

# Solar PV Plant Energy Loss Attribution Report

Dashboard Link: <https://alamba570-pv-test-kuwrmj.streamlit.app/>

## 1. Objective

Quantify and explain the gap between **theoretical solar energy generation** and **actual energy output** using sensor data, and attribute losses to specific causes:

- Cloud Cover
- Shading
- Temperature Effects
- Soiling
- Other/Unexplained Factors

## 2. Data & Preprocessing

- **Data Source:**  
15-minute interval SCADA/monitoring CSV from a solar PV plant. Includes actual generation, GHI, temperature, tracker angles, inverter data, etc.
- **Preprocessing Steps:**
  - Missing values imputed using **median**.
  - Outliers removed using the **IQR method**.
  - Key numeric features normalized using **MinMaxScaler**.

## 3. Theoretical Generation Model

- **Approach:**  
Theoretical generation is computed using:
  - GHI (Global Horizontal Irradiance)
  - Panel area = 10 m<sup>2</sup>
  - Panel efficiency = 15%
  - **Temperature correction** using module temperature with coefficient -0.004/°C
- **Formula Used:**
  - Theoretical Generation=GHI×10×(0.15+(Module Temp−25)×−0.004)

## 4. Loss Attribution Methodology

Loss Type	Flag Condition	Loss Calculation
Cloud Cover	GHI < 25th percentile	Theoretical - Actual (during flag)
Temperature	Ambient Temp > 75th percentile	Theoretical - Actual (during flag)
Soiling	Rolling mean of (Theoretical - Actual) > 0.05	Theoretical - Actual (during flag)
Shading	Sharp dip: diff of loss > 0.1	Theoretical - Actual (during flag)
Other Losses	Residual loss after above attributions	Same

## 5. Modeling & Experiments

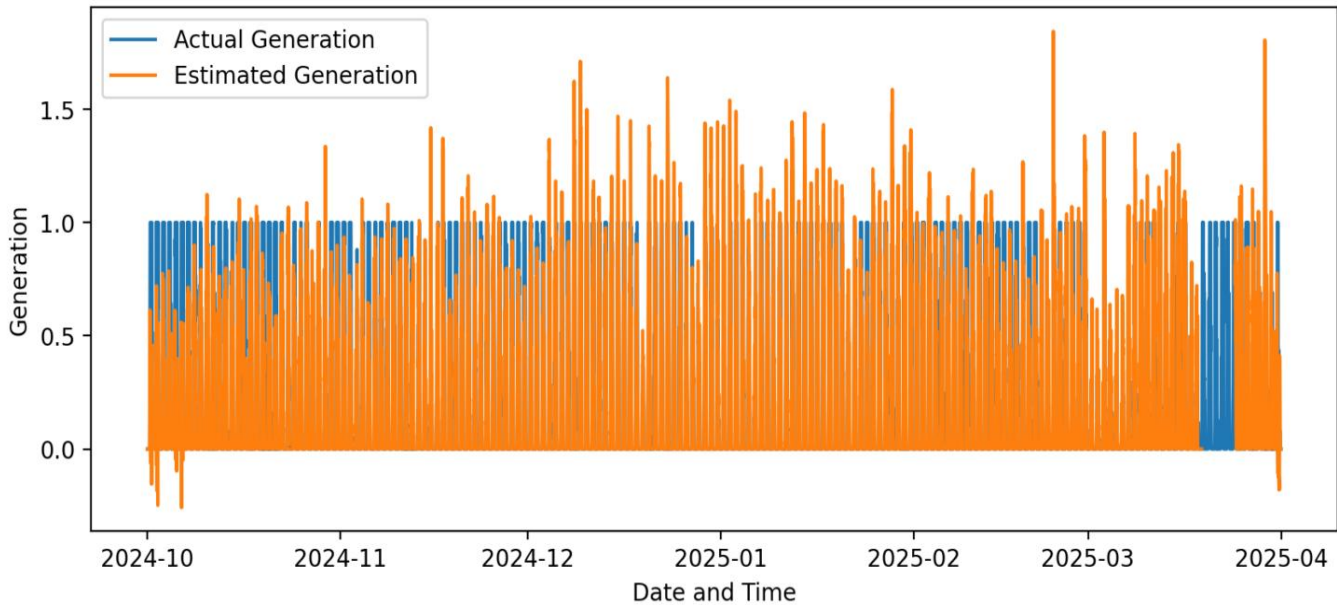
- **Models Used:**
  - Linear Regression (with imputed features)
  - HistGradientBoostingRegressor (robust to NaNs)
- **Performance Metrics:**

Model	MAE	RMSE
Linear Regression	0.0763	0.1725
Gradient Boosting Regressor	0.0391	0.1059

- **Top Features by Importance:**  
GHI, module temperature, hour of day, and day of week.

6. Results & Visualizations

- Actual vs. Theoretical Generation Plot



- Loss Breakdown Charts

Aggregated over:

- 15-min
- Hourly
- Daily
- Weekly
- Monthly

*Insert screenshot here*

- Downloadable CSVs:

- Loss flags
- Quantified losses
- Multiple time resolutions (15-min to monthly)

7. Key Insights

- Cloudy days, high temperatures, and soiling correlate strongly with major losses.
- ML models showed high accuracy, with RMSE within acceptable thresholds.
- Multiple concurrent loss types observed — handled using layered flagging logic.
- Visualization shows strong seasonal and hourly patterns in loss distributions.

8. Assumptions & Limitations

- Percentile thresholds (25th/75th) are **data-driven**; can be tuned.
- Soiling/shading detection based on statistical heuristics due to lack of direct sensors.
- Inverter/String-level breakdown **only possible** if data is present.

9. Appendix

- Dashboard:

Interactive dashboard with:

- Visual loss breakdowns
- Actual vs theoretical graphs
- CSV export
- Project summary