Smart Factory Energy Prediction

Rajan Aryan R.
Mechademy

May 8, 2025

1. Introduction and Approach

Understanding and optimizing equipment energy consumption is critical for reducing operational costs and environmental impact. In this project, we leveraged time-series data to build a predictive model that estimates energy consumption using historical patterns, operational variables, and temporal indicators.

Our approach followed a structured pipeline:

1. **Data Loading and Cleaning:**
   - Imported timestamped sensor data from CSV.
   - Parsed datetime features and handled missing or corrupted entries.
   - Removed uninformative variables and encoded categorical fields.
2. **Feature Engineering:**
   - Extracted features such as hour of the day, day of week, month, and weekend indicators.
   - Created lag features (`lag1_energy`, `lag2_energy`) and rolling averages to capture temporal dependencies.
3. **Modeling:**
   - Scaled data using `StandardScaler` and split into training and test sets without shuffling (preserving time order).
   - Tuned a `Random Forest Regressor` using `TimeSeriesSplit` and `GridSearchCV`.
   - Developed a `Stacking Regressor` with Random Forest, XGBoost, and Gradient Boosting as base models, and Ridge regression as meta-model.
4. **Evaluation and Visualization:**
   - Evaluated the model using RMSE, MAE, and R^2 metrics.
   - Generated feature importance and prediction vs. actual plots for interpretation.

2. Data Insights

From the processed dataset, several patterns emerged:

2.1 Feature Importance

The Random Forest model revealed the most impactful predictors:

- **Lag features:** Previous energy values (`lag1_energy`, `lag2_energy`) were highly predictive, indicating strong autocorrelation.
- **Rolling mean:** The 3-hour rolling average helped smooth short-term fluctuations.
- **Temporal variables:** Hour of day and day of week played critical roles in characterizing operational cycles.

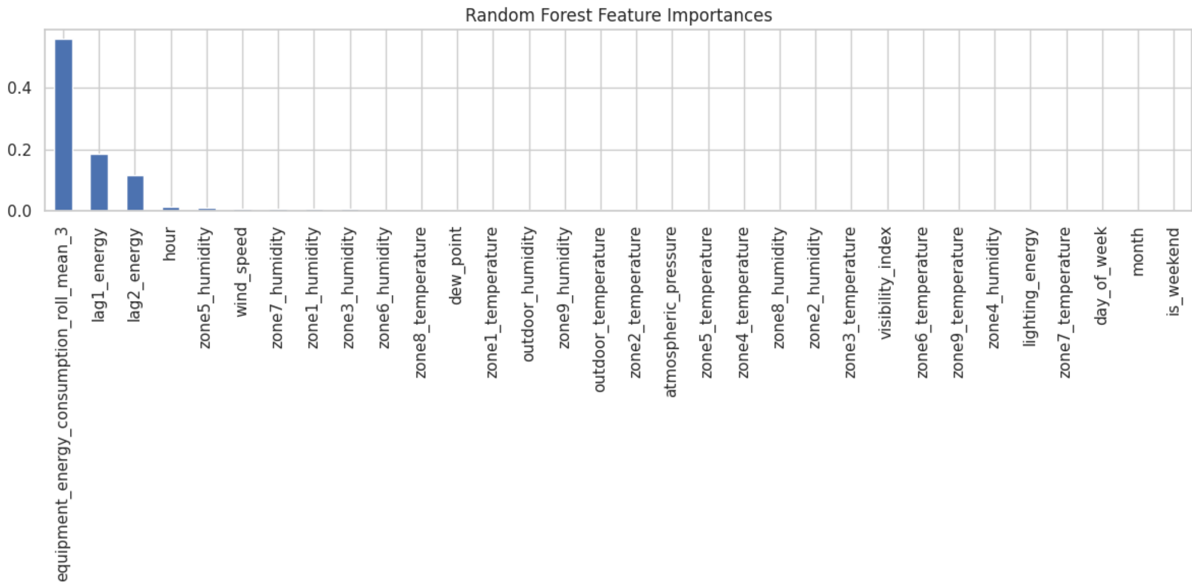

Figure 1: Feature Importance from Random Forest Regressor

2.2 Temporal Influence

- Energy consumption showed clear trends tied to the **hour of the day**, with peak usage typically observed during business hours.
- **Weekday vs. weekend behavior** differed significantly, with weekends showing lower energy use.
- Monthly trends suggested **seasonality**, possibly due to climate-driven operations or maintenance schedules.

3. Model Performance Evaluation

We evaluated the ensemble model on a hold-out test set using the following metrics:

- **Root Mean Squared Error (RMSE)**: `86.39`
- **Mean Absolute Error (MAE)**: `28.49`
- R^2 **Score**: `0.768`

These results indicate that the model captures the variance in energy consumption well, with minimal overfitting due to the use of cross-validation and model ensembling.

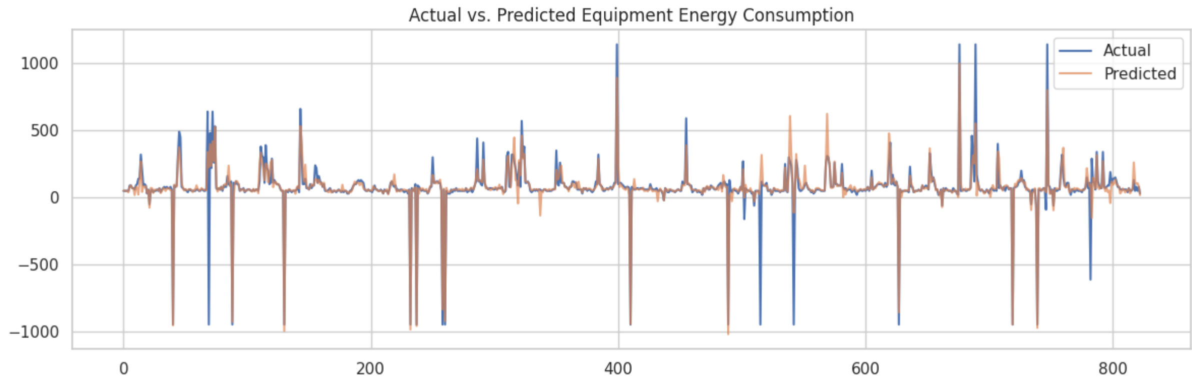

Figure 2: Predicted vs Actual Equipment Energy Consumption

4. Recommendations for Reducing Energy Consumption

Based on the data analysis and model interpretability, we propose the following operational strategies:

1. **Optimize Scheduling:** Shift energy-intensive operations to hours with lower historical consumption to balance load.
2. **Enable Predictive Maintenance:** Monitor spikes in lagged energy use, which may signal malfunctioning equipment or inefficiencies.
3. **Reduce Weekend Usage:** Since weekend patterns differ significantly, consider shutting down or reducing operations during those periods.
4. **Leverage Forecasting:** Integrate the model to generate next-day/hour energy forecasts and use them in planning and budgeting tools.
5. **Implement Seasonal Adjustments:** Adjust baseline consumption targets based on observed monthly trends.

5. Conclusion

This project demonstrated the effective application of ensemble machine learning methods to predict equipment energy consumption using time-series data. The analysis highlighted the value of temporal features and lagged indicators in understanding operational behavior. By interpreting the model and trends, actionable recommendations were derived to reduce energy usage, lower costs, and enhance efficiency.