**Project Report: Solar Panel Efficiency Prediction**

**🔍 Project Overview**

This project aims to develop a machine learning-based model to predict the **efficiency of solar panels** using various sensor and environmental data. The motivation stems from the critical role solar energy plays in sustainable development and the need to optimize the performance of photovoltaic (PV) systems.

**📘 Brief Introduction to Solar Panels**

**Solar panels**, or photovoltaic (PV) panels, convert sunlight into electricity using the photovoltaic effect. They are composed of semiconductor materials—typically silicon—that absorb photons and release electrons, generating an electric current.

**Key Parameters Affecting Solar Panel Efficiency:**

* **Temperature**: Excess heat can reduce panel output.
* **Solar Irradiance**: The intensity of sunlight directly impacts energy generation.
* **Humidity & Dust**: Environmental factors can block or scatter sunlight.
* **Panel Angle & Orientation**: Determines how effectively panels capture sunlight.
* **Material Degradation**: Over time, panels degrade and lose efficiency.

**🎯 Project Goal**

To **predict the efficiency** of a solar panel using historical data comprising:

* Environmental attributes (e.g., temperature, irradiance)
* Panel characteristics (e.g., current, voltage)
* Sensor-generated time-series readings

This allows:

* **Proactive Maintenance**: Detecting degradation before failure
* **Performance Monitoring**: Ensuring optimal operation
* **System Optimization**: Enhancing long-term energy output

**📊 Dataset Description**

* **Source**: Provided dataset simulating real-time operational and environmental data.
* **Target Variable**: efficiency
* **Features**: Includes irradiance, temperature, voltage, current, and possibly timestamps.

### 📊 Dataset Description

**Source**: Zelekstra ML Challenge  
**Description**: The dataset includes various environmental and panel-specific parameters to predict the efficiency of solar panels. Below is a detailed explanation of each feature:

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| id | Unique identifier for each row. |
| temperature | Ambient air temperature in °C; affects panel efficiency. |
| irradiance | Solar energy received per unit area (W/m²); key driver of energy generation. |
| humidity | Air moisture content; may influence long-term panel performance. |
| panel\_age | Age of the panel in years; older panels may be less efficient. |
| maintenance\_count | Number of previous maintenance events recorded. |
| soiling\_ratio | Ratio (0–1) indicating efficiency loss due to dirt or debris. |
| voltage | Voltage (V) output from the panel. |
| current | Current (A) output, influenced by irradiance and panel condition. |
| module\_temperature | Surface temperature of the panel. |
| cloud\_coverage | Percentage of sky covered by clouds; impacts irradiance. |
| wind\_speed | Wind speed (m/s); may cool panels and affect temperatures. |
| pressure | Atmospheric pressure in hPa. |
| string\_id | Identifier for the group of connected panels (e.g., A1, B2). |
| error\_code | Diagnostic error codes from the panel (e.g., E00, E01). |
| installation\_type | Mounting setup: fixed, tracking, or dual-axis. |
| efficiency | **Target variable**; actual output efficiency of the panel. |

**🧠 Model Development**

**1. Data Preprocessing**

* Missing value handling
* Feature scaling
* Feature engineering (like time-based trends if applicable)

**2. Model Training**

Multiple models were experimented with:

* **Linear Regression**
* **Random Forest Regressor**
* **Gradient Boosting Regressor**
* **XGBoost**

After evaluation, the model providing the best performance on the validation set was selected.

**3. Evaluation Metric**

The model was evaluated using a **custom score**:

Score=100×(1−MSE(actual, predicted))\text{Score} = 100 \times \left(1 - \sqrt{\text{MSE(actual, predicted)}}\right)Score=100×(1−MSE(actual, predicted)​)

This score emphasizes minimizing the root mean squared error (RMSE), normalized for better interpretation.

**📄 Research References**

The project references cutting-edge research for theoretical backing and practical insights:

1. **IEEE Paper**: “Predicting Photovoltaic Efficiency Using Data-Driven Models”  
   🔗 <https://ieeexplore.ieee.org/document/9696845>
   * Discusses real-world ML applications on PV data
   * Compares performance of various regression models
2. **Elsevier Paper**: “Machine learning methods for performance modeling of solar panels”  
   🔗 Direct PDF
   * Explores ensemble methods for PV prediction
   * Offers performance improvement tips using weather datasets

1. **Title**: Solar Panel Performance Prediction using ML Models  
   **Available at**: https://ssrn.com/abstract=4771205

### Model Results

|  |  |
| --- | --- |
| **Model** | **Accuracy (%)** |
| Linear Regression | **89.37** |
| Gradient Boosting | 89.11 |
| Random Forest | 88.78 |
| XGBoost | 88.33 |

**💡 Conclusion**

**Predicting solar panel efficiency** is vital for the growing solar industry. Using machine learning:

* Reduces downtime by identifying underperformance
* Improves planning and resource allocation
* Enhances return on investment for solar infrastructure