Smart Factory Energy Prediction - Final Report

1. Project Approach

Objective: Develop a predictive model for equipment energy consumption using historical sensor and operational data.

Methodology

1. Exploratory Data Analysis (EDA)
    - Analyzed data distributions, trends, and correlations.
    - Identified non-linear relationships between variables (e.g., temperature/humidity vs. energy use).
2. Data Preprocessing
    - Missing/Invalid Data: Retained all data points using time-series-friendly methods:
        - Forward/backward filling (superior to median imputation for temporal data).
    - Outlier Handling: Applied IQR with a factor of 9 to minimize data loss.
3. Feature Engineering
    - Extracted time-based features: hour, month, weekend from timestamps.
    - Removed irrelevant features (*random_variable1/2*) with low correlation to target (≤0.47).
4. Model Selection
    - Candidate Models: Random Forest, XGBoost, Multiple Linear Regression (discarded due to non-linearity).
    - Final Model: Random Forest (outperformed XGBoost in cross-validation).

---

2. Key Insights

Data Patterns

- Energy Consumption Peaks:
    - High in extreme temperature/humidity zones (see *Scatter Plot*).
    - Non-linear relationship with sensor data (confirmed via heatmaps).

Feature Impact

- Critical Features: zone_temperature, humidity, hour_of_day.
- Discarded Features: Low-correlation variables to reduce noise.

---

3. Model Performance

Metric	Score	Interpretation
R^2	0.56	Explains 56% of energy variance
RMSE	59.66	Moderate prediction error
MAE	30.30	Average absolute error is low

Performance is acceptable given sensor data noise.

---

4. Energy-Saving Recommendations

Operational Adjustments

- Maintenance Scheduling: Prioritize low-energy hours (per model predictions).
- Climate Control: Pre-emptively adjust equipment based on temperature/humidity forecasts.

Infrastructure Improvements

- Insulation/Ventilation: Target high-energy zones to stabilize ambient conditions.