Smart Factory Energy Consumption Prediction Report

Problem Overview

The goal of this project was to build a machine learning model that can predict the energy consumption of industrial equipment in a smart factory setting. This prediction helps factory managers make better decisions to reduce energy costs and improve operational efficiency.

Approach

1. **Data Understanding**
   We worked with sensor data collected from 9 zones in a factory, outdoor weather readings and timestamped energy measurements.
2. **Data Preprocessing**
   - Converted timestamps into features like hour, day and month.
   - Handled missing values using column-wise averages.
   - Applied domain-based filtering to remove invalid or extreme values.
   - Log-transformed the target variable (equipment_energy_consumption) for stability.
3. **Feature Engineering**
   - Created zone-wise summary statistics: mean, min, max, std for temperature and humidity.
   - Added derived features like temp-humidity index and cyclic encodings for hour of day.
   - Included lighting energy and outdoor weather data.
4. **Modeling**
   - Used a RandomForestRegressor wrapped in a pipeline with StandardScaler.
   - Performed 5-fold cross-validation and hyperparameter tuning using GridSearchCV.

Key Insights

- Zone conditions (especially average zone temperature and humidity) were among the most influential features.
- Lighting energy and outdoor temperature also impacted equipment energy use.
- The two random_variable columns were highly similar in distribution. After testing both, we chose to use random_variable1, which slightly improved the model's R^2 score. This suggests that while the variable is synthetic, it carries some weak predictive signal and may help with regularization or noise balancing.

Model Performance

Metric	Cross-Validation (log scale)	Test Set (original scale)
RMSE	~0.57 (mean)	~113.57 Wh
MAE	~0.40 (mean)	~50.94 Wh
R^2 Score	~0.34 (mean)	~0.25

Note: RMSE and MAE values on the test set were obtained after reversing the log transformation.

Recommendations

- Monitor zone-level temperature, lighting energy and humidity closely—they are strong indicators of energy consumption.
- Use this model to build a dashboard for real-time energy forecasting and alerts.

Limitations

- Time-series modeling (e.g., LSTM) could potentially improve predictions.

Conclusion

This project successfully delivered a working predictive model that estimates energy consumption using available sensor data. It can serve as a valuable tool for optimizing energy use in smart manufacturing environments.