Equipment Energy Consumption Prediction Report

Date: May 09, 2025

1. Problem Statement

The goal of this project is to build a machine learning model that accurately predicts equipment energy consumption based on various sensor and contextual inputs. The dataset contains numerical and time-series features, as well as a few categorical variables. Two additional variables (`random_variable1`, `random_variable2`) were evaluated for usefulness.

2. Approach

- **Exploratory Data Analysis (EDA):** Performed using visualizations such as histograms, boxplots, and correlation heatmaps.
- **Data Preprocessing:** Included handling missing values, converting timestamps, encoding categorical variables, and scaling features.
- **Feature Selection:** Random variables were analyzed for correlation and predictive importance. Unimportant features were dropped.
- **Model Development:** Several models were tested including Random Forest, Ridge Regression, and Gradient Boosting with hyperparameter tuning using GridSearchCV and cross-validation.
- **Evaluation Metrics:** RMSE, MAE, and R^2 score were used for performance evaluation.

3. Key Insights

- Some features exhibited high correlation with equipment energy consumption.
- A few columns had missing values, which were imputed using median or dropped based on percentage missing.
- The dataset was found to be moderately complex, requiring regularization to avoid overfitting.
- Random variables provided no predictive power and were excluded after RFECV evaluation.

4. Model Performance

Best Model Achieved:

- MAE : ~0.2584
- RMSE : ~0.0854

This suggests the model has room for improvement but captures some meaningful patterns.

5. Recommendations

- **Reduce Peak Usage:** Encourage operating during non-peak hours to reduce load.
- **Feature Monitoring:** Regularly analyze top influencing features (temperature, humidity, load sensor values) for optimization.
- **Data Quality:** Improve logging for timestamp data and fix missing value issues in real-time.
- **Predictive Maintenance:** Use the model to pre-emptively detect spikes in energy consumption, signaling need for maintenance.

6. Future Work

- Incorporate more contextual data such as external temperature, weather, or production schedules.
- Use deep learning models if data volume increases significantly.
- Deploy model to monitor energy usage in real-time.
