**Machine Learning Approach for Predicting Solar Energy Generation**

**Chosen ML Paradigm: Supervised Learning**

This project focuses on predicting solar energy generation using historical weather and solar irradiance data. Given that we have labeled data — inputs (weather features like temperature, humidity, irradiance) and corresponding outputs (solar energy produced) — this is a classic **supervised learning** problem. The goal is to learn a mapping from features to target values, enabling the model to make accurate future predictions.

Supervised learning is especially suitable because:

* The relationship between inputs and outputs is mappable with enough historical data.
* Performance can be quantitatively evaluated using standard metrics (e.g., RMSE, MAE).
* It allows us to provide actionable insights for energy planning and storage optimization.

**Proposed Model: Gradient Boosting Regressor (e.g., XGBoost or LightGBM)**

We propose using **Gradient Boosting Regression**, a robust ensemble technique that combines multiple weak learners (typically decision trees) into a strong predictive model. It’s particularly powerful for tabular data and excels in capturing non-linear relationships.

**Why XGBoost or LightGBM?**

* Handles missing values and noisy data effectively.
* Offers high accuracy without requiring deep neural networks.
* Supports model interpretability, which is critical for real-world energy planning.
* Easily tunable for performance improvement.

Alternative models like **Linear Regression** may underperform due to the likely non-linear interactions in environmental data. Neural networks (e.g., CNNs or RNNs) are more complex and data-hungry, making them less ideal unless very large time-series datasets are available.

**Model Pipeline Overview:**

1. **Data Collection**  
   Collect historical weather and solar energy generation data for Ghana and Kenya.
2. **Data Preprocessing**  
   Clean missing values, normalize/standardize features, engineer time-based features (e.g., hour, season).
3. **Feature Selection**  
   Select key features such as irradiance, temperature, humidity, cloud cover, etc.
4. **Train/Test Split**  
   Divide dataset into training and testing sets (e.g., 80/20 split).
5. **Model Training**  
   Train XGBoost/LightGBM regressor on training data.
6. **Model Evaluation**  
   Evaluate using metrics like RMSE, MAE, and R².
7. **Deployment/Prediction**  
   Use the trained model to predict solar generation for future weather forecasts.