🌞 Solar Power Generation Prediction

🗓️ Week 1 Report – Data Loading, Cleaning, Merging & EDA

Intern: *Rishabh Singh Papola*
Duration: Week 1 (Project Theme: Energy)

1️⃣ Objective of Week 1

The goal of Week 1 was to perform **data understanding, cleaning, and exploration** on the provided datasets to prepare them for the prediction model in Week 2.

The tasks included:

- Loading and merging power generation and weather sensor datasets for both solar plants.
- Performing data cleaning, handling missing values, and feature engineering.
- Conducting **Exploratory Data Analysis (EDA)** to understand patterns and relationships between weather conditions and power generation.

2️⃣ Datasets Used

The project uses four CSV files:

1. Plant_1_Generation_Data.csv
2. Plant_1_Weather_Sensor_Data.csv
3. Plant_2_Generation_Data.csv
4. Plant_2_Weather_Sensor_Data.csv

Each generation file contains **power output readings** from multiple inverters, while the weather data files include **solar irradiance, temperature, and wind speed** readings from sensors.

3️⃣ Data Preprocessing Steps

- **Datetime Conversion:** Converted DATE_TIME columns to proper datetime format and extracted DATE, HOUR, and MONTH.
- **Merging:** Combined each plant's generation and weather datasets using the DATE_TIME column.

- **Missing Value Handling:** Filled missing data using **forward fill (ffill)** to ensure time-series continuity.
- **Feature Engineering:**
    - Created TOTAL_POWER as the sum of DC and AC power.
    - Extracted new columns for MONTH and WEEKDAY.

4️⃣ Exploratory Data Analysis (EDA)

📊 a. Basic Statistics

- Generated descriptive statistics for power and weather features.
- Identified consistent ranges for DC_POWER and AC_POWER across both plants.

🔥 b. Correlation Heatmap

- Created a heatmap showing correlations between numeric features.
- Observed that **solar irradiance and DC power** had a strong positive correlation.
- Ambient temperature showed a moderate negative effect on DC power (higher temperature, slightly reduced efficiency).

📈 c. Power Generation Over Time

- Plotted both **DC Power** and **AC Power** over time.
- Observed clear daily patterns—power generation increases during the day and drops at night.

🌡️ d. Temperature vs Power

- Scatterplot showed that as **ambient temperature** rises, DC power increases up to a point, then stabilizes or slightly drops (indicating optimal operating range).

⚡ e. Daily Energy Comparison

- Grouped data by DATE and summed daily DC power for each plant.
- Found that **Plant 2** generally produced slightly more energy than **Plant 1**, possibly due to better location or irradiance conditions.

5️⃣ Insights from Week 1

✅ Solar power generation is highly dependent on **irradiance** and **time of day**.
✅ Temperature impacts power output efficiency.
✅ Both plants exhibit similar daily generation patterns but differ in total output.
✅ The datasets are now **clean, merged, and feature-enriched**, ready for model building.

---

6 Plan for Week 2

In Week 2, we will:

- Prepare final feature sets for training and testing.
- Build and compare multiple regression models:
    - **Linear Regression**
    - **Random Forest**
    - **XGBoost**
- Evaluate models using **MAE**, **RMSE**, and **R^2 score**.
- Visualize prediction performance and deploy a small **Streamlit app** to demonstrate real-time prediction

7 Summary

Step	Task	Status
1	Data Loading & Inspection	Completed
2	Data Cleaning & Merging	Completed
3	Feature Engineering	Completed
4	Exploratory Data Analysis	Completed
5	Insights Documentation	Completed