# Analysis and Application of Power System Data in Electrical Engineering

## Introduction

The integration of data-driven analysis in power systems has become crucial for optimizing energy management, fault detection, and system reliability. This project focuses on analyzing a large-scale solar power dataset, which includes inverter outputs, environmental conditions, and power meter readings. By validating and analyzing this dataset, we aim to derive meaningful insights that enhance power system efficiency and sustainability.

## Dataset Overview

This dataset serves as a valuable resource for understanding the performance, efficiency, and reliability of solar power systems. Spanning a four-year period (2015–2018), it contains 21,03,810 records with 117 different parameters that cover key aspects of power generation, fault detection, and environmental conditions. By analyzing this data, we can uncover critical insights into how solar energy systems respond to changing environmental factors and identify potential inefficiencies or failures.

One of the major strengths of this dataset is its high temporal resolution, with records being logged every 60 seconds. This level of granularity allows for precise tracking of power fluctuations, system errors, and external influences such as temperature, wind speed, and solar radiation levels. Additionally, the dataset underwent rigorous validation, confirming its accuracy, completeness, and reliability.

Key Parameters in the Dataset:

• Power System Parameters: Inverter voltage, current, frequency, and power output.

• Fault Detection Metrics: Inverter fault logs, system warnings, and grid-related errors.

• Environmental Data: Solar radiation, ambient temperature, and wind speed.

• Geographical Information: Latitude, longitude, and timestamps.

• Energy Storage: Battery voltage and charge state.

The primary objectives of analyzing this dataset include:  
• Understanding power output trends based on inverter performance and weather conditions.  
• Identifying fault patterns in the system to enhance predictive maintenance.  
• Evaluating the impact of environmental factors like solar radiation and wind speed on energy efficiency.  
• Enhancing battery storage efficiency by monitoring charge and discharge cycles.  
• Developing data-driven strategies to optimize renewable energy systems for future smart grids.

This dataset is crucial for researchers, engineers, and energy analysts who aim to improve the sustainability, efficiency, and reliability of solar power systems. The insights drawn from this study can contribute to better decision-making in the design, operation, and maintenance of renewable energy infrastructures.

## Validation and Data Quality Analysis

### 1. Temporal Validation

• The dataset has an estimated resolution of 60 seconds per record.

• No duplicate timestamps were found.

• 9 large time gaps were detected, which may indicate missing data during certain periods.

### 2. Spatial Validation

• The dataset contains 2 unique coordinate points, with latitude fixed at 39.25 and longitude at 77.15m.

• These values suggest that all data was collected from a single location or a small number of sites.

### 3. Physical Validation

• Identified Photovoltaic (PV) system parameters, including inverter power output and DC input values.

• Average night power was calculated as 51.65 kW, suggesting possible measurement errors or data artifacts.

• Average night radiation was recorded at 2.81 W/m², which is within expected limits.

### 4. Missing Data Analysis

The dataset contains missing values in several key columns:

• Pyra1\_Wm2\_Avg, Pyra2\_Wm2\_Avg, and RECORD have approximately 0.46% missing data.

• Inverter current and voltage values have 6.10% missing data, potentially affecting power analysis.

• Wind speed and direction data have missing values up to 3.13%.

• Battery and charging parameters have minor missing values (<0.5%).

### 5. Recommendations

• Investigate 9 large time gaps to determine if data recovery is possible.

• Address missing values in inverter readings, as they affect power output analysis.

• Validate night-time power readings to confirm system behavior during non-operational hours.

## Methodology

After validating the dataset, we applied the following preprocessing techniques to ensure accuracy:

### 1. Data Cleaning

• Removed erroneous values (-999, -7999) that indicate sensor errors.

• Replaced missing data in critical columns using interpolation where possible.

### 2. Normalization and Standardization

• Standardized power output, frequency, and wind speed values to ensure consistency.

• Normalized environmental factors for better comparison across different conditions.

### 3. Feature Engineering

• Derived efficiency ratios by comparing inverter input and output power.

• Identified correlation between environmental conditions and power generation efficiency.

## Analysis and Key Findings

### 1. Power Output Trends

• The dataset showed fluctuations in inverter power output, influenced by solar radiation and temperature changes.

• Extreme temperatures led to reduced efficiency, highlighting the impact of thermal effects on PV performance.

### 2. Fault Detection and System Stability

• System logs revealed frequent faults in grid connectivity and inverter operations.

• Correlation analysis showed that higher inverter temperatures increased failure rates.

### 3. Renewable Energy Insights

• Higher wind speeds correlated with increased power generation, suggesting potential for hybrid energy models.

• Wind and solar energy data together provided insights into seasonal power variations.

## Conclusion and Future Scope

Through this study, we successfully validated and analyzed a large-scale solar power dataset, leading to insights on efficiency, fault detection, and energy optimization. Our findings highlight the need for real-time monitoring and predictive maintenance to improve power system performance.  
  
Future work can focus on implementing machine learning models for predictive fault detection and automated energy management. Additionally, integrating AI-driven anomaly detection can further enhance system reliability and reduce downtime in large-scale solar installations.

## Detailed Dataset Parameters

The dataset used in this project contains 117 parameters related to power generation, environmental conditions, fault detection, and system performance. Below is a categorized breakdown of key parameters and their significance.

### 1. Power System Parameters

These parameters represent the electrical characteristics of the system, including voltage, current, and power output.

InvVDVoltage\_V\_Avg, InvVa\_Avg, InvVb\_Avg, InvVc\_Avg, InvIa\_Avg, InvIb\_Avg, InvIc\_Avg, InvVDCin\_Avg, InvIDCin\_Avg, InvFreq\_Avg, InvPAC\_kW\_Avg, InvEtot\_kWh\_Max, InvPDC\_kW\_Avg, PwrMtrIa\_Avg, PwrMtrIb\_Avg, PwrMtrIc\_Avg, PwrMtrFreq\_Avg, PwrMtrP\_kW\_Avg, PwrMtrEdel\_kWh\_Max, PwrMtrErec\_kWh\_Max

### 2. Fault Detection and System Warnings

These columns indicate system faults, errors, and warnings encountered during power generation.

InvMainFault\_Max, InvDriveFault\_Max, InvVoltageFault\_Max, InvGridFault\_Max, InvTempFault\_Max, InvSystemFault\_Max, InvSystemWarn\_Max, InvPVMStatus\_Max, InvOpStatus\_Avg

### 3. Environmental Conditions

These parameters capture external conditions affecting power generation, including temperature, wind speed, and radiation.

SEWSAmbientTemp\_C\_Avg, SEWSModuleTemp\_C\_Avg, SEWSPOAIrrad\_Wm2\_Avg, WindSpeedAve\_ms, WindDirAve\_deg, WindDirStdDev\_deg, WindSpeed\_ms\_Max, WindRef\_V\_Min, AmbTemp\_C\_Avg, CR1000Temp\_C\_Avg

### 4. Energy Storage and Battery Parameters

These columns represent battery voltage levels, charging states, and system energy storage behavior.

Battery\_V\_Min, Battery\_A\_Avg, ChgState\_Min, ChgSource\_Min, CkBatt\_Max, Qloss\_Ah\_Max

### 5. Location and Timestamp Data

These parameters help track the geographical source and time of each recorded data point.

TIMESTAMP, latitude, longitude, valid\_time, timestamp, source\_file

### 6. Miscellaneous System Parameters

Additional columns related to grid stability, power quality, and phase monitoring.

PwrMtrPhaseRev\_Avg, PwrMtrVa\_Avg, PwrMtrVb\_Avg, PwrMtrVc\_Avg, PwrMtrP\_kVAR\_Avg, PwrMtrP\_kVA\_Avg, PwrMtrPF\_Avg, ShuntPDC\_kW\_Avg\_1, ShuntPDC\_kW\_Avg\_2, ShuntPDC\_kW\_Avg\_3, ShuntPDC\_kW\_Avg\_4, ShuntPDC\_kW\_Avg\_5, ShuntPDC\_kW\_Avg\_6, ShuntPDC\_kW\_Avg\_7