**Solar Energy: Power Accuracy, Efficiency, Generation, and Consumption Prediction [Timeseries Data]**

## **Problem Statement**

This project focuses on developing a predictive system for **solar energy generation** and **load consumption** using real-world data from **SkyElectric** and **SkyLabs**. The system should analyze historical PV (Photovoltaic) system data to:

* Predict future PV power generation and load consumption at **10-minute intervals** within a look-ahead period of **1 to 4 hours**.
* Calculate **energy efficiency** (generation-to-consumption ratio).
* Recommend strategies to optimize system performance based on predicted results.

The approach to building the system is **flexible** but the **core prediction and efficiency calculation pipeline is mandatory**.

## **Key Objectives**

* Predict **solar generation** and **load consumption** for the specified look-ahead period.
* Calculate **energy efficiency** = (Generation ÷ Consumption) ratio.
* Apply and compare multiple prediction models such as **LSTM**, **ARIMA**, and **XGBoost**.
* Select and optimize the best-performing model(s) for accurate results.

## **Data Acquisition**

**Important:** The dataset is the property of SkyElectric and SkyLabs. It must **not** be shared outside the competition, and no derivative works or publications are allowed using this data.

The dataset contains **one year** of data at 10-minute intervals from multiple PV systems.

**Files:**

1. **systems\_new.csv** – Metadata about each PV system (system type, location, capacity).
2. **train\_data.csv** – Historical PV generation and load consumption data.
3. **test\_data\_masked.csv** – Test set with masked future values (-1 for required predictions, -2 for beyond the look-ahead).
4. **sample\_submission.csv** – Template for prediction submissions.

**Features:**

* **Timestamp**: Date & time of the sample (YYYY-MM-DD HH:MM:SS)
* **System ID**: Unique ID for each PV system
* **Location**: City of installation
* **Connection Type**: Commercial / Residential
* **Panels Capacity (kW)**
* **Load Capacity (kW)**
* **PV Generation (W)**
* **Load Consumption (W)**

## **System Components**

1. **Data Preprocessing & EDA**
   * Handle missing values and anomalies.
   * Detect trends, seasonality, and outliers.
   * Feature engineering (e.g., time-based features, weather-based features if available).
2. **Model Development**
   * Models: Students may use a justified model of their choice
   * Predict both **generation\_W** and **load\_W** for each 10-minute interval in the look-ahead period.
   * Optimize hyperparameters for each model.
3. **Energy Efficiency Calculation**
   * Formula: **Efficiency = Generation ÷ Consumption**.
   * Identify patterns in high and low efficiency periods.
4. **Performance Evaluation**
   * **Metric**: Mean Absolute Error (MAE) for both PV generation and load consumption.
   * Compare models to select the most accurate.

## **Deliverables**

* **Functional prediction system** for PV generation and load consumption.
* **Energy efficiency calculation module**.
* Submissions to Kaggle Leaderboard

**Participants will be provided access to data and competition if they select this project.**