Solar Power Generation Prediction

Week 2 Report – Model Building, Training & Evaluation

Intern: *Rishabh Singh Papola*
Duration: Week 2 (Project Theme: Energy)

Objective:

To build, evaluate, and compare multiple regression models for predicting solar power generation using the processed dataset prepared in Week 1.

Work Completed:

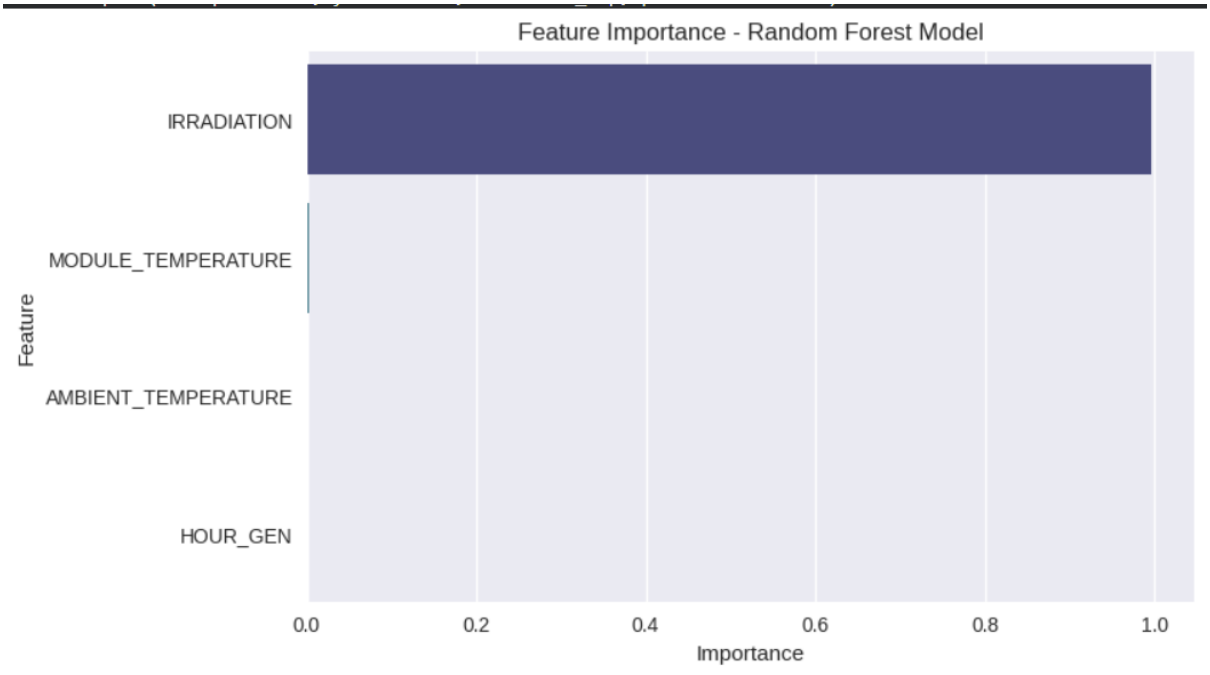
In Week 2, the focus was on developing machine learning models to accurately predict solar power output. The following tasks were performed:

1. **Data Preparation:**
    - Finalized the feature set for model training and testing, including *Ambient Temperature, Module Temperature, Irradiation,* and *Hour of Generation*.
    - Handled missing values and ensured proper data formatting.
    - Split the dataset into training (80%) and testing (20%) subsets for model evaluation.
2. **Model Building:**
    - Developed and trained three regression models:
        - **Linear Regression** – for baseline performance comparison.
        - **Random Forest Regressor** – to capture non-linear relationships.
        - **XGBoost Regressor** – for efficient and accurate gradient boosting.
3. **Model Evaluation:**
    - Assessed model performance using **Mean Absolute Error (MAE)**, **Root Mean Square Error (RMSE)**, and **R^2 Score**.
    - Compared results to identify the most accurate model for solar power prediction.
4. **Visualization:**
    - Created plots showing **Actual vs Predicted Power Generation** for visual comparison.
    - Displayed **Feature Importance** using Random Forest to highlight key influencing factors.

5. **Deployment Preparation:**

- Designed a simple **Streamlit app** structure to demonstrate real-time solar power prediction.

Linear Regression	266.264856	566.999792	0.980106
Random Forest	168.379704	469.242796	0.986374
XGBoost	175.684995	474.674222	0.986057



Results & Insights:

- The **Random Forest** and **XGBoost** models outperformed Linear Regression, achieving higher R^2 and lower error values.
- **Irradiation** was the most significant factor affecting power generation, followed by **Module Temperature**.
- The models successfully captured variations in daily power output trends.

Improvisations Done:

Enhanced prediction accuracy by comparing multiple models, optimizing key parameters, and introducing feature importance analysis for better interpretability.

Next Steps (Week 3):

- Perform hyperparameter tuning for model optimization.
- Finalize and deploy the Streamlit application for real-time prediction visualization.
- Prepare a summarized report comparing all model performances.