PowerPulse: Household Energy Usage Forecast

PowerPulse: Household Energy Usage Forecast

A machine learning pipeline for predicting household energy consumption using historical data. This system helps consumers understand usage patterns and enables providers to forecast demand.

Project Overview

Features

- End-to-end ML pipeline from raw data to predictions
- Multiple model comparison (XGBoost, Random Forest, etc.)
- Feature engineering for temporal patterns
- Comprehensive evaluation metrics
- Visualizations of key insights

Project Structure

📂 **Project Structure**

energy-consumption/ ├── data/ │ ├── raw/ # Original dataset │ ├── processed/ # Cleaned and normalized data │ └── features/ # Data with engineered features ├── models/ # Trained model binaries ├── reports/ # Evaluation reports ├── visualizations/ # Generated plots and charts └── scripts/ ├── 1_data_understanding.py ├── 2_data_preprocessing.py ├── 3_feature_engineering.py ├── 4_model_training.py ├── 5_model_evaluation.py └── utils.py # Helper functions

Technologies Used

Programming and Data Science:

Python 3.8+ - Core programming language

Pandas - Data manipulation and analysis

NumPy - Numerical computing

Scikit-learn - Machine learning algorithms

XGBoost - Gradient boosting framework

Matplotlib - Data visualization

Seaborn - Statistical data visualization

Joblib - Model serialization

Machine Learning Models

- **Linear Regression** - Baseline model
- **Random Forest** - Ensemble learning
- **Gradient Boosting** (HistGradientBoosting) - Advanced ensemble
- **XGBoost** - Optimized gradient boosting
- **Neural Networks** (MLP) - Deep learning approach

Data Processing & Feature Engineering

- **Scikit-learn Preprocessing** - Data normalization (MinMaxScaler)
- **Pandas DateTime** - Time series feature extraction
- **Lag Features** - Temporal pattern capture
- **Rolling Statistics** - Moving averages and standard deviations

Run a code

1. Exploratory Data Analysis
python scripts/1_data_understanding.py

2. Data Cleaning and Normalization
python scripts/2_data_preprocessing.py

3. Feature Engineering
python scripts/3_feature_engineering.py

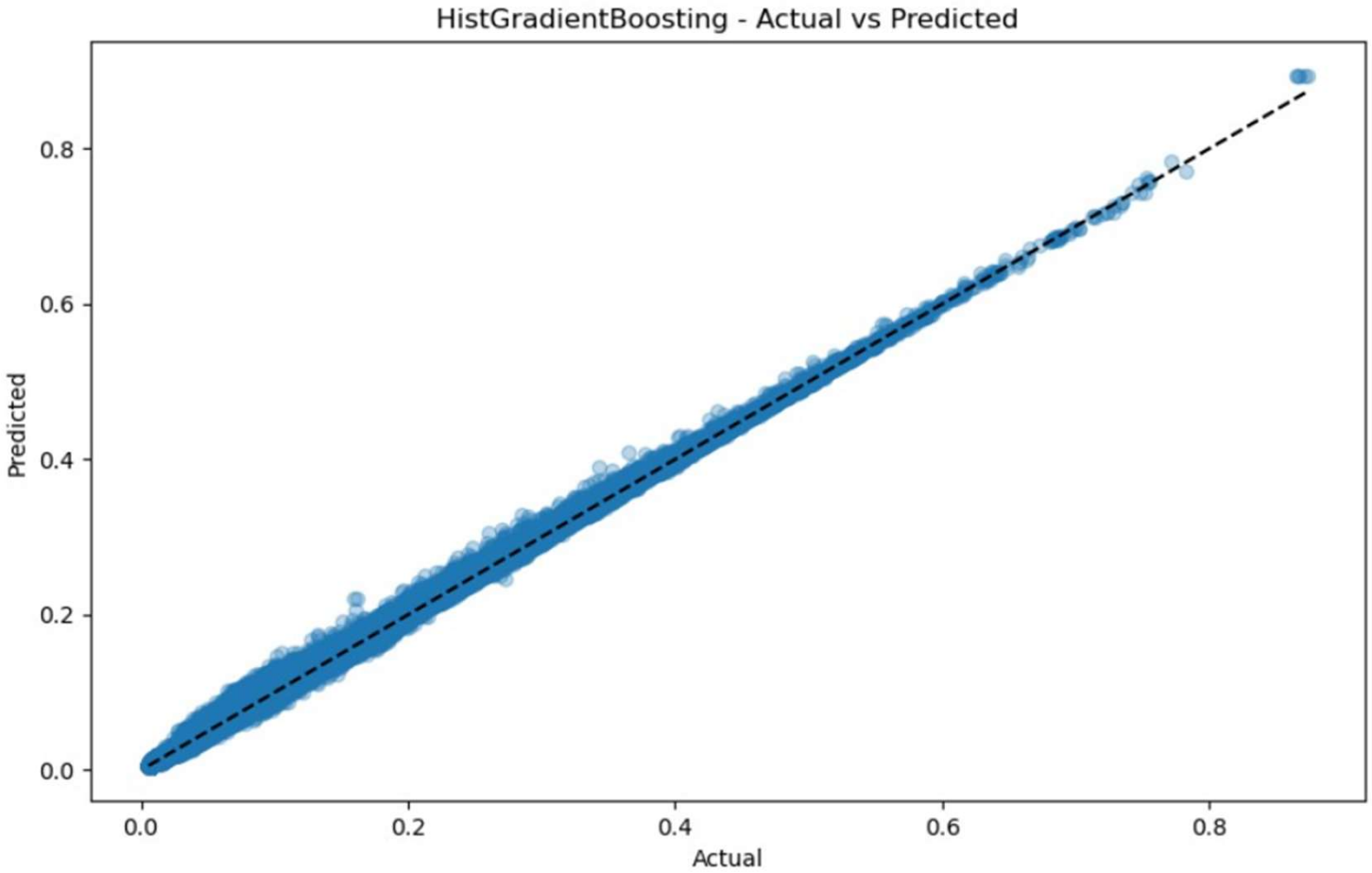
4. Model Training
python scripts/4_model_training.py

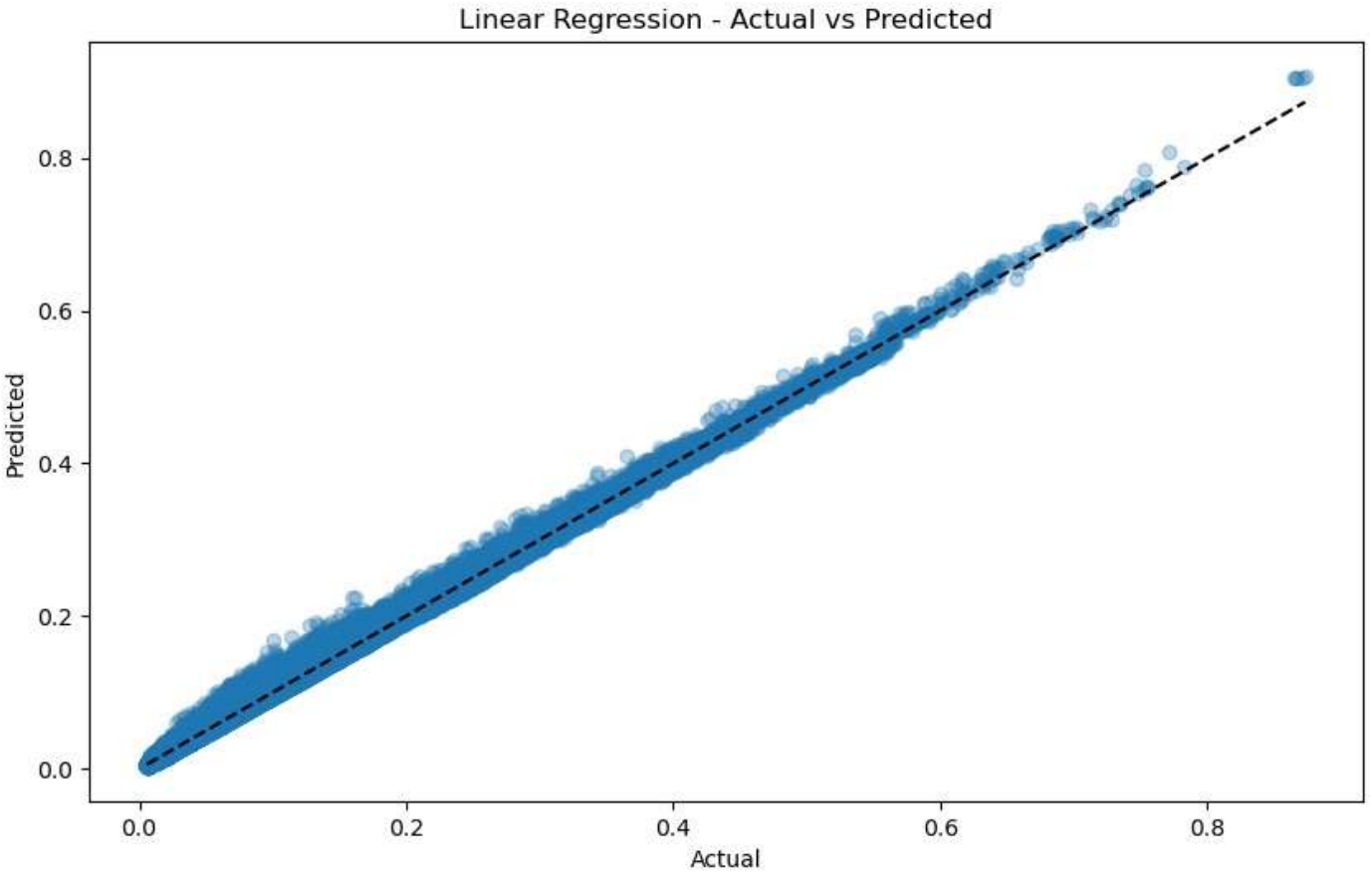
5. Model Evaluation
python scripts/5_model_evaluation.py

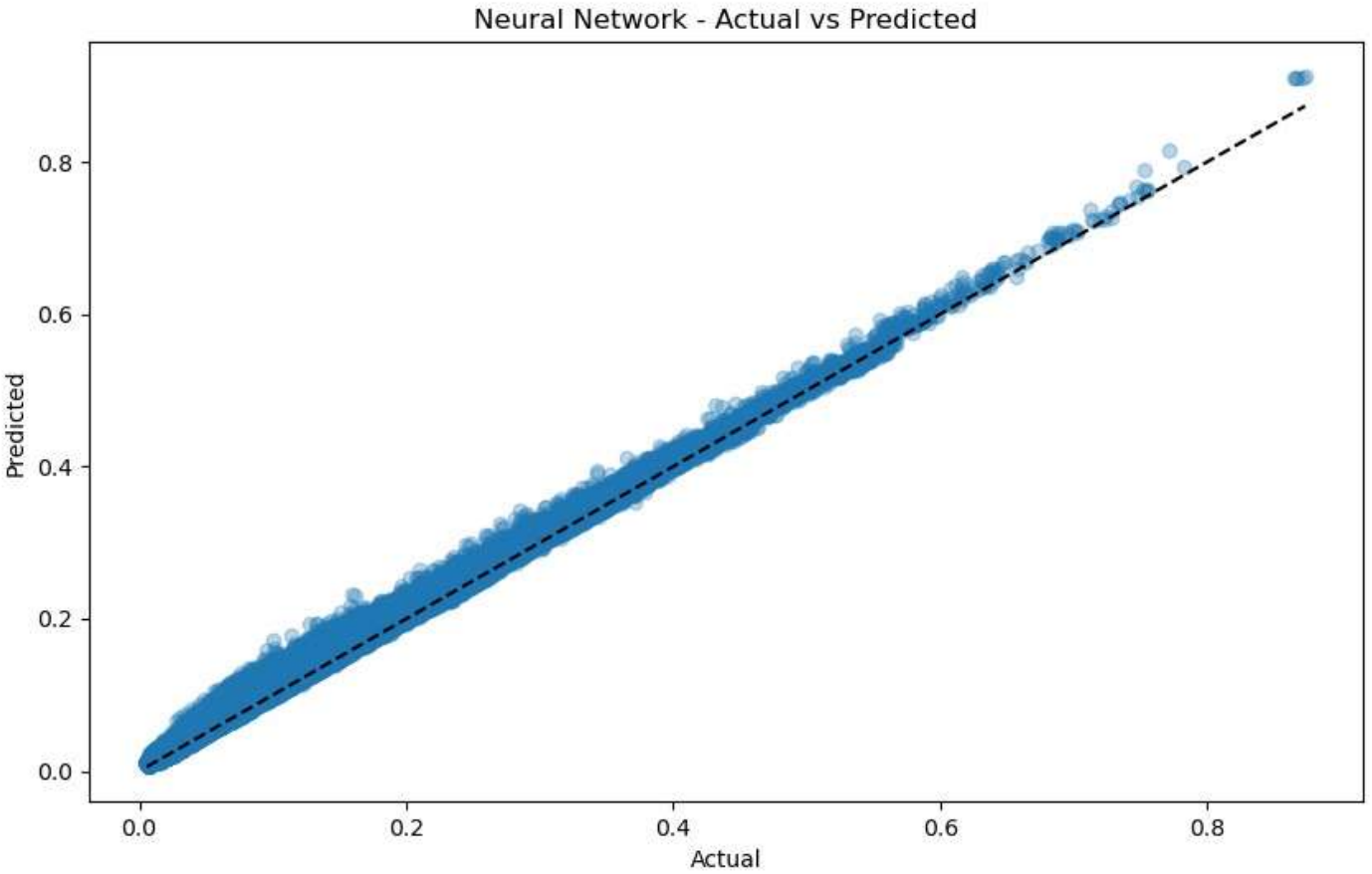
Results

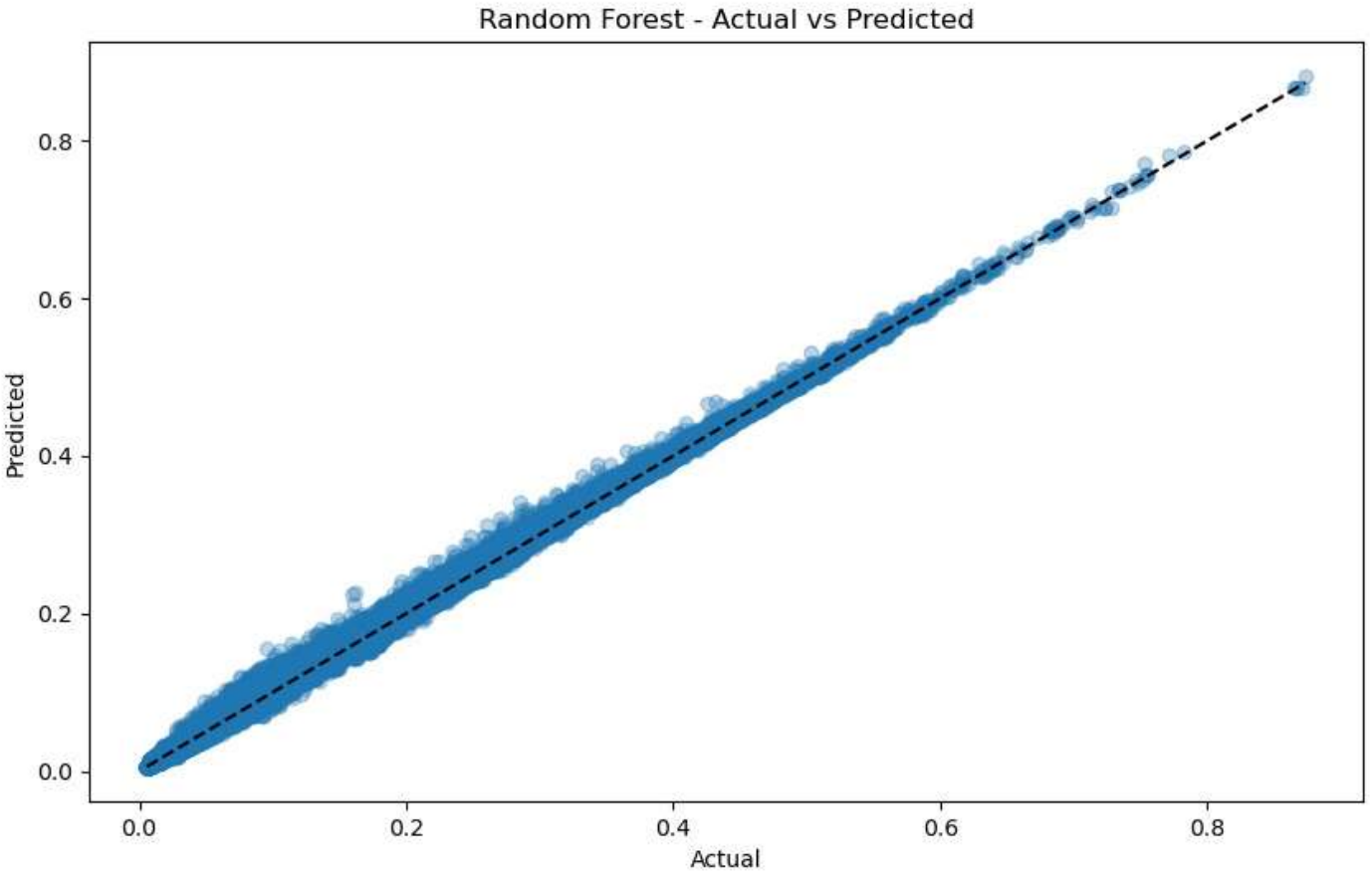
Evaluation Complete!

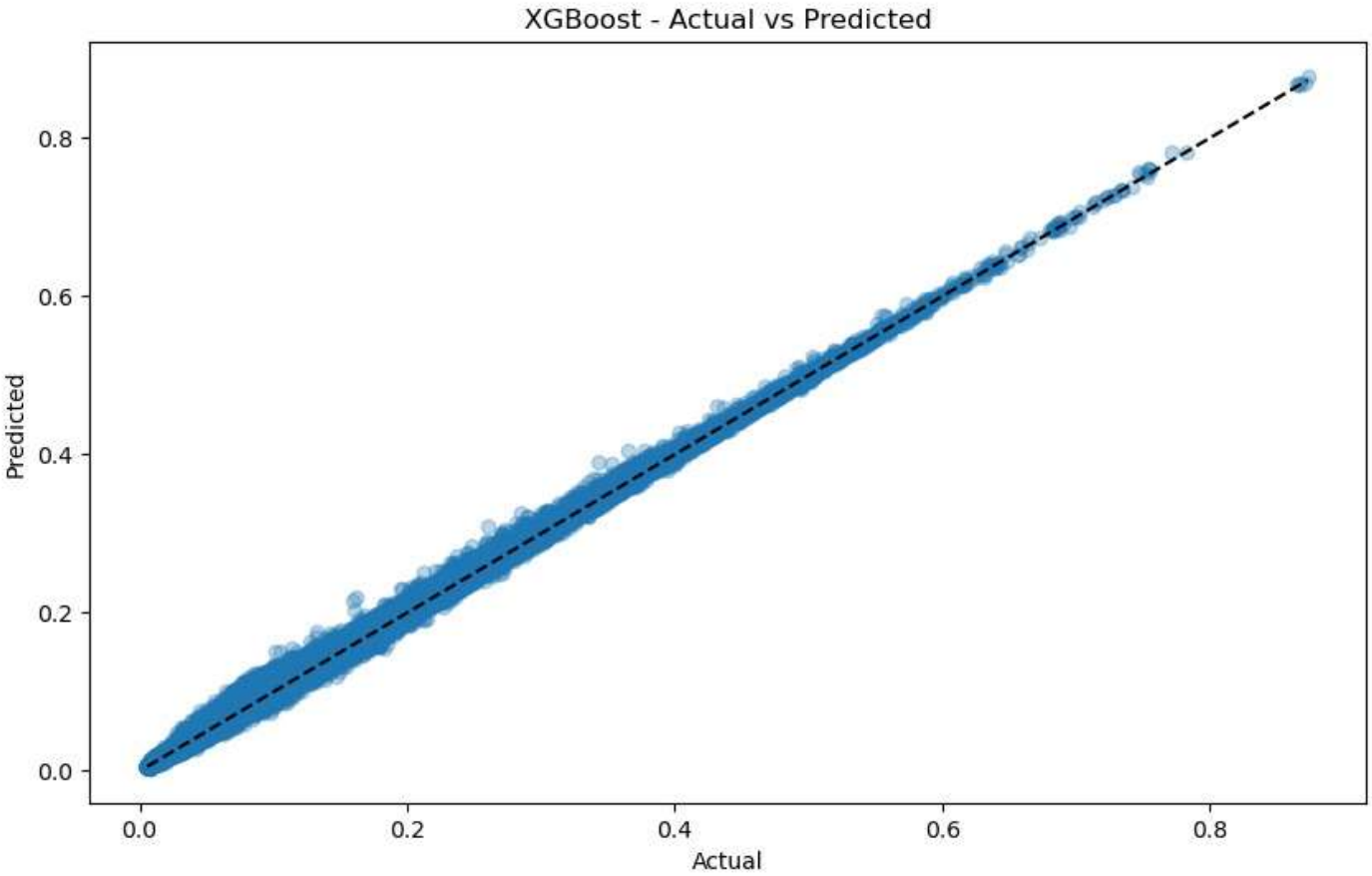
	Model	RMSE	MAE	R2
0	Linear Regression	0.00347165	0.00210404	0.998096
1	Random Forest	0.00293257	0.00152246	0.998641
2	HistGradientBoosting	0.00233599	0.0013802	0.999138
3	XGBoost	0.00223969	0.00129153	0.999208
4	Neural Network	0.00692777	0.00576164	0.992418

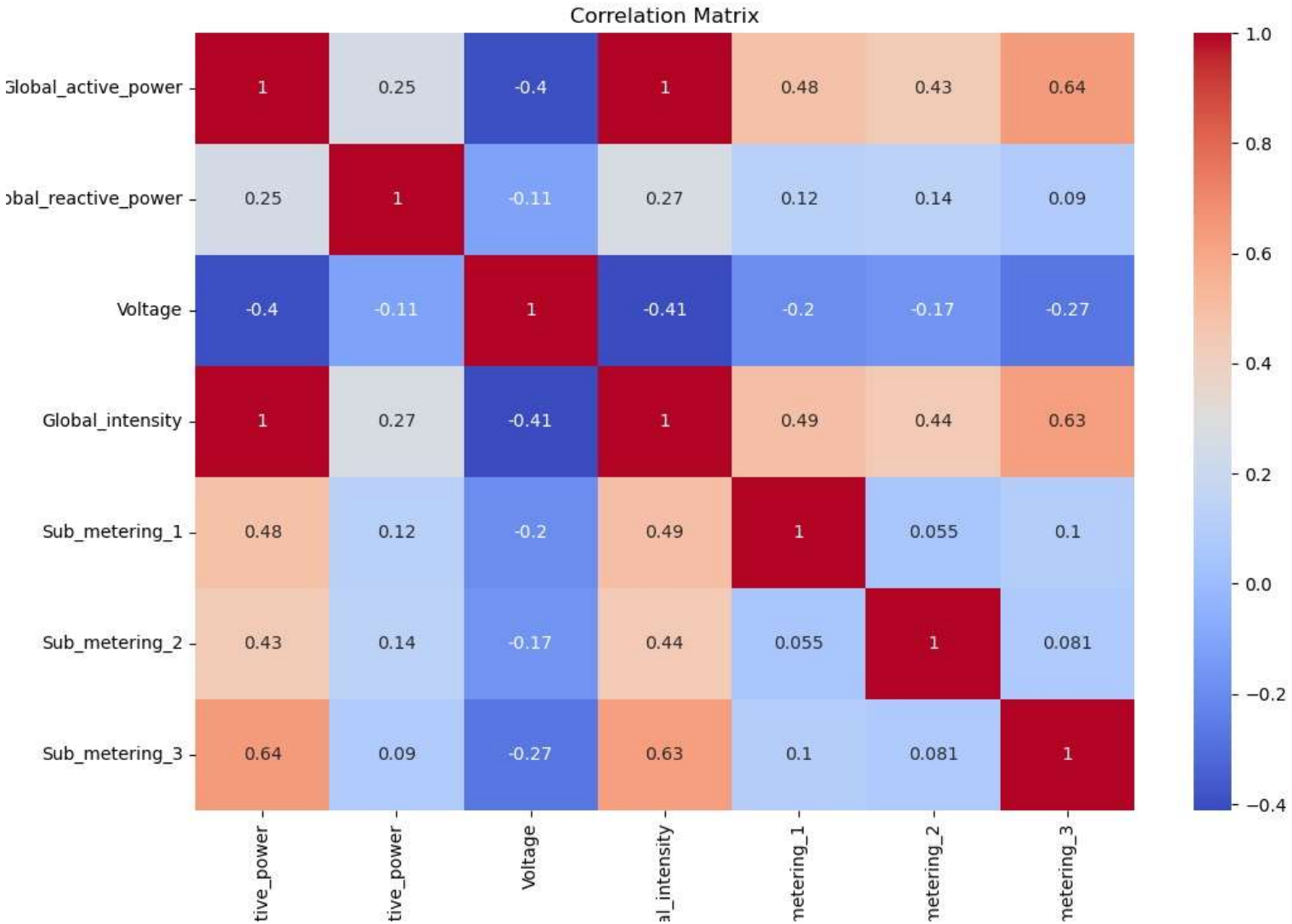

HistGradientBoosting - Actual vs Predicted
Predicted
Actual
0.0
0.2
0.4
0.6
0.8

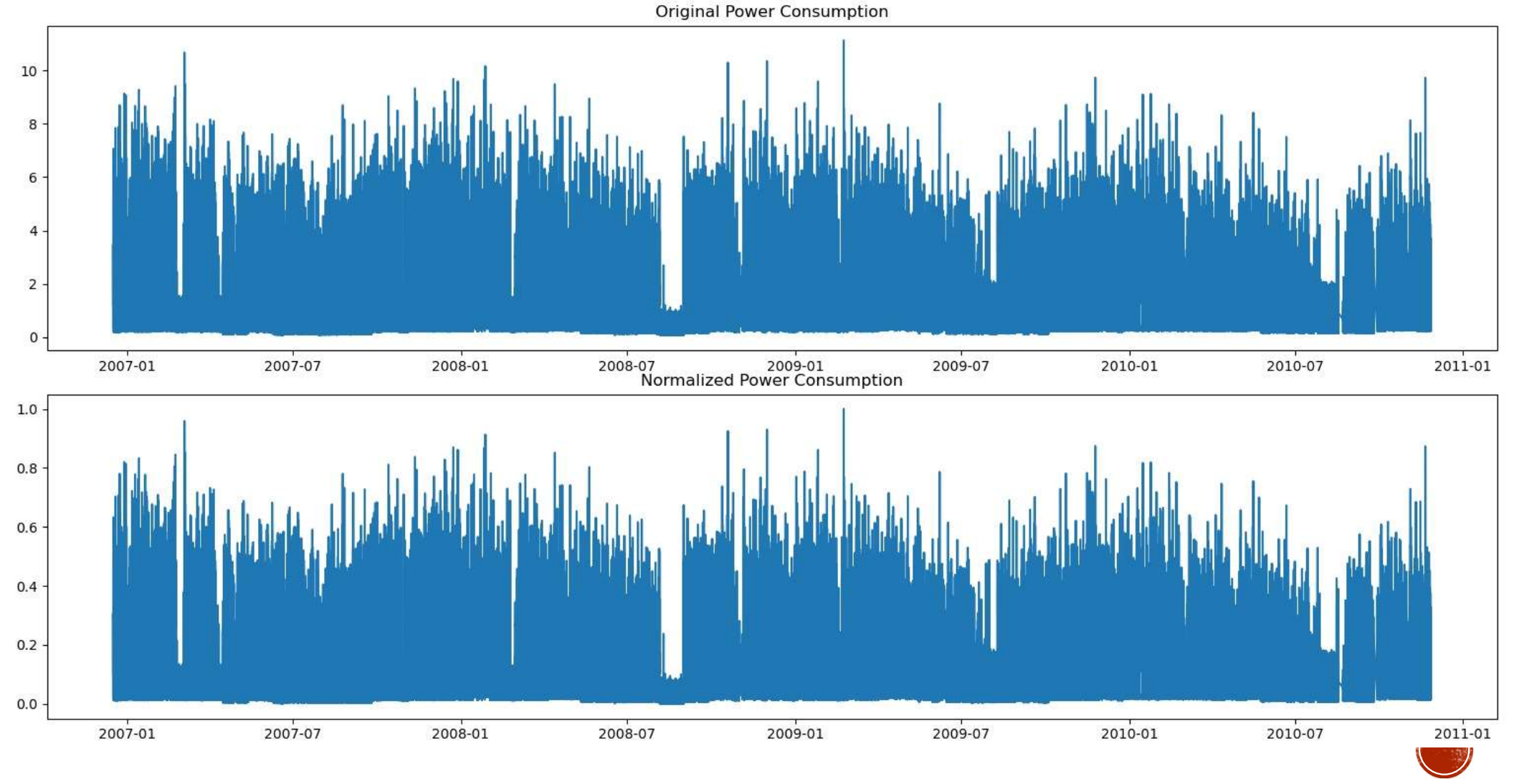

Linear Regression - Actual vs Predicted
Predicted
Actual
0.0
0.2
0.4
0.6
0.8

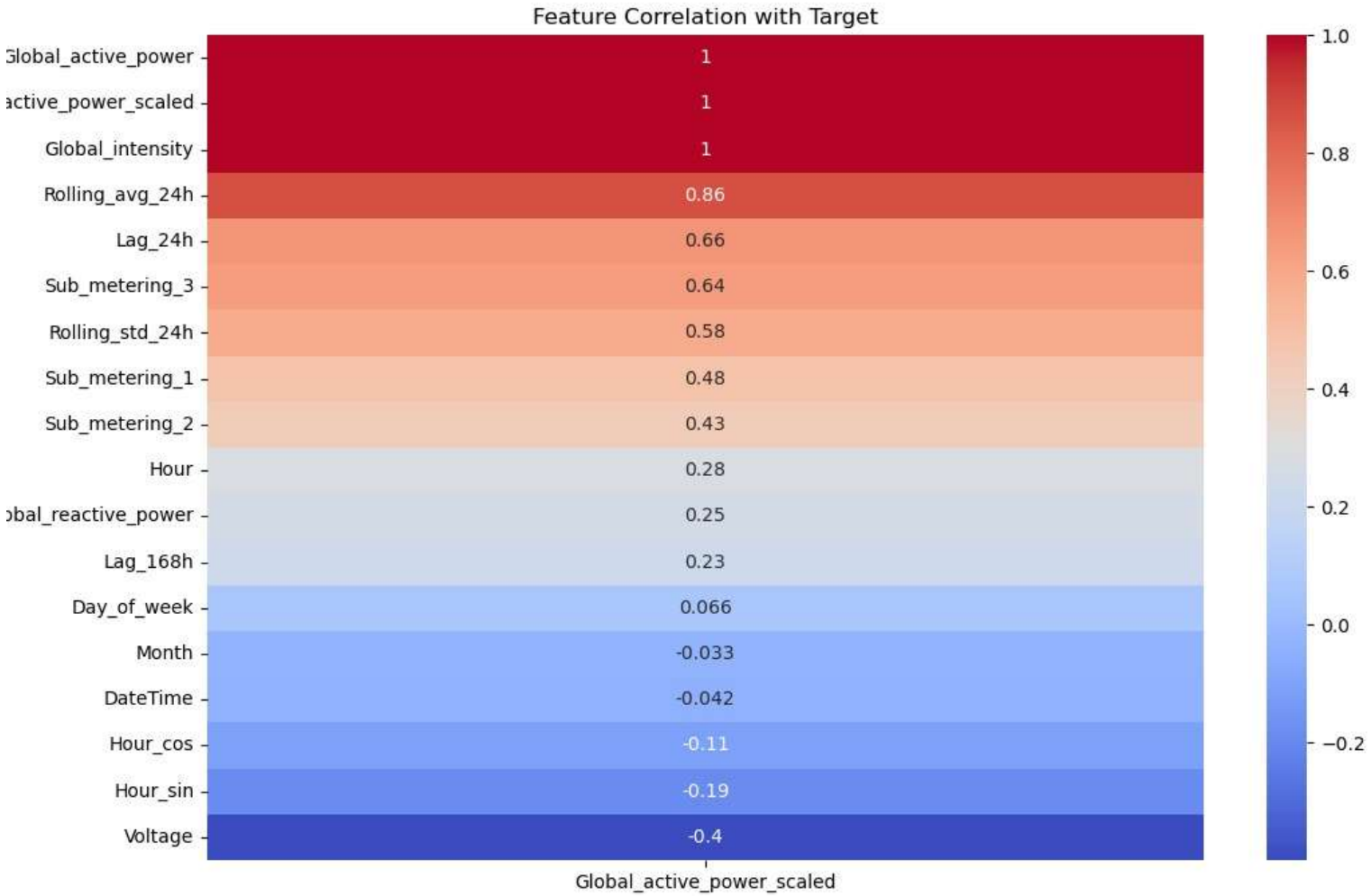

Neural Network - Actual vs Predicted
Predicted
Actual
0.0
0.2
0.4
0.6
0.8


Random Forest - Actual vs Predicted
Predicted
Actual
0.0
0.2
0.4
0.6
0.8


XGBoost - Actual vs Predicted
Predicted
Actual
0.0
0.2
0.4
0.6
0.8


Correlation Matrix
Global_active_power
obal_reactive_power
Voltage
Global_intensity
Sub_metering_1
Sub_metering_2
Sub_metering_3
1 0.25 -0.4 1 0.48 0.43 0.64
0.25 1 -0.11 0.27 0.12 0.14 0.09
-0.4 -0.11 1 -0.41 -0.2 -0.17 -0.27
1 0.27 -0.41 1 0.49 0.44 0.63
0.48 0.12 -0.2 0.49 1 0.055 0.1
0.43 0.14 -0.17 0.44 0.055 1 0.081
0.64 0.09 -0.27 0.63 0.1 0.081 1
tive_power
tive_power
Voltage
al_intensity
metering_1
metering_2
metering_3
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4


Original Power Consumption
10
8
6
4
2
0
2007-01
2007-07
2008-01
2008-07
2009-01
2009-07
2010-01
2010-07
2011-01
Normalized Power Consumption
1.0
0.8
0.6
0.4
0.2
0.0
2007-01
2007-07
2008-01
2008-07
2009-01
2009-07
2010-01
2010-07
2011-01


Feature Correlation with Target
Global_active_power
active_power_scaled
Global_intensity
Rolling_avg_24h
Lag_24h
Sub_metering_3
Rolling_std_24h
Sub_metering_1
Sub_metering_2
Hour
obal_reactive_power
Lag_168h
Day_of_week
Month
DateTime
Hour_cos
Hour_sin
Voltage
1
1
1
0.86
0.66
0.64
0.58
0.48
0.43
0.28
0.25
0.23
0.066
-0.033
-0.042
-0.11
-0.19
-0.4
Global_active_power_scaled
1.0
0.8
0.6
0.4
0.2
0.0
−0.2