POWERPULSE: HOUSEHOLD ENERGY USAGE FORECAST

PROBLEM STATEMENT:

In the modern world, energy management is a critical issue for both households and energy providers. Predicting energy consumption accurately enables better planning, cost reduction, and optimization of resources. The goal of this project is to develop a machine learning model that can predict household energy consumption based on historical data. Using this model, consumers can gain insights into their usage patterns, while energy providers can forecast demand more effectively.

By the end of this project, learners should provide actionable insights into energy usage trends and deliver a predictive model that can help optimize energy consumption for households or serve as a baseline for further research into energy management systems

BUSINESS USE CASES:

- Household Energy Management – Monitor usage and reduce electricity bills
- Demand Forecasting – Helps energy providers manage loads and pricing
- Anomaly Detection – Identify faults or unauthorized power usage
- Smart Grid Integration – Real-time energy optimization using AI
- Environmental Sustainability – Support carbon footprint reduction

DATASET:

• Source: UCI Machine Learning Repository (Individual Household Electric Power Consumption)

• Size: ~2 million rows

• Time Frame: December 2006 to November 2010

• Frequency: 1-minute intervals (resampled to hourly)

Key Variables:

- Date & Time

- Global Active Power (target)

- Voltage, Global Intensity, Sub Meterings (1, 2, 3)

EXPLORATORY DATA ANALYSIS (EDA)

EDA Highlights:

- Resampled data to hourly intervals for smoother trends
- Identified seasonal and daily patterns in energy consumption
- Detected outliers and handled missing values
- Visualized trends using time-series and box plots

DATA PREPROCESSING & FEATURE ENGINEERING

Data Preparation:

- Handled missing values (NA removal)
- Extracted datetime features (hour, day, weekday)
- Applied scaling using StandardScaler for model training
- Engineered lag features: 1-hour and 24-hour lagged power usage

MODEL BUILDING

Models Used:

Linear Regression

Random Forest Regressor

Gradient Boosting Regressor

Training Approach:

- Train-test split (80-20)
- Feature scaling
- Cross-validation and hyperparameter tuning
- Saved best models using joblib

EVALUATION METRICS

Evaluation Metrics:

• Root Mean Squared Error (RMSE)

• Mean Absolute Error (MAE)

• R-Squared (R^2)

Best Performing Model: Random Forest Regressor

Achieved:

- RMSE: ~0.27

- MAE: ~0.18

- R^2 Score: ~0.93

ACTUAL VS PREDICTED

Prediction Performance:

- Plotted actual vs predicted energy consumption
- Smooth trend line confirms close tracking
- Residuals show minimal and symmetric error distribution

Lag features improved model precision further

MODEL DEPLOYMENT READINESS

Model Deployment Steps:

- Saved the model and scaler with joblib
- Validated reusability by reloading and predicting
- Sample prediction tested from unseen data
- Pipeline ready for integration in real-time applications

TAKEAWAYS

Key Takeaways:

- Successfully predicted energy consumption using ML
- Achieved high accuracy with Random Forest + Lag Features
- Provided interpretable insights via feature importance
- Built reusable model pipeline for deployment

Real-world impact:

- Enables smarter homes
- Helps energy companies optimize grid performance