**Project Title:**

PowerPulse: Household Energy Usage Forecast

**Problem Statement:**

This Project invloves predicting the global active power using various ML Models & visualising the model performance.

**Tools Used:**

* + Python 3.12.8

**IDE Used:**

* + MS Visual Studio Code

**Approach:**

1.Initially the given dataset is loaded using pandas library

2.The given dataset is checked for datatypes correction ,null values & duplicate values.

3.The required data type conversion is done & null & duplicate values are removed.

4.Addtional columns for Daily average , IsPeakHour & Season is added.

5. Correlation analysis is performed to understand the relationship between different parameters in the dataset.

6.The analysis clearly indicate that there is linear relationship between global intensity & global active power. Global intensity is highly affected by submetering1,2,3 values.

7.The input & output variables are determined as Global\_Intensity,Sub\_Metering\_1,Sub\_Metering\_2,Sub\_Metering\_3,IsPeak\_Hour,Season and Global\_active\_power respectively.

8.Testing & training data is split uing train\_test\_split.

9.Basic Linear regression model is built & its performance is evaluated using root men squared error, r2 score & mean absolute error values as the metrics. The same is visualized by plotting a histogram for residuals(True value-predicted value).

10.Further , Decision Tree , Random Forest Regressor , Gradient Boosting ,neural networks & SVR models are also built on the same data in a similar manner & their performances are evaluated

**Model Selection ,Insights & Recommendations:**

Initially , a simple linear regression model was built as the global intensity & global active power had a linear relationship. The model predictive performance was really good , subsequently Decision tree , Random forest , gradient boosting & neural networks was also tried . All models performed very well with little effort in hyperparameter tuning. With gradient boosting it was bit higher compared to other models

The given dataset is a very clean dataset with no duplicates & correlation analysis helped in identifying the parameter relationship.

For this dataset, a simple linear regression model itself gave good output. For this Dataset , either a simple linear Regression or Gradient Boosting Model is sufficient.