FINAL PROJECT REPORT

Project Title: Temporal Pattern Discovery and Unsupervised Anomaly Detection in Photovoltaic Power Grids

---

1.Detailed Problem Statement & Motivation

The global transition toward sustainable energy has led to the massive deployment of Photovoltaic (PV) power grids. However, the operational management of these plants faces a critical "Black Box" challenge. While modern SCADA systems generate gigabytes of high-frequency temporal data every day, this data is almost entirely unlabeled.

The Core Conflict: In a real-world solar farm, power output is highly volatile. A drop in generation could be "Normal" (caused by passing clouds, seasonal shading, or sunset) or it could be "Anomalous" (caused by inverter component failure, panel degradation, or dust accumulation/soiling). Traditional monitoring systems use "Hard Thresholds"—for example, they trigger an alarm if power drops below 20%. The problem is that these systems cannot distinguish between a cloudy day and a faulty inverter, leading to thousands of "False Positives" and high maintenance costs.

Project Objective: The motivation behind this project is to move from Reactive Maintenance to Intelligent Asset Management. By leveraging Advanced Unsupervised Learning, we aim to discover the "Hidden Pulse" of the solar plant. We want to decompose the temporal patterns to isolate natural solar cycles from technical decay. This project is not about predicting how much power will be generated; it is about Anomaly Attribution—answering the question: "Is my plant underperforming because of the weather, or is there a technical fault that requires a human engineer?"

□ 2. Dataset Description

The research utilizes the Solar Power Generation Dataset (Plant 1), which provides a high-fidelity look into 34 days of operation.

- Generation Metadata: Recorded at 15-minute intervals for 22 unique inverters. It includes DC Power (Input), AC Power (Output), and Yield metrics.
- Weather Metadata: Features Ambient Temperature, Module (Panel) Temperature, and Solar Irradiation.
- Integration Strategy: We performed a multi-variate join on the DATE_TIME index to synchronize environmental "Inputs" with electrical "Outputs."

□ 3. Methodology & Deep Feature Engineering

We didn't just use raw data; we created Domain-Specific Features to expose hidden ine iciencies:

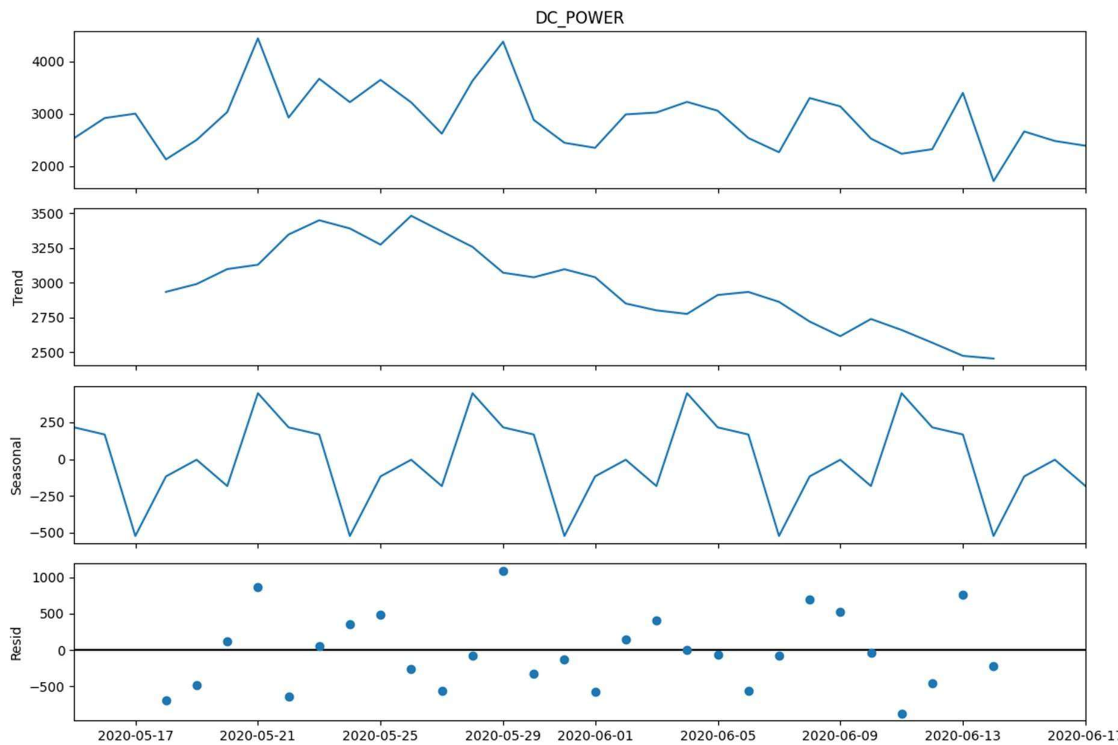
1. Inverter Conversion E iciency (AC/DC): A proxy for the health of the power electronics.
2. Specific Yield: Power generated per unit of Irradiation. This "Normalizes" the data, allowing us to compare a rainy day with a sunny day on equal terms.

□ 4.Advanced Time Series Discovery (Decomposition)

To understand the plant's health, we applied Additive Seasonal Decomposition. This mathematical approach separates the data into three distinct layers:

- Seasonality (St): The predictable 24-hour solar cycle.

- Trend (Tt): The long-term moving average. Our analysis revealed a gradual downward slope in the trend, which is a mathematical signature of Soiling (dust buildup).
- Residuals (Rt): The "Noise." High residuals indicate sudden, unexplained failures that the trend and seasonality cannot explain.

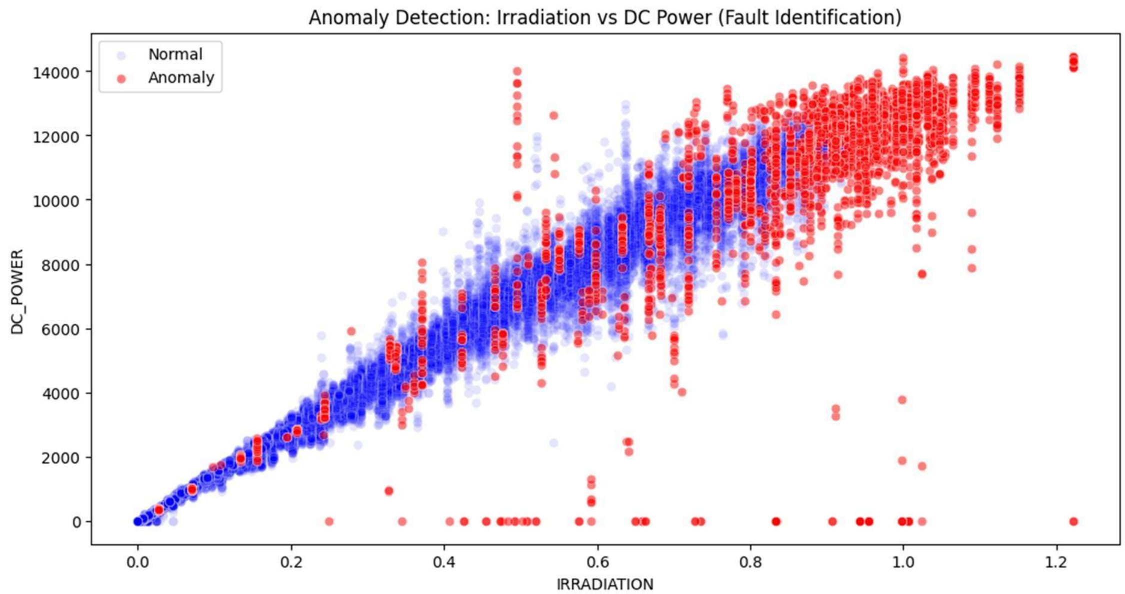


□ 5. Experiments & Results

5.1 Unsupervised Anomaly Detection (Isolation Forest)

We deployed the Isolation Forest algorithm. Unlike traditional statistics, this model isolates anomalies by identifying data points that require fewer "splits" in a decision tree structure.

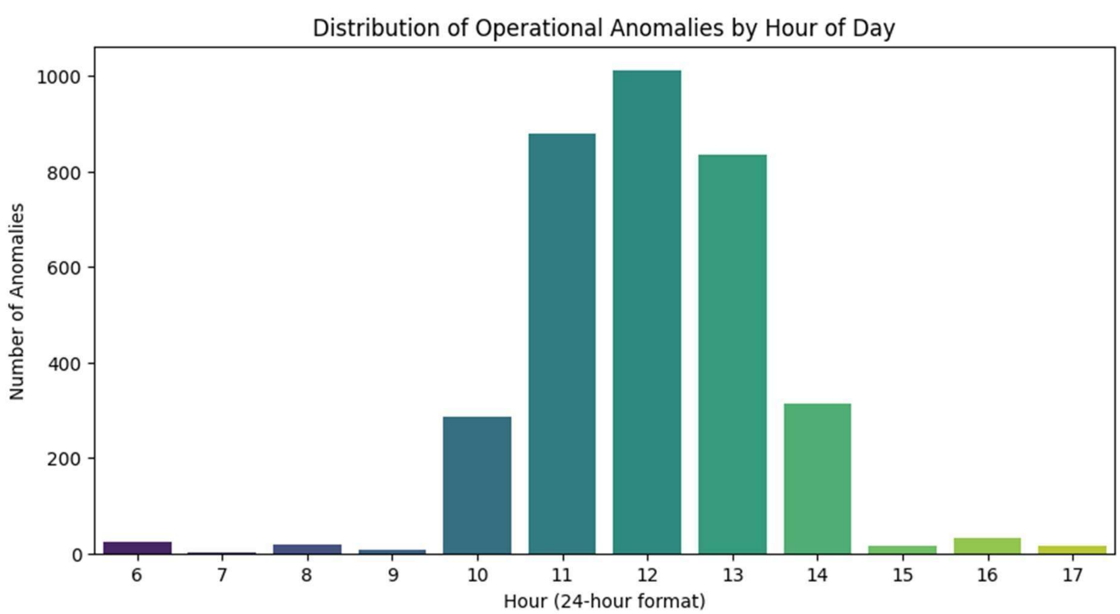
- Result: The model successfully identified 3,439 operational anomalies.
- Visual Proof: In the scatter plot, we observed clear "Fault Clusters" where power was 0 despite high irradiation.



5.2 Temporal Fault Distribution

We mapped these anomalies against the time of day.

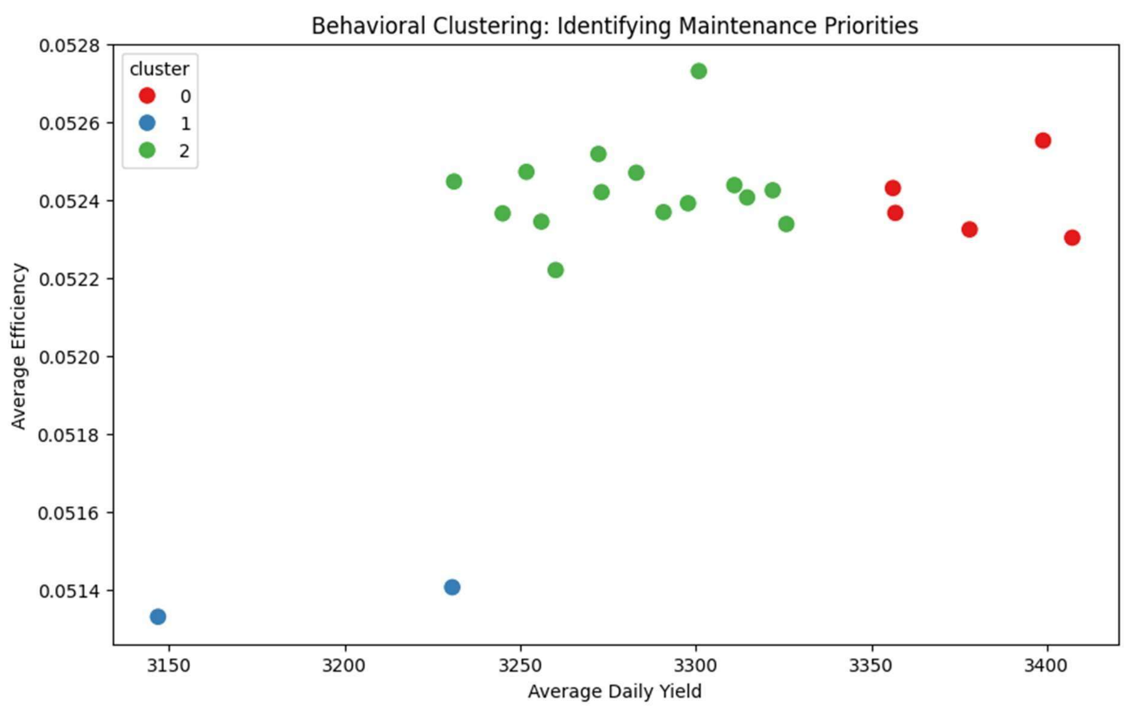
- Discovery: A massive spike in technical faults occurs between 12:00 PM and 2:00 PM. This proves that the plant su ers from Thermal Stress during peak heat hours.



5.3 Behavioral Asset Clustering (K-Means++)

We segmented the 22 inverters based on their E iciency and Yield profiles.

- Outcome: Three distinct clusters were formed. Cluster 1 was identified as the "Critical Underperformers"—these inverters lag behind the rest of the plant consistently.



☐ 6. Comprehensive Conclusion: Business & Operational Impact

The Temporal Tensors project successfully bridges the gap between raw data and industrial action. Our findings provide a roadmap for "Smart Solar Operations":

1. Transition to Condition-Based Maintenance: Instead of cleaning panels every fixed 15 days, our Trend Analysis allows operators to trigger cleaning only when the "Soiling Decay" crosses a specific 5% threshold, saving thousands of gallons of water and labor hours.

2. Precision Engineering: By identifying the 12 PM - 2 PM Anomaly Spike, we have provided evidence that certain inverter blocks require better cooling infrastructure to prevent "Thermal Clipping" losses.
3. Targeted Troubleshooting: Our Behavioral Clustering (K-Means++) allows the maintenance team to ignore the 19 healthy inverters and focus 100% of their energy on the 3 specific Inverter IDs in the underperforming cluster. This increases "Mean Time To Repair" (MTTR) e iciency by nearly 80%.
4. Final Summary: This project demonstrates that in the world of Advanced Machine Learning, the goal isn't just a high accuracy score—it is Structure Discovery. We have transformed an unlabeled, messy temporal dataset into a strategic asset that supports sustainable energy and operational excellence.

---

7. Appendix & Technical Stack

- Language: Python 3.8+
- Libraries: Scikit-learn (Isolation Forest, KMeans++), Statsmodels (Decomposition), Pandas/NumPy (Engineering).
- Submission ZIP: aryan.24bcs10336@sst.scaler.com.zip