**1. Approach**

This project aims to predict household energy consumption using a machine learning model. The goal is to understand patterns in historical energy data and develop a reliable model to forecast future energy usage.

**Steps followed:**

* Data cleaning and preprocessing
* Exploratory data analysis (EDA)
* Model selection and training
* Hyperparameter tuning
* Performance evaluation
* Visualization of results and interpretation

**Dataset:**

Contains minute-wise energy consumption metrics from a household

Key variables include:

* Global\_active\_power (target)
* Global\_reactive\_power
* Voltage
* Sub\_metering\_1, Sub\_metering\_2, Sub\_metering\_3
* Time-based features (hour, day, weekday)

**Preprocessing:**

* Converted Datetime to datetime object
* Handled missing values
* Extracted additional time features (hour, day, month)

**Findings:**

* Energy usage peaks during evening hours
* Weekdays have higher average consumption
* Energy use peaks in evenings

**Model Selection and Evaluation**

Three models were evaluated on both training and testing datasets.

**Linear Regression**

| **Metric** | **Training** | **Test** |
| --- | --- | --- |
| MAE | 0.3583 | 0.3580 |
| RMSE | 0.5308 | 0.5303 |
| R² | 0.7478 | 0.7492 |

**Random Forest Regressor**

| **Metric** | **Training** | **Test** |
| --- | --- | --- |
| MAE | 0.2717 | 0.2730 |
| RMSE | 0.4345 | 0.4365 |
| R² | 0.8310 | 0.8301 |

**Gradient Boosting Regressor**

| **Metric** | **Training** | **Test** |
| --- | --- | --- |
| MAE | 0.2469 | 0.2492 |
| RMSE | 0.4022 | 0.4065 |
| R² | 0.8552 | 0.8526 |

**Performance Visualizations:**

* Actual vs Predicted Plot: Shows how closely predictions align
* Residual Plot: Detects systematic prediction errors
* Feature Importance: Highlights top contributing features