**Solar Energy Prediction using Stacked Ensemble Models**

**Submitted by:  
Name: Taha Azmat**

**1. Tools and Libraries**

* **Pandas & NumPy: Data handling and computation.**
* **LightGBM, XGBoost, CatBoost: Gradient boosting models (GPU).**
* **Scikit-learn: Linear models, cross-validation, preprocessing.**
* **GroupKFold & MAE: Model evaluation.**
* **LabelEncoder: Encoding categorical variables.**

**2. Dataset Description**

* **train\_data.csv: Historical solar energy generation (training).**
* **test\_data\_masked.csv: Features without labels (testing).**
* **systems\_new.csv: Metadata of solar panel systems.  
  Merged system metadata with training and testing data for richer features.**

**3. Data Preprocessing**

* **Merge: Joined metadata (system\_id) with main datasets.**
* **Feature Engineering:**
  + **Time features (hour, day, month, etc.).**
  + **Lag features (1, 2, 3, 6 steps).**
  + **Rolling averages (3, 6 steps).**
* **Encoding: connection\_type and location encoded with LabelEncoder.**
* **Targets & Features:**
  + **Targets → generation\_W, load\_W.**
  + **Dropped timestamp and targets.**
  + **Remaining features used for training.**

**4. Model Development**

* **Cross-Validation: 5-fold GroupKFold (grouped by system\_id).**
* **Base Models:**
  + **LightGBM, XGBoost, CatBoost (GPU).**
  + **ElasticNet (CPU, with imputation).**
* **Stacking: Combined predictions with Ridge Regression as meta model.**
* **Evaluation: Mean Absolute Error (MAE) used on out-of-fold predictions.**

**5. Results and Submission**

* **Final predictions generated for both generation\_W and load\_W.**
* **Submission file submission.csv created with columns: system\_id, timestamp, generation\_W, load\_W.**