Intelligent Solar Energy Forecasting From Power Generation Analytics to Renewable Insights

Manas Selukar, Anant Singh, Parv Goyal, Vaibhav Rajput
Department of Computer Science
Roll Nos: 2401010259, 2401010067, 2401010322, 2401010487

***Abstract*—This project presents a machine learning-based approach for forecasting solar AC power generation using historical plant generation data and weather parameters. The objective is to develop a regression model capable of capturing solar variability and temporal patterns. A Random Forest Regressor was implemented with engineered temporal and lag features. Model performance was evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Results demonstrate strong predictive capability and alignment with real-world photovoltaic behavior.**

I. Introduction

The increasing integration of renewable energy sources into modern power grids requires accurate forecasting systems. Solar energy generation is highly dependent on environmental and temporal factors such as irradiation, temperature, and time of day. Accurate prediction helps optimize grid stability and energy planning.

This project focuses on developing a traditional machine learning model (without generative or agentic AI methods) to forecast short-term solar AC power output using historical data.

II. Dataset Description

Two datasets were used:

- **Generation Data:** AC and DC power output, daily yield, and total yield.
- **Weather Data:** Ambient temperature, module temperature, and irradiation.

The datasets were merged using the `DATE_TIME` column to align generation with environmental conditions.

III. Preprocessing

Preprocessing steps included:

- Forward filling missing values
- Chronological sorting
- Feature scaling using StandardScaler
- Removal of null values after lag feature creation

Time-series integrity was maintained by disabling shuffling during the train-test split.

IV. Feature Engineering

To capture temporal dynamics:

- Hour and month features extracted from timestamps
- Autoregressive lag features (previous 1, 2, and 3 AC power values)

These features allow the model to learn daily seasonality and short-term dependencies.

V. Model Selection

A Random Forest Regressor was selected due to:

- Ability to model non-linear relationships
- Robustness to noise
- Strong performance on structured tabular datasets

Model parameters:

- 200 decision trees
- Maximum depth of 15
- Random state = 42

VI. Evaluation Metrics

Performance was evaluated using:

$$MAE = \frac{1}{n}\sum |y_i - \hat{y}_i| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum (y_i - \hat{y}_i)^2} \tag{2}$$

VII. Results

The model achieved:

- MAE $\approx$ 15.36
- RMSE $\approx$ 52.52

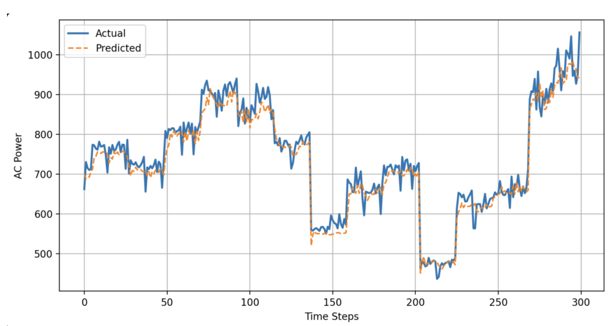

Fig. 1. Actual vs Predicted Solar Power Output

VIII. Feature Importance

Feature importance analysis showed irradiation as the dominant predictor, aligning with photovoltaic principles.

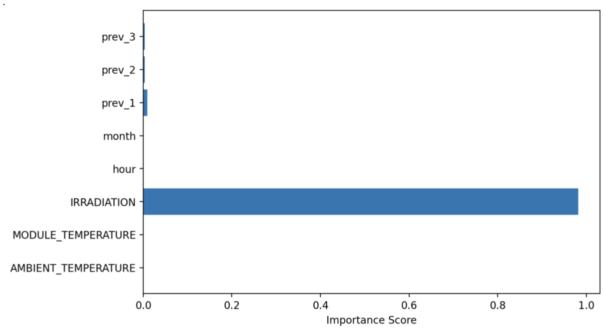

Fig. 2. Feature Importance Scores

IX. Seasonality Analysis

Average power generation peaked during midday hours, confirming expected solar behavior.

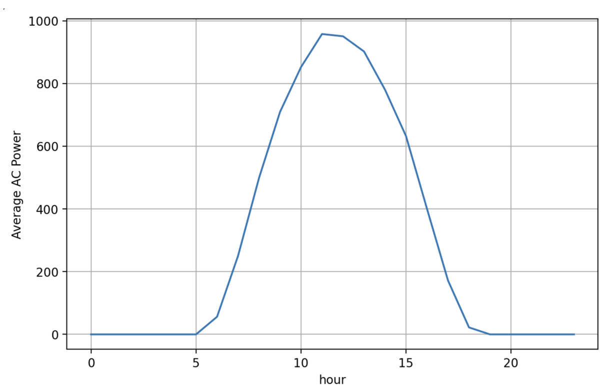

Fig. 3. Average Power Generation by Hour

X. Discussion

The results demonstrate that traditional machine learning methods can effectively forecast solar energy output. Lag features significantly improved predictive stability. While Random Forest performed strongly, future work may explore LSTM-based deep learning models for enhanced temporal modeling.

XI. Conclusion

This project successfully developed and deployed a solar power forecasting system using traditional regression techniques. The system achieved strong performance and demonstrated applicability for renewable grid integration scenarios.

Deployment

https://intelligent-solar-energy-forecasting-y6fpajnbvsdecdhgbhuvkf.streamlit.app/