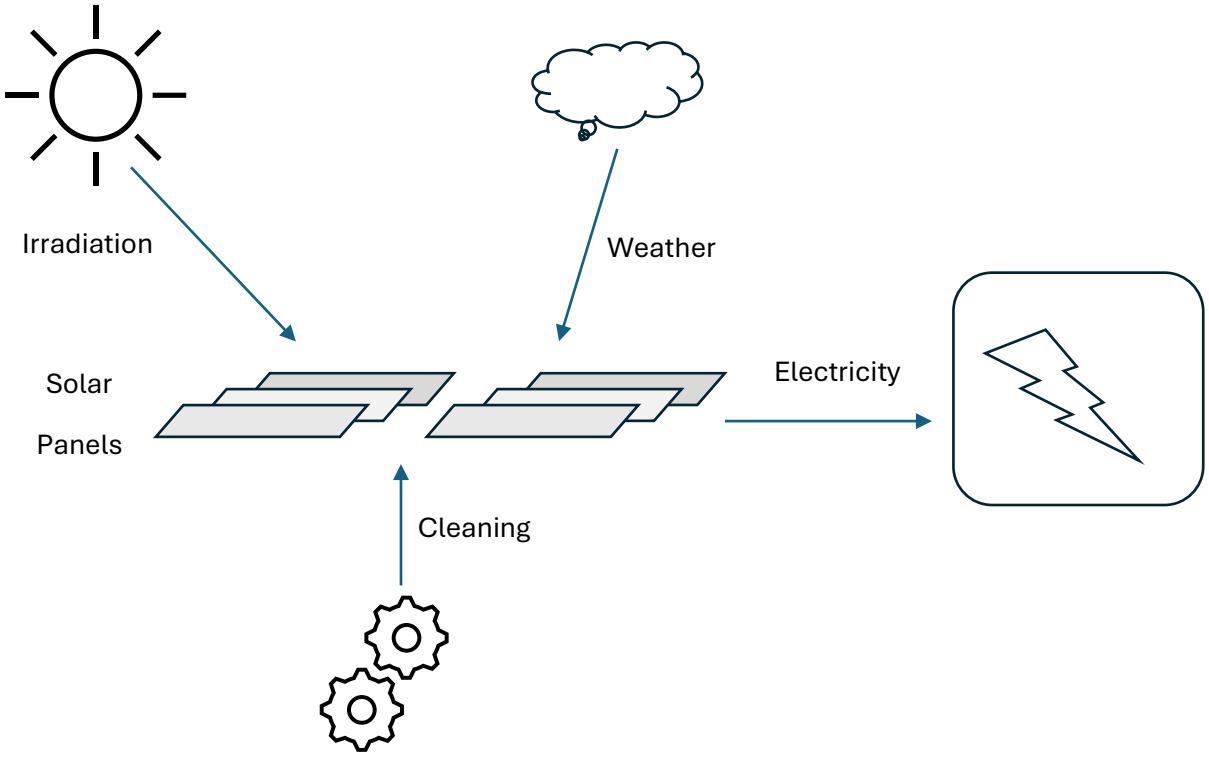
Solar Farm – Model Card

Predicting the Electricity Output based on Sensors and Weather Data

Overview



1. Model Name
    - Solar-Electric-Output
2. Version
    - 1.1
3. Model Type
    - Regression
4. Primary Use Case
    - Predicting the Output of a Solar Farm

Intended Use

5. Target Audience
    - Solar Farm owners or Students(for research)
6. Limitation
    - Some of the data has been created synthetically to simulate real life scenarios
7. Ethical considerations
    - N/A

Training Data

8. Data Sources: curated synthetic data based on real measurements patterns
    - Time series of measurements for 34 days every 15 minutes
    - Data points:
        - Date/Time of measurement
        - DC_Power
        - AC_Power
        - Irradiation
        - Ambient Temperature
        - Module Temperature
        - Days since last cleanup
        - Electricity Output

Model Architecture

9. Algorithms: Random Forest Regressor
10. Train/Test Split: 70/30
11. Cross Validation: 5 folds
12. Hyperparameters tuning: Bayesian Optimization
    - Max_depth = 24
    - Max_features = 0.28
    - Min_samples_leaf = 2
    - Min_split = 4
    - N_estimators = 158
13. Libraries
    - sklearn
    - RandomForestRegressor
    - BayesianOptimization
    - Numpy, Panda

Performance Metrics

14. Precision: based on RMSE (last training = 0.027781)
15. Benchmark Comparisons: compared against Polynomial, KNN and Gradient Boosting
16. Generalisation: it handles well unseen data
17. Bias Analysis: no major biases

Limitations

18. It assumes that all solar panels are the same. In reality a Solar Farm can have multiple brands of hardware where each of them has specific manufacturer parameters.

Deployment and Maintenance

19. The trained model is saved in pickled file
20. The Scaler of Input and Output are saved in separate files
21. To get the real output the values needs to be inverse transformed
22. It needs regular maintenance/training especially when different makes of Solar Panels are installed
23. Computing usage during training is lightweight considering that Random Forest is quite efficient. One powerful CPU and 64M RAM will be sufficient to perform calculations.
24. Recommended training: every week to capture seasonality in the weather pattern

---