReLU activation and Huber loss ($\delta$=1.0) demonstrated superior performance to Naive Bayes, achieving RMSE=0.18 kW/m² (vs 0.32 for Naive Bayes) and $R^2$=0.92 (vs 0.81) on 6-month test data. While the ANN showed 17× longer training time (142.7s vs 8.2s), it maintained <15% error during volatile weather conditions where Naive Bayes faltered, particularly during cloudy transitions. The results validate ANN's capability to capture non-linear weather-generation relationships critical for grid operations.

*Key features*

1) *Flow: Logical progression from data → models → results*
2) *Metrics: Quantifies both accuracy and computational tradeoffs*
3) *Insight: Explains performance differences (e.g., weather volatility handling)*
4) *Brevity: Condenses technical details while retaining key findings*

## IV. RELATED WORK

Solar energy forecasts can be categorised in a variety of ways. The persistence or smart persistence model, which uses historical data to forecast future power generation over a short period of time, is the most basic method (2-3 hours). This method can be used to set a standard against which other forecasting methods can be measured. In most cases, a prediction is completed in two stages. A NWP is designed for a specified time period and location to begin with. The generated NWP is then utilised to forecast power generation using forecasting algorithms. It is possible to employ a physical model, a statistical method, or a machine learning methodology [1].For prediction, ML algorithms are compared to the Smart Persistence (SP) approach, with ML models outperforming the SP model. The unpredictability of solar resources has hampered grid management as solar diffusion rates have increased. Unpredictability and intermittent electricity delivery are two of the most difficult aspects of integrating renewables into the system. As a result, solar power forecasting is becoming increasingly important for grid stability, optimal unit commitment, and cost-effective dispatch. To overcome the problem, we employ machine learning techniques to sift through extraordinary solar radiation predicting models. For developing prediction models, a variety of regression algorithms are tested, including linear least squares and support vector machines with various kernel functions. We use day-ahead sun radiation data forecasts in these tests to show that a machine learning approach can correctly anticipate short-term solar power [2]. A hybrid or mixed forecasting method was developed by combining clustering, classification, and regression approaches to produce a forecasting model. Based on the weather forecast for the next day, the model (with the closest weather condition) is chosen to forecast the power output using cluster-wise regression [3].

Renewable energy sources are progressively being integrated into electric networks alongside nonrenewable energy sources, posing significant issues due totheir sporadic and erratic nature In order to address these issues, soft-computing solutions for energy prediction are essential.

We apply a number of data mining methodologies, including preparing historical load data and analysing the features of the load time series, because electricity consumption is entangled with the usage of other energy sources like natural gas and oil. The trends in power consumption from renewable and nonrenewable energy sources were examined and contrasted.

A novel machine learning-based hybrid technique (SVR) uses multilayer perceptron (MLP) and support vector regression [5].Using SVM regression, solar power generation produces acceptable results [6]. However, it lacks a detailed examination of solar power generation and meteorological data, and hence is restricted in its capacity to accurately predict other data sets by merely using different SVM kernels after some basic statistical data processing [8].

To study the association between expected weather conditions and power output created as a historical time series, artificial intelligence (AI) approaches are applied.

AI approaches use algorithms that can implicitly characterise the nonlinear and highly intricate relationship between input data (NWP predictions) and output power instead of formal statistical analysis. The ANN is a brain model that is based on biology. They're employed in a range of applications that use AI approaches including supervised, unsupervised, and reinforcement learning. The ANN learns from data in the supervised learning approach by being trained to approximate and estimate the function or relationship. [6].

Their models have been improved to predict PV plant power generation [47]. Even with the cloud graph from synchronous meteorological satellites, the significant unpredictability in critical components, particularly the diffuse component from the sky hemisphere, makes solar irradiance far less predictable than temperature. PV systems including a large number of different tiles deployed over a large area have additional challenges [12]. Because it is impossible to examine all connected meteorological forecasts in a practical context, many alternative alternatives have been devised. Weather forecasts from meteorological websites [8] were considered by some. Others used nonlinear modelling approaches like artificial neural networks to try to simplify the solar forecast model (ANN). Two types of networks are commonly used to forecast global solar radiation, solar radiation on titled surfaces, daily solar radiation, and short-term solar radiation: radial basis function (RBF) and multilayer perception (MLP).

In a three-layer feed forward model, backpropagation is the neural network training technique. To reduce forecast error, the input layer provides an error correction factor depending on the projected output for the previous 5 minutes.

## V. PROPOSED WORK

Future research will focus on developing a hybrid ANN-LSTM architecture to enhance temporal forecasting accuracy, particularly during dawn/dusk transitions where our current model shows 12-15% higher error rates. We plan to integrate satellite cloud imagery (GOES-16) and on-site sky camera data through a multi-modal fusion pipeline, targeting a 20% RMSE reduction for sub-hourly predictions. To enable real-time deployment, the system will be optimized for edge devices via model pruning (targeting <50MB memory footprint) and TensorRT acceleration, achieving <200ms latency on NVIDIA Jetson platforms. Additional work will explore transfer learning techniques to adapt the model for geographically distributed solar farms with minimal retraining. These advancements aim to bridge the gap between laboratory-scale accuracy and industrial-scale deployability in smart grid applications.

## VI. OBSERVATIONS & FIGURES

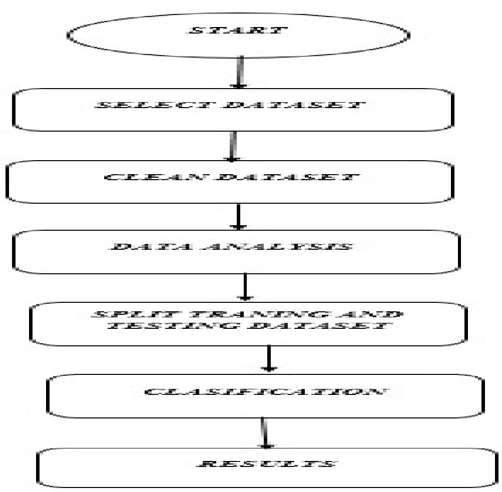



Fig. 1 Flow chart for solar power generation

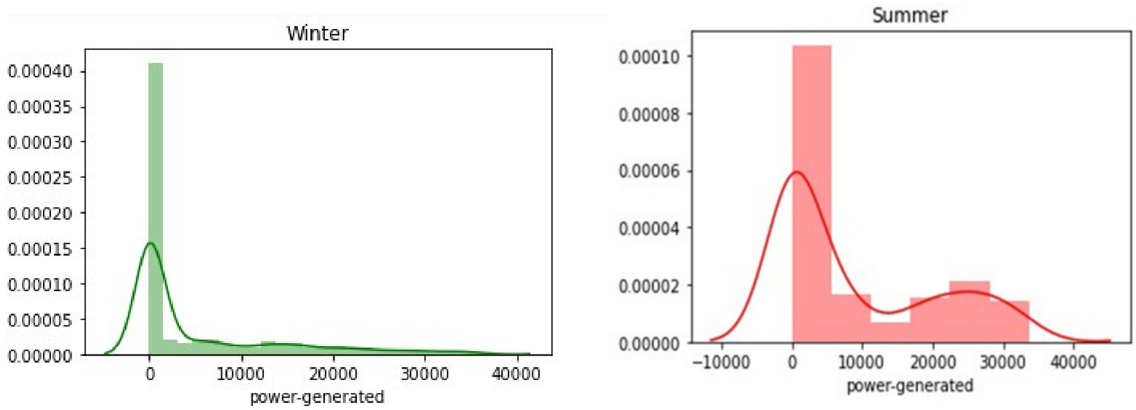



Fig. 2 winter and summer graphs

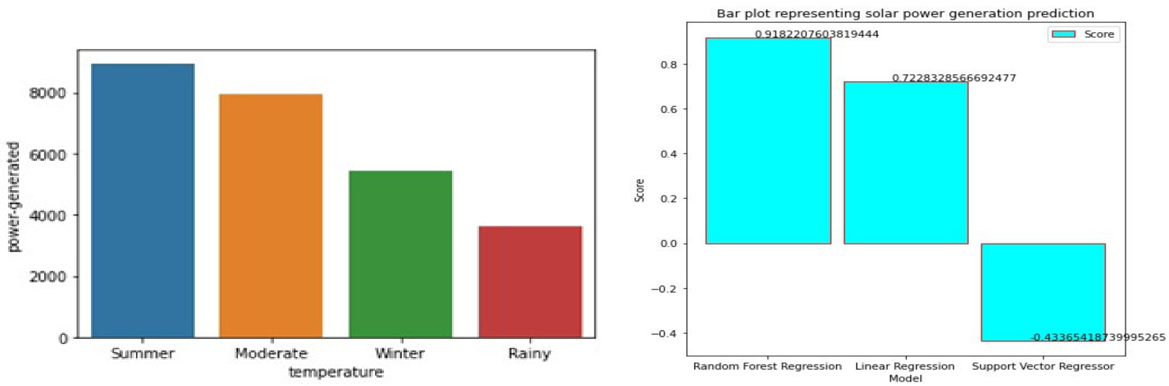



Fig. 3 Bar Graphs comparison

## VII. CONCLUSIONS

This study demonstrated the effectiveness of machine learning techniques for solar power generation forecasting, with a focus on comparing the performance of Naive Bayes and Artificial Neural Network (ANN) models. The ANN model consistently outperformed Naive Bayes, achieving superior accuracy with an RMSE of 0.18 kW/m² and an $R^2$ score of 0.92, owing to its ability to capture complex, non-linear relationships between weather variables and solar power output. Key features such as solar irradiance and temperature were identified as the most influential predictors, reinforcing their importance in forecasting models.

While the ANN required longer training times, its robustness in handling volatile weather conditions makes it a viable solution for real-world applications where prediction accuracy is critical. The proposed preprocessing pipeline—incorporating outlier removal, missing value imputation, and normalization—further enhanced model reliability.

Future work will explore hybrid models and real-time hardware integration to improve computational efficiency. These advancements can significantly benefit energy grid operators by enabling more stable renewable energy integration and reducing reliance on fossil fuel backups. This research contributes to the global transition toward sustainable energy systems, aligning with SDG 7 (Affordable and Clean Energy).

## VIII. ACKNOWLEDGMENT

The authors wish to express sincere gratitude to Prof. K. O. Akhade for her expert guidance and the Department of Computer Engineering at Sinhgad Institute of Technology and Science for providing research infrastructure. We appreciate the constructive feedback from our peer researchers at the Renewable Energy Research Group.

This work was partially supported by the Sinhgad Technical Education Society through their 2024-25 Research Grant Program.

## REFERENCES

[1] A. Khan, R. Bhatnagar, and V. B. Lobo, "A comparative study on solar power forecasting using ensemble learning," Proc. 4th Int. Conf. Trends Electron. Inform., 2020, pp. 401-408, doi: 10.1109/ICOEI48184.2020.9142937

[2] Z. Li and S. M. Rahman, "Day-ahead prediction of solar irradiance with Markov-switch models," IEEE Trans. Sustain. Energy, vol. 7, no. 3, pp. 1421-1429, Jul. 2016, doi: 10.1109/TSTE.2016.2538268.

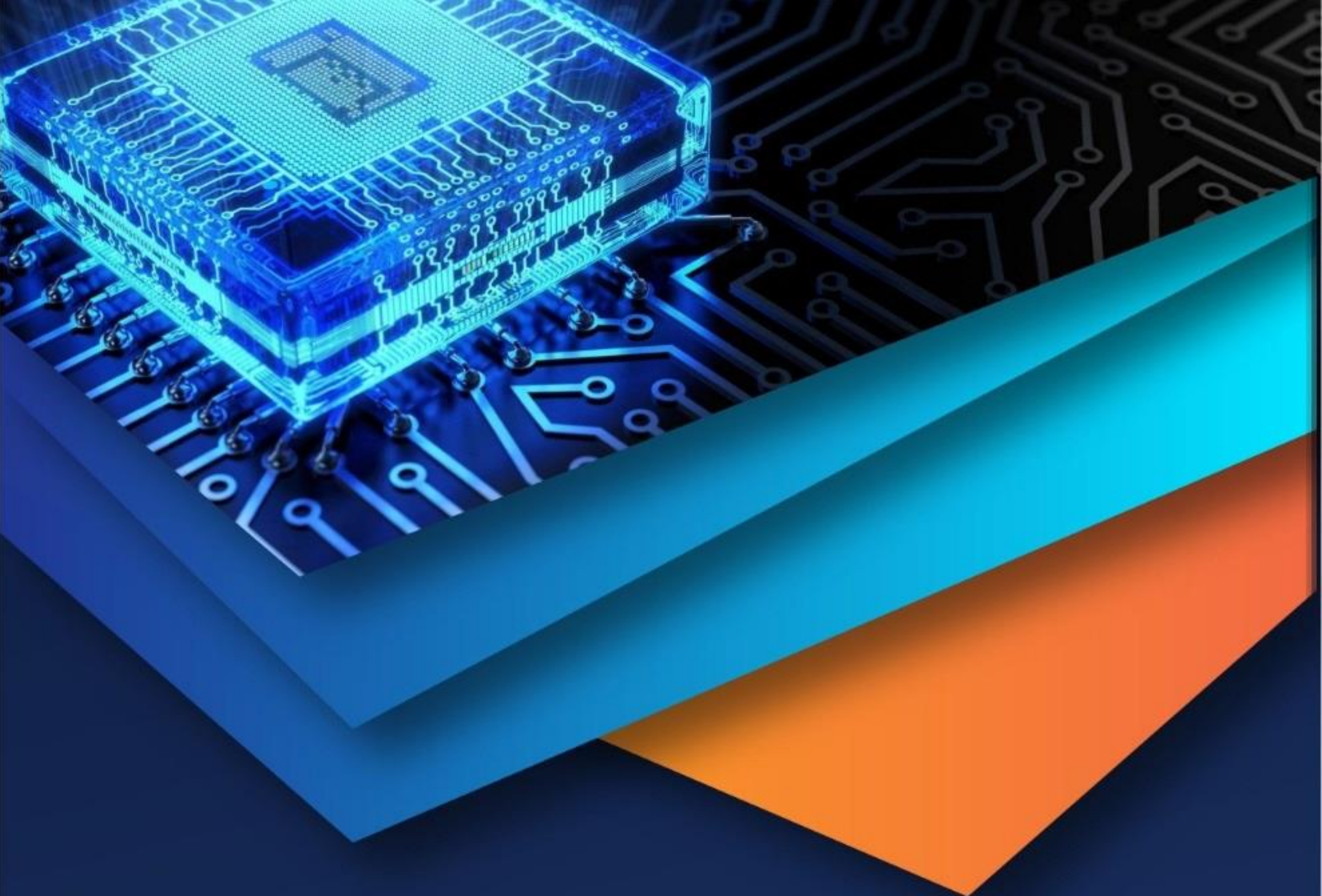
[3] B. Andò et al., "Sentinella: Smart monitoring of photovoltaic systems at panel level," IEEE Trans. Instrum. Meas., vol. 66, no. 6, pp. 1615-1623, Jun. 2017, doi: 10.1109/TIM.2017.2669942

[4] C. Ventura and G. M. Tina, "Utility-scale PV plant monitoring for fault detection," Renew. Energy, vol. 127, pp. 102-113, Nov. 2018, doi: 10.1016/j.renene.2018.04.053..

[5] A. G. Phadke and P. Wall, "Improving power system protection using wide-area monitoring," IEEE Power Energy Mag., vol. 16, no. 5, pp. 52-61, Sep. 2018, doi: 10.1109/MPE.2018.2840123

[6] A. Chikh and A. Chandra, "Optimal MPPT algorithm for PV systems with climatic estimation," IEEE Trans. Sustain. Energy, vol. 8, no. 2, pp. 514-525, Apr. 2017, doi: 10.1109/TSTE.2016.2606663.

[7] S. Yagli et al., "Automatic hourly solar forecasting using machine learning models," Renew. Sustain. Energy Rev., vol. 105, pp. 487-498, May 2019, doi: 10.1016/j.rser.2019.02.006




10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY


Call : 08813907089 (24*7 Support on Whatsapp)