Energy Consumption Prediction Report – SmartManufacture Inc.

1. Approach to the Problem

We developed a machine learning pipeline to predict equipment energy consumption using sensor and environmental data. The workflow involved:

- Loading and preprocessing the dataset
- Conducting exploratory data analysis (EDA)
- Applying feature scaling
- Training and evaluating both Linear Regression and Random Forest models
- Extracting feature importance for actionable insights

---

2. Key Insights from the Data

- The target variable equipment_energy_consumption showed a right-skewed distribution, which is typical for consumption data.
- Correlation analysis revealed strong relationships between energy consumption and environmental factors such as zone temperature, humidity, and lighting energy usage.
- Feature importance from the Random Forest model confirmed these observations, highlighting climate control and lighting as dominant predictors.

---

3. Model Performance Evaluation

Two models were trained and tested on an 80/20 train-test split:

- Linear Regression served as a baseline.

- Random Forest Regressor significantly outperformed the linear model.

📊 Sample Metrics (example placeholders):

- Linear Regression:
  - MAE: 74.870
  - RMSE: 163.275
  - R^2: 0.0069
- Random Forest:
  - MAE: 68.9103
  - RMSE: 159.085
  - R^2: 0.0572

---

4. Recommendations for Reducing Equipment Energy Consumption

Based on model insights and feature importance:

- Optimize climate control in zones with high temperature/humidity impact.
- Reduce unnecessary lighting load, especially in low-traffic or non-critical areas.
- Consider removing or deprioritizing less relevant sensors/features to streamline data collection and model simplicity.
- Implement predictive maintenance strategies using model outputs to identify inefficient conditions early.