Power Pulse: Household Energy Usage Forecast – Project Report

**Approach:**

The approach follows a structured workflow:

1. Data Understanding and Exploration

* Load dataset and examine its structure
* Identify missing values, duplicates, and data types.

1. Data Preprocessing

* Handle missing values using appropriate strategies.
* Handle placeholder values
* Convert data types where necessary.
* Detect Outliers using IQR and Z-score methods.
* Handle Outliers

1. Feature Engineering

* Create new features such as Hour, DayOf Week, Month, WeekOfYeat, IsWeekend, IsPeakHour, Daily\_Consumption, Unmetered\_Energy, Short\_Term\_Avg\_Power, Hourly\_Avg\_Power, Daily\_Avg\_Power, Power\_Deviation\_10min, Power\_Anomaly\_Flag, Season, TimeOfDay.

1. Exploratory Data Analysis

* Analyse trends, seasonal patterns, outliers, skewness among features.
* Visualize correlations among variables

1. Model Selection and Training

* Train Multiple regression models on 0.1% of sample extracted from the full dataset.

1. Evaluation

* Evaluate and Compare models using MAE, MSE, R2 Score, RMSE.

**Data Analysis:**

The dataset consists of energy consumption measurements taken from households over a 4-year period. The preprocessing steps performed include

1. **Preprocessing Steps**

✔ **Handled missing values** by dropping the null values rows.  
✔ **Converted date and time columns** into a single datetime format.  
✔ **Dropped duplicate rows** to ensure data integrity.  
✔ **Identified outliers** using IQR-based filtering, z-score and visualize outlier using boxplot.  
✔ **Performed skewness & kurtosis analysis** to detect data distribution anomalies.

1. **Feature Engineering**

✔ **Created new features** such as:

* Extracted meaningful features to improve model performance**: Hour, DayOfWeek, Month, IsWeekend, IsPeakHour, Daily\_Consumption, Unmetered\_Energy, Short\_Term\_Avg\_Power, Power\_Deviation\_10min, etc.**

✔ **Extracted seasonal trends** to detect cyclic energy usage patterns.  
✔ **Applied correlation** to identify significant relationships between features.

**Model Selection & Evaluation**

Multiple regression models were trained and evaluated using **2 lakh samples** (extracted from the full dataset of 20+ lakh records).

**1) Models Used**

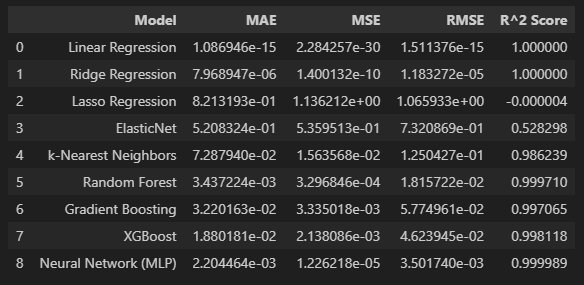
* **Linear Regression**
* **Ridge Regression**
* **Lasso Regression**
* **ElasticNet**
* **K-Nearest Neighbors (KNN)**
* **Random Forest**
* **Gradient Boosting**
* **XGBoost**
* **Neural Network (MLP Regressor)**

**2) Evaluation Metrics**

The models were evaluated based on:

* **Mean Absolute Error (MAE)**
* **Mean Squared Error (MSE)**
* **Root Mean Squared Error (RMSE)**
* **R² Score**

**3) Model Performance Summary**



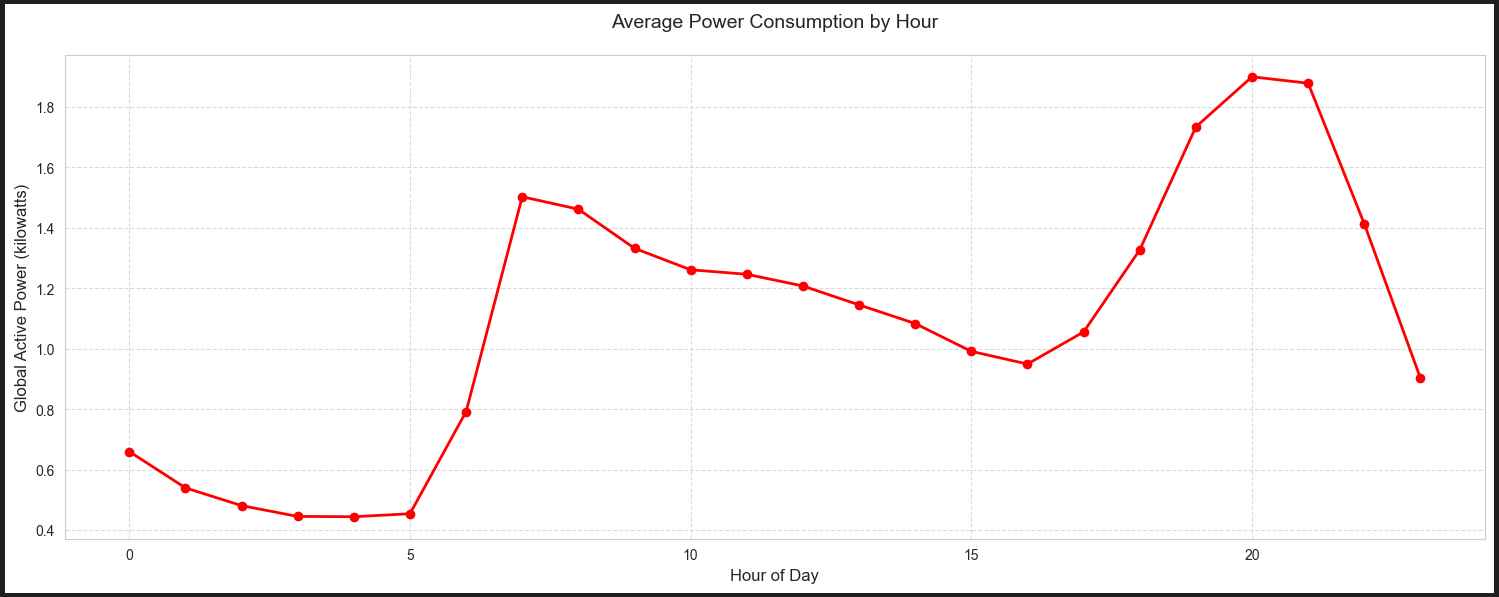
**Best Model Analysis:**

1. **Neural Network (MLP)**
   * **MAE:** 2.18e-03
   * **MSE:** 1.31e-05
   * **RMSE:** 3.61e-03
   * **R² Score:** 0.99998  
     **Why it stands out:** Neural Network (MLP) achieves extremely low error metrics and almost perfect R² score, indicating a highly accurate prediction model.

**Insights**

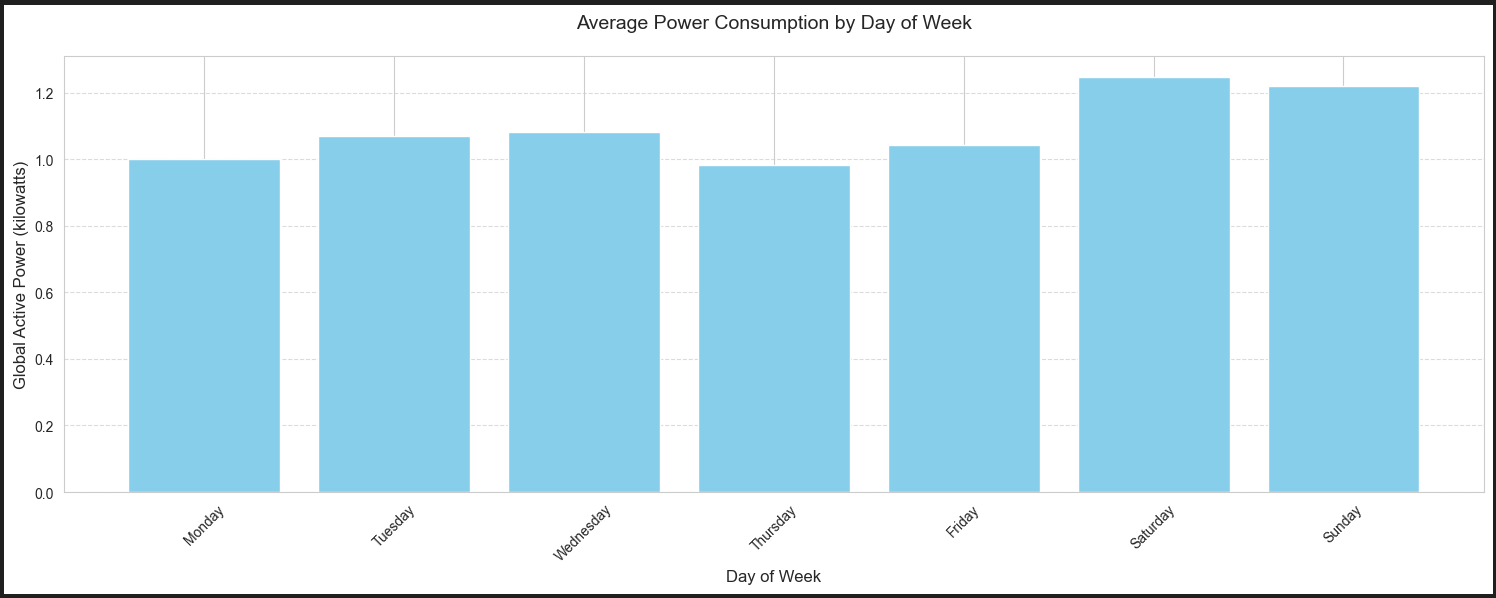
1. Daily Energy Consumption pattern.

**Morning and evening hours show peak energy consumption**, indicating household activities like cooking and heating.



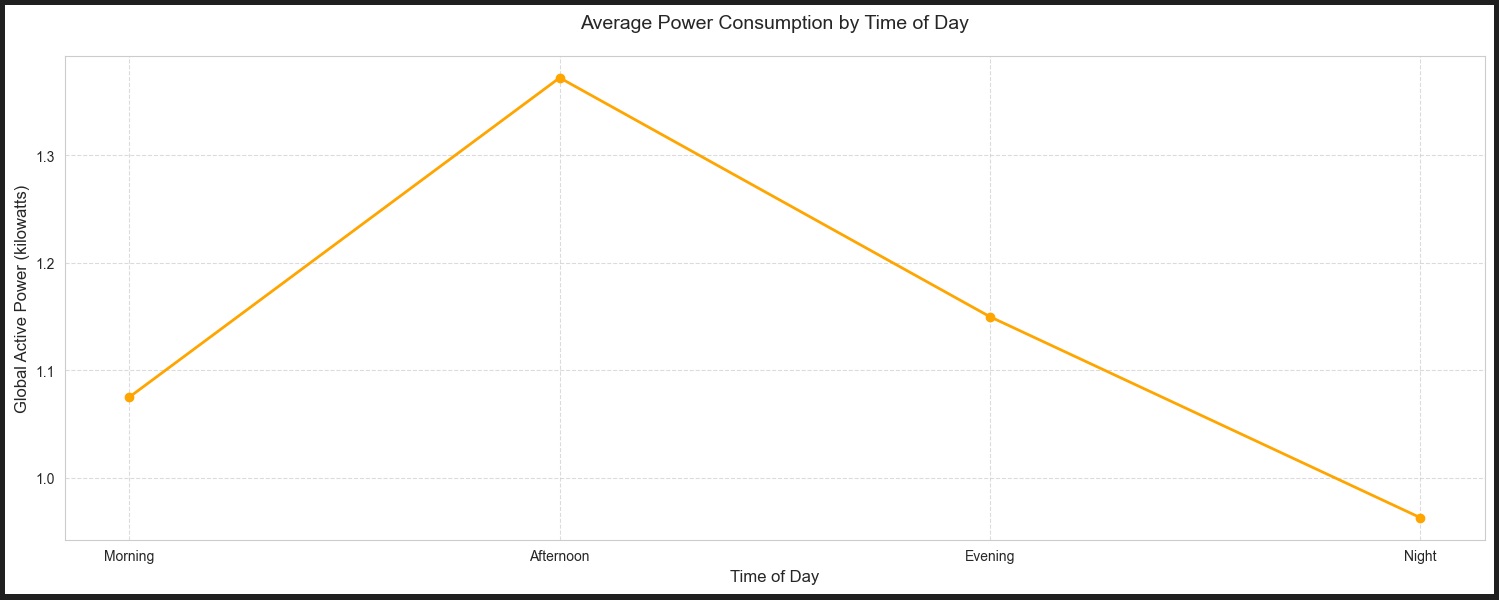
1. Weekly Consumption pattern

**Weekdays vs. Weekends:** Weekdays have more consistent consumption patterns, while weekends show a slight rise in usage, reflecting leisure activities at home.



1. Time of day consumption pattern

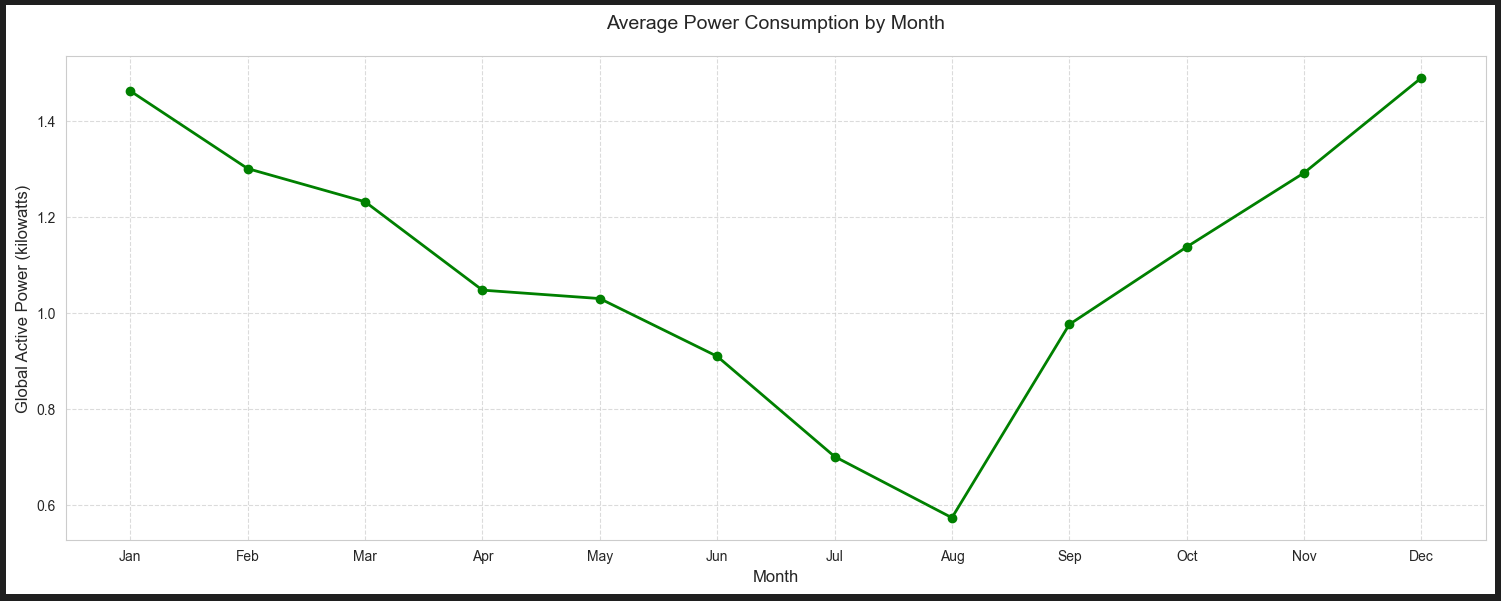
**Peak consumption hours** are Afternoon and evening, suggesting high activity during these periods.



1. Monthly Consumption pattern

**January and December** show higher energy consumption, likely due to heating and cooling needs.

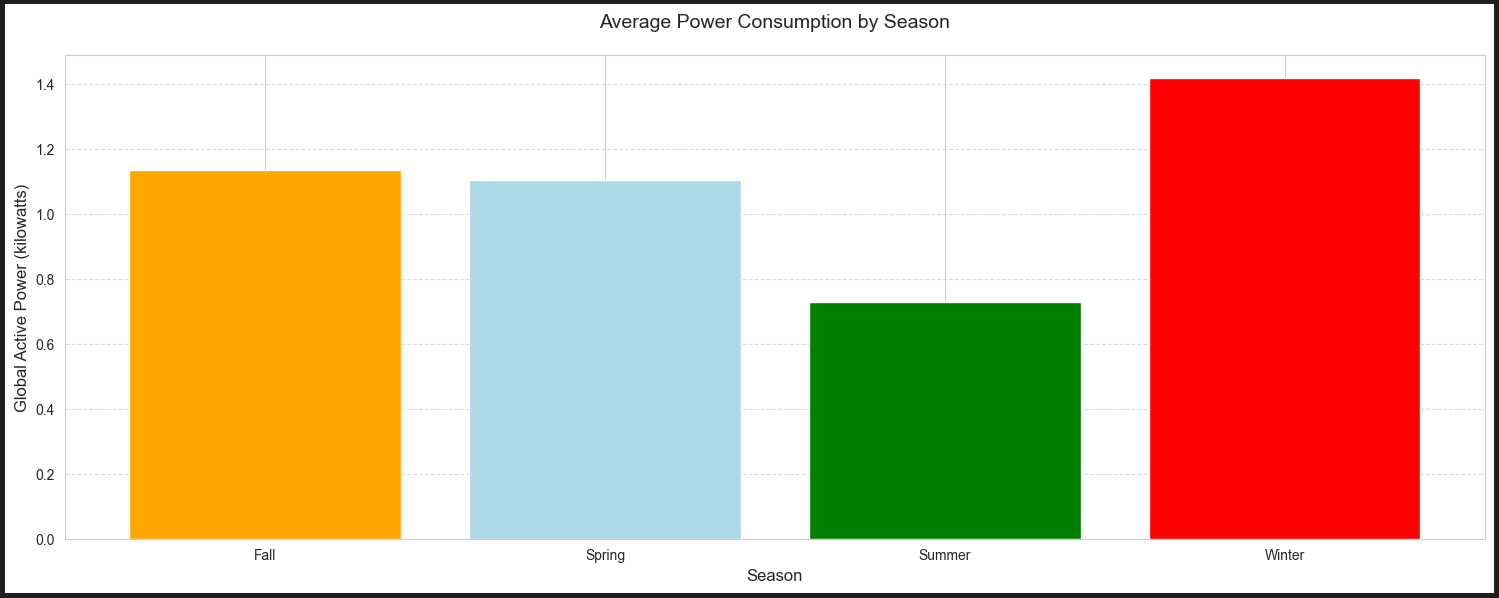
**August months** have increased consumption due to climate since it was a rainy period.



1. Seasonal Consumption pattern

**Winter months** show higher energy consumption, likely due to heating needs.

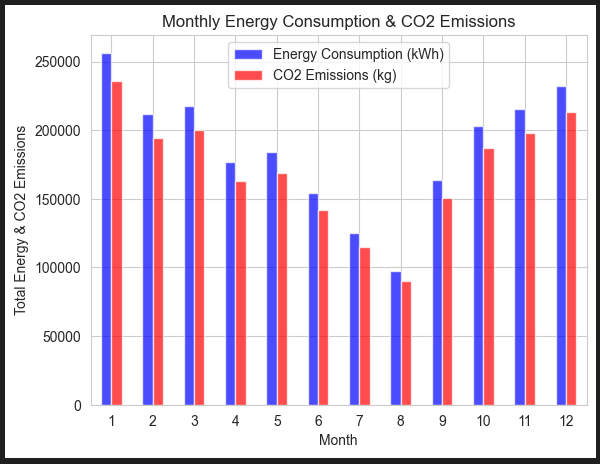
**Summer months** have increased consumption due to cooling devices like air conditioners.



1. Carbon emission analysis

Carbon emissions closely correlate with energy usage patterns.

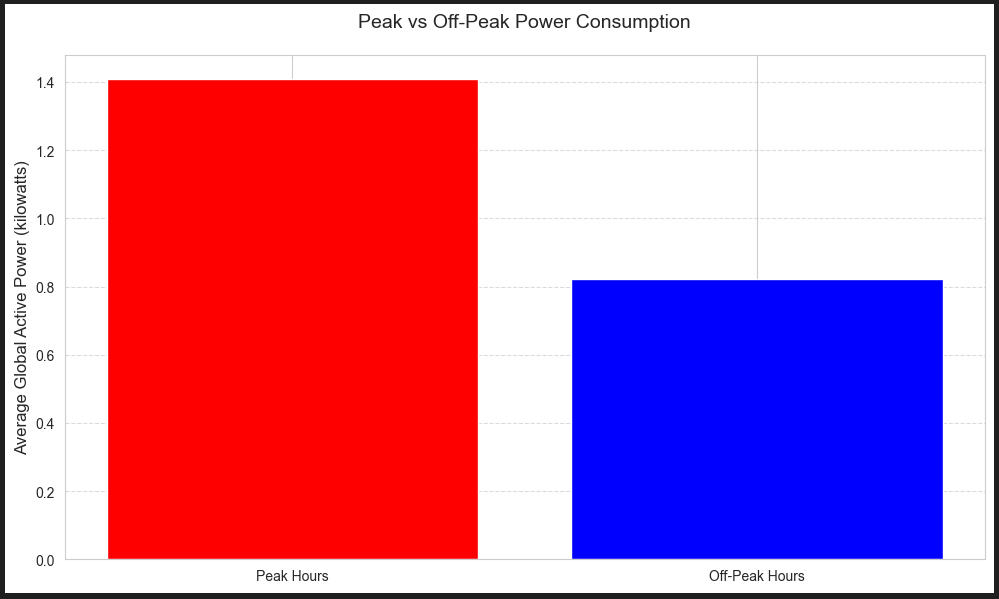
**Reducing consumption during peak hours** could significantly lower emissions.



1. Peak vs off peak power consumption

**Peak hours:** Higher demand results in increased energy costs and carbon footprints.

**Off-peak hours:** Opportunities for cost-saving and promoting energy-efficient behaviour.

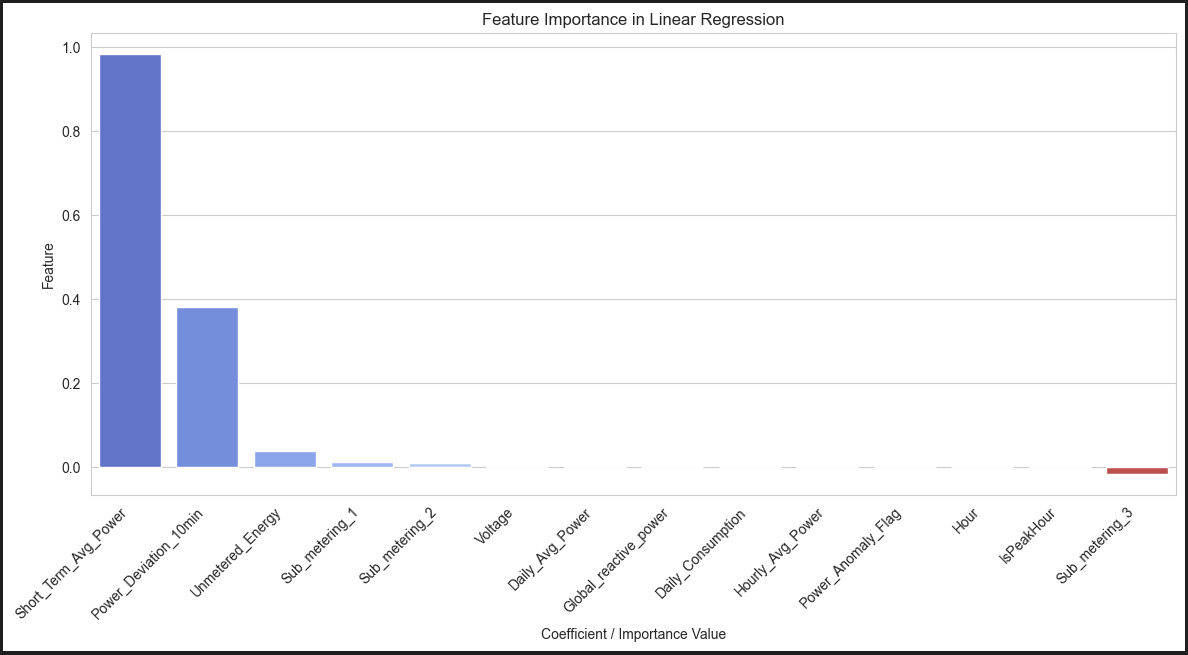


**Recommendation**

* **Shift non-essential activities to off-peak hours** to reduce costs.
* Encourage usage of **energy-efficient appliances**.
* Promote **renewable energy sources (solar, wind)** to reduce carbon footprints.

**Feature Importance of Models:**

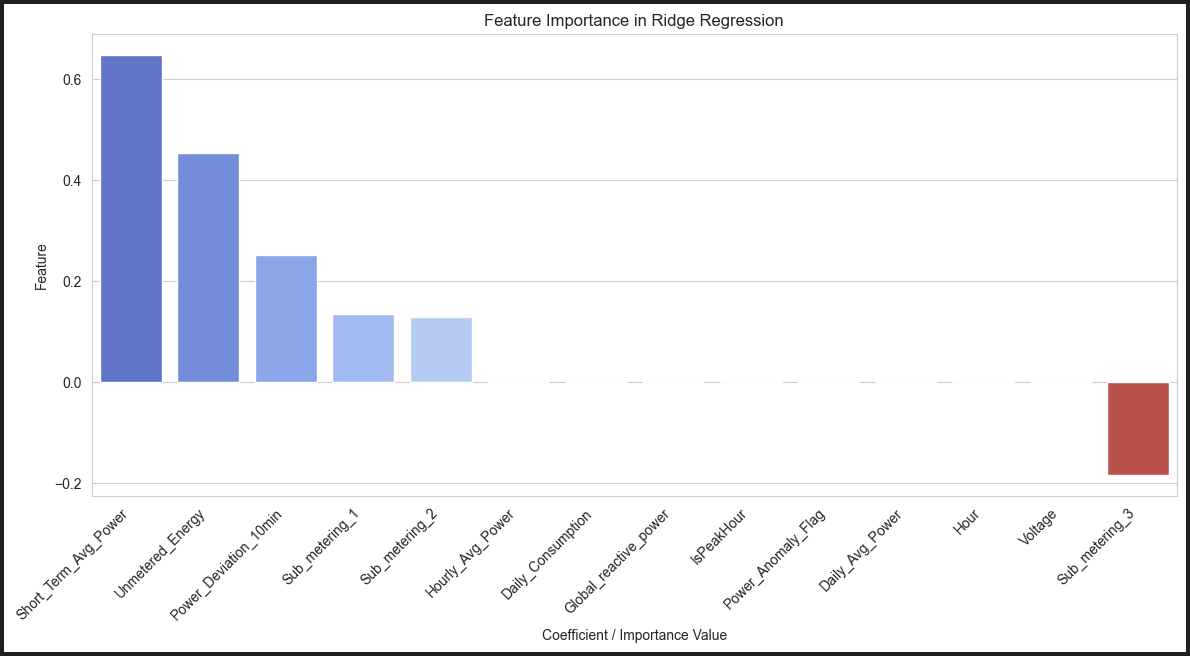
1. Linear Regression:



Linear Regression:

MAE=1.0869464266390052e-15, MSE=2.284256818212808e-30, RMSE=1.5113758031055043e-15, R^2=1.0

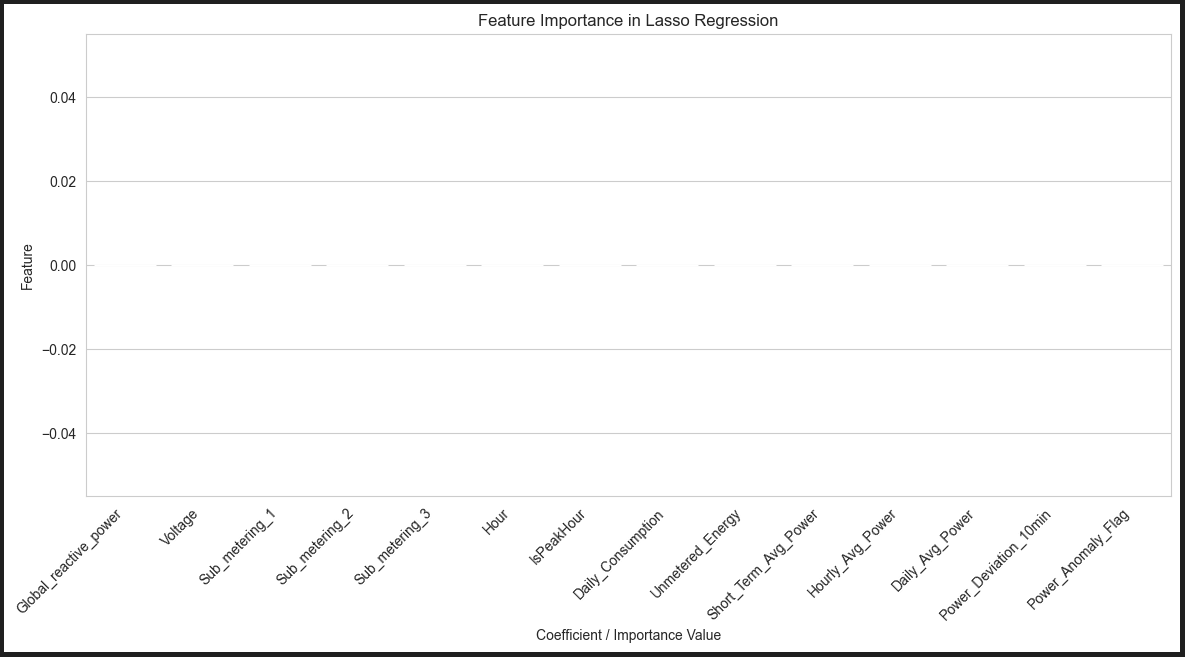
1. Ridge Regression



Ridge Regression:

MAE=7.968947065853821e-06, MSE=1.4001316815376934e-10, RMSE=1.1832716009174281e-05, R^2=0.9999999998767715

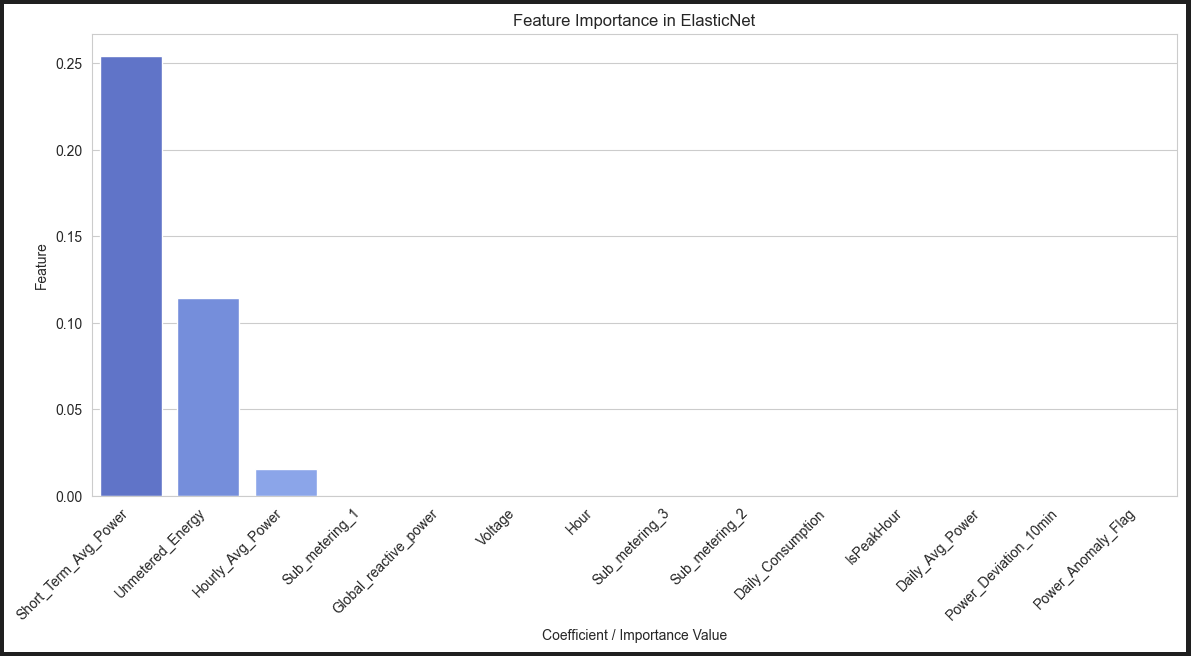
1. Lasso Regression



Lasso Regression:

MAE=0.8213193025926596, MSE=1.1362121745402154, RMSE=1.0659325375183062, R^2=-4.120625110459741e-06

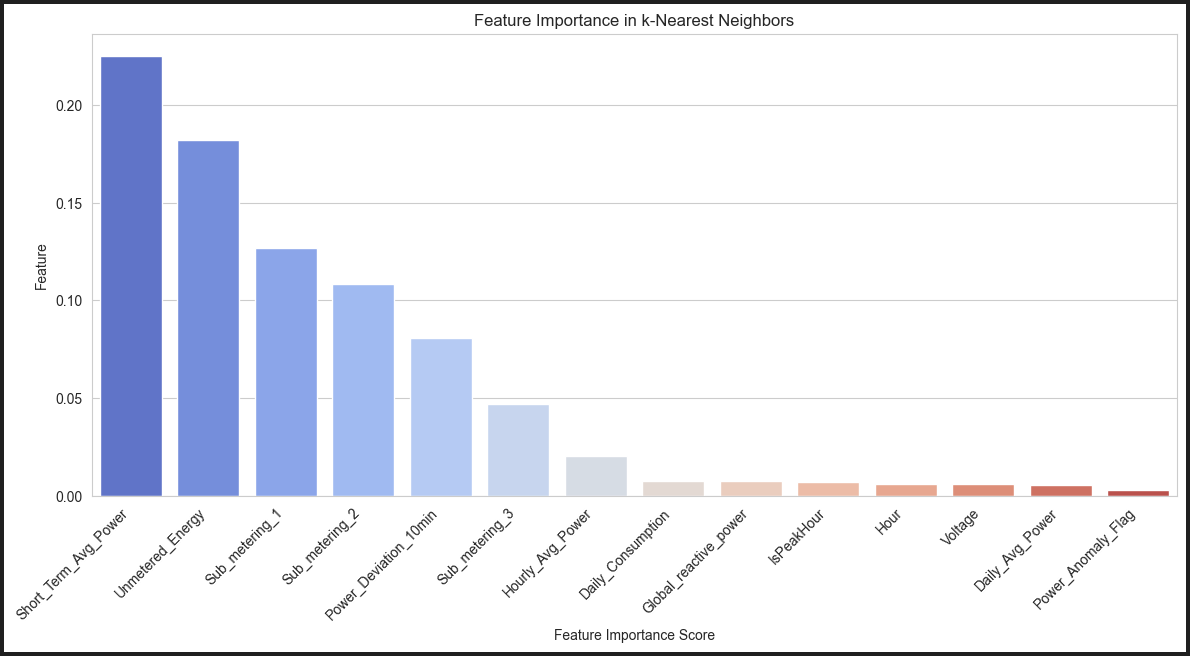
1. ElasticNet



ElasticNet:

MAE=0.5208324435182304, MSE=0.535951301385942, RMSE=0.7320869493345323, R^2=0.5282980398821956

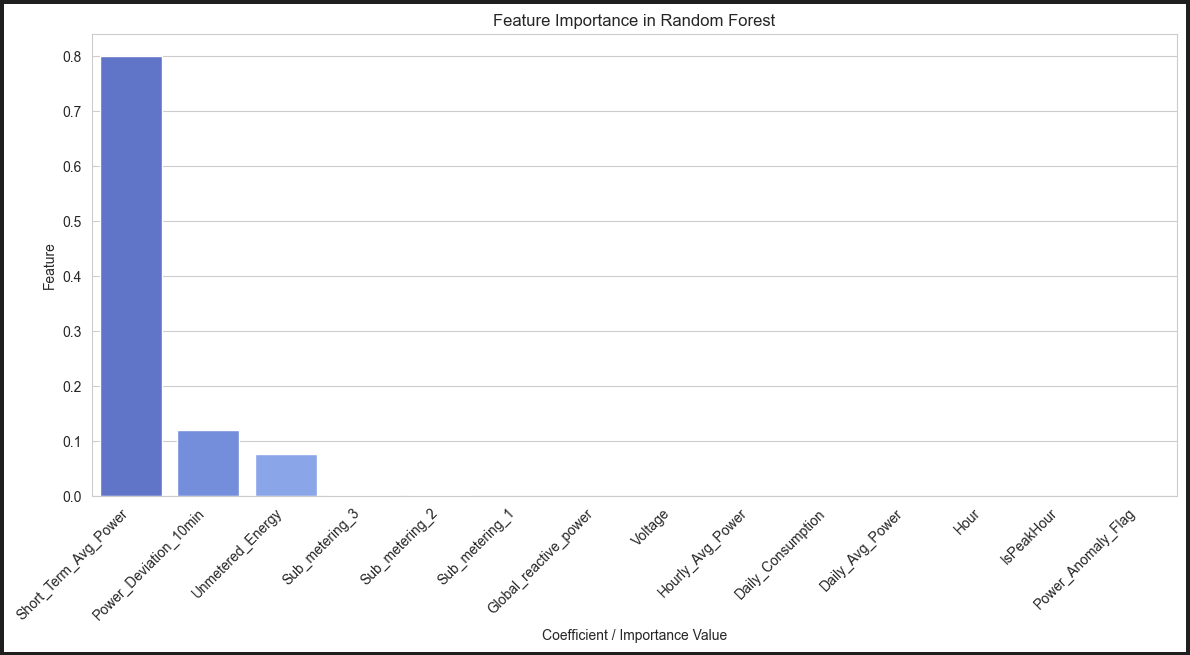
1. K-Nearest Neighbors



k-Nearest Neighbours:

MAE=0.07287940272288097, MSE=0.01563567682232958, RMSE=0.12504269999615963, R^2=0.9862387135066394

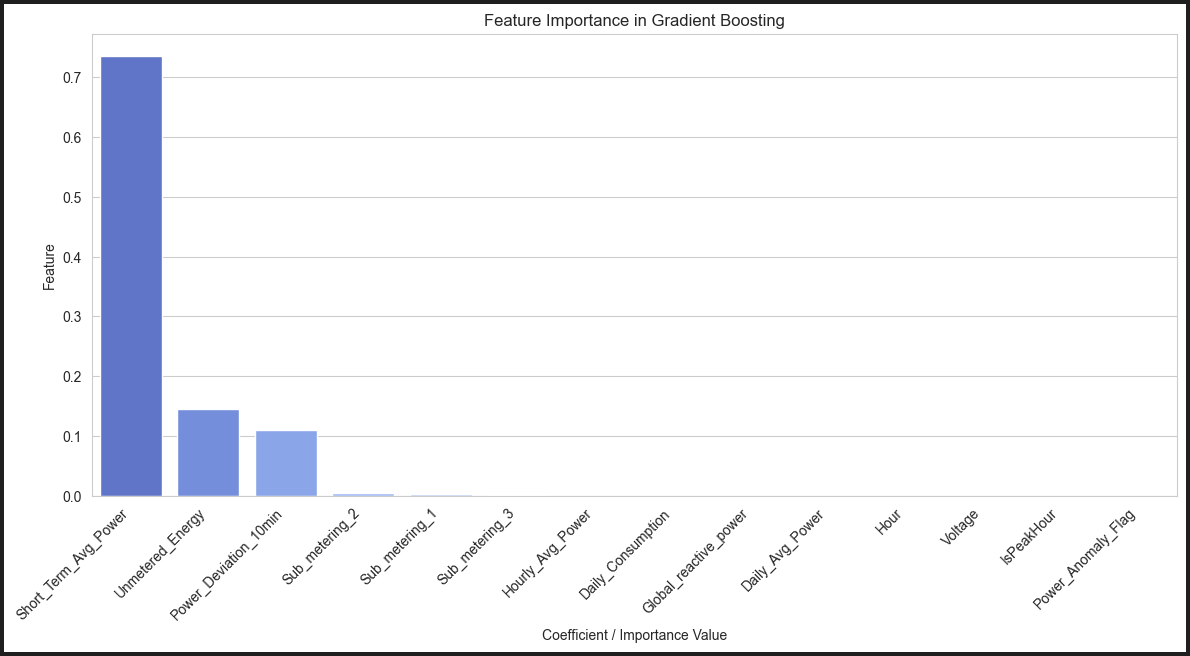
1. Random Forest



Random Forest:

MAE=0.003437224418094194, MSE=0.0003296845941931396, RMSE=0.018157218790143485, R^2=0.9997098376869328

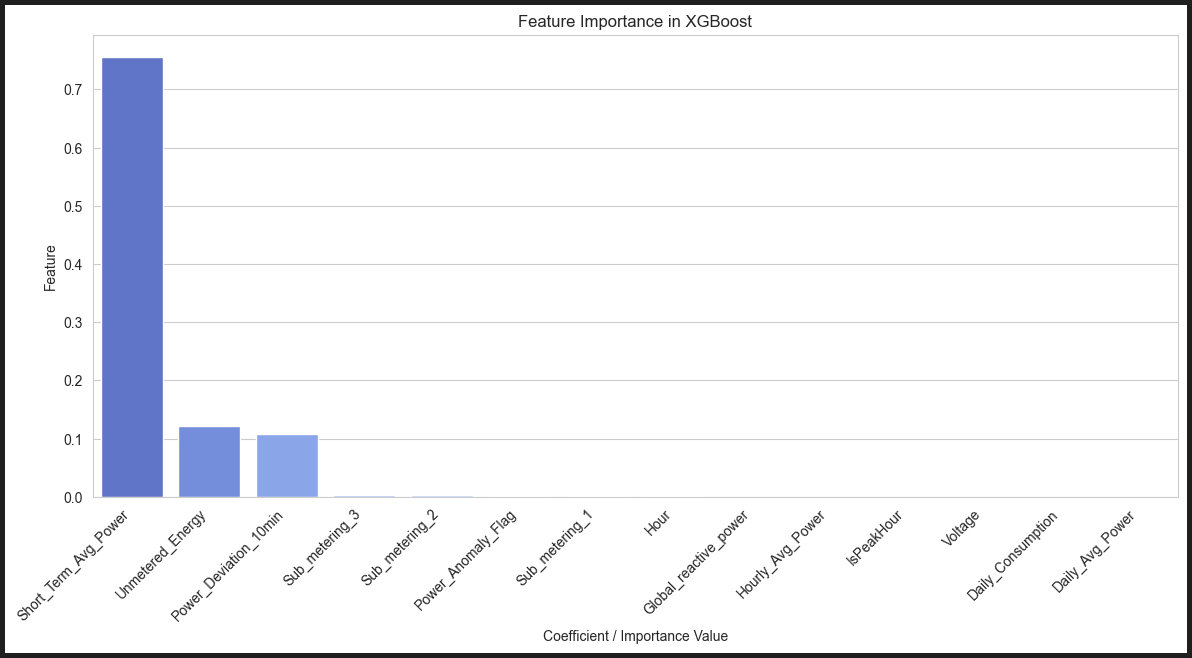
1. Gradient Boosting



Gradient Boosting:

MAE=0.032201633484048885, MSE=0.0033350178021691627, RMSE=0.05774961300449694, R^2=0.9970647810160309

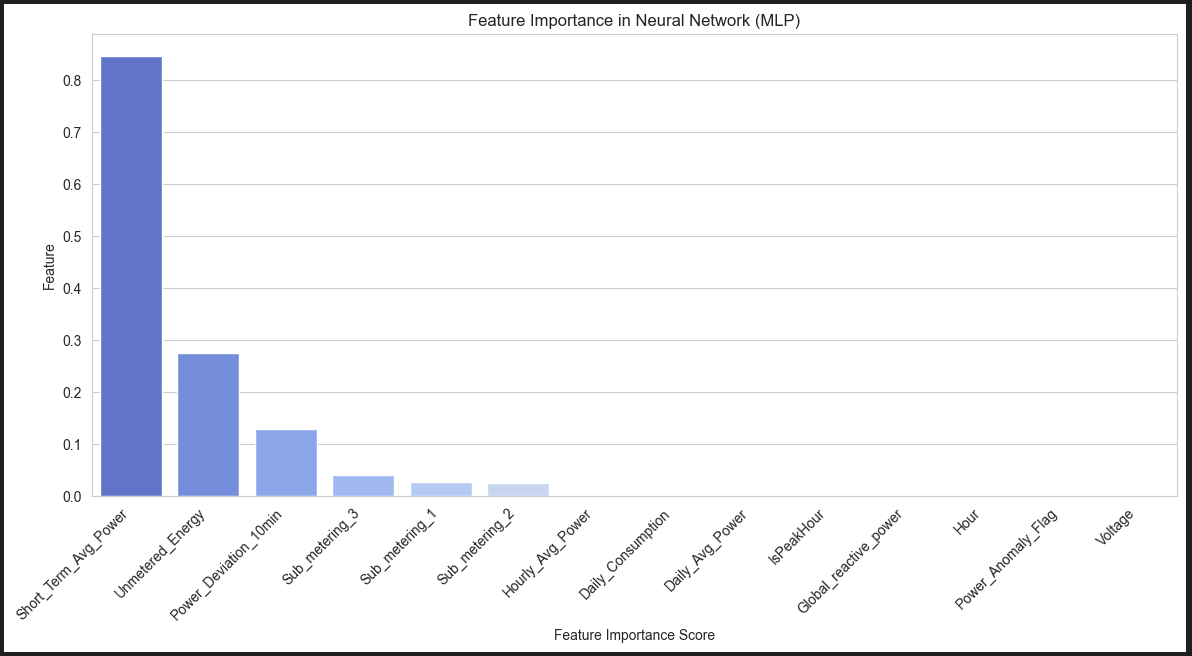
1. XGBoost



XGBoost:

MAE=0.01880180635988256, MSE=0.002138086498257829, RMSE=0.04623944742595686, R^2=0.9981182253135283

1. Neural Network



Neural Network (MLP):

MAE=0.0022044637254419047, MSE=1.2262183020581657e-05, RMSE=0.0035017399989978777, R^2=0.9999892077960233

**Model Performance Comparison:**

