**Project 1:** Machine Learning-Based Prediction of Power Output in Solar PV Systems

**Description:**

The increasing reliance on renewable energy sources necessitates accurate forecasting models to optimize power generation and resource allocation. This project aims to develop a machine learning (ML) model to predict the power output of solar photovoltaic (PV) systems. Specifically, the project will:

1. Develop an ML model to predict power output based on historical solar PV data and weather condition.
2. Implement hyper-parameter tuning techniques to enhance model performance.
3. Extract seasonal information from the date
4. Analyze the influence of seasonal data on prediction effectiveness.
5. Examine the most influential features affecting various prediction models.

**Question:**

1. Data Preparation:
   1. Handle missing and duplicates data
   2. Use date to extract seasonal information (Season, Month and day of the month).
2. Dataset and Exploratory Data Analysis (EDA) The dataset contains multiple attributes related to solar PV power generation. To ensure data quality and uncover insights, the following EDA techniques should be applied:

* Histogram Analysis: To visualize the distribution of individual features.
* Box Plots: To detect outliers and understand data spread.
* Scatter Plots & Correlation Matrix: To explore relationships between variables and identify dependencies.

1. Feature Selection Techniques Feature selection plays a crucial role in improving model accuracy and efficiency. This project should compare different techniques, including:

* Principal Component Analysis (PCA): To reduce dimensionality while preserving variance.
* Chi-Square (Chi2) Test: To determine feature importance for categorical data.

1. Machine Learning Models and Hyper-Parameter Tuning The project should employ two powerful ML models:

* XGBoost (Extreme Gradient Boosting): A robust decision-tree-based ensemble method.
* Random Forest (RF): An ensemble learning method that builds multiple decision trees.

To optimize performance, hyper-parameter tuning techniques like Grid Search should be utilized.

5. Performance Metrics Model performance should be evaluated using four key metrics:

1. Mean Absolute Error (MAE): Measures average absolute prediction error.
2. R-Square (R2) Score: Assesses model goodness-of-fit.
3. Root Mean Square Error (RMSE): Evaluates error magnitude.
4. Normalized RMSE (NRMSE): Compares RMSE to the range of observed data.

6. Report Structure (Max 5 Pages) The final report will document the following:

* Introduction & Objectives
* EDA Results & Visualizations
* Feature Selection & Comparison of Techniques
* Model Development & Hyper-Parameter Tuning
* Performance Evaluation & Seasonal Influence Analysis
* Conclusion & Key Findings