# Smart Factory Energy Consumption Prediction - Analysis Report

## 1. Overview

This report analyzes the energy consumption of equipment in a smart factory using machine learning models. The goal was to:

• Clean and preprocess sensor data.

• Identify key factors affecting energy consumption.

• Build and evaluate predictive models to forecast energy usage.

• Provide recommendations for reducing energy consumption.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## 2. Data Preprocessing & Cleaning

✔ Handled Missing Values

• Used time-based interpolation (best for time-series data).

• Compared forward-fill, linear interpolation, and time-based methods (time-based was most effective).

✔ Removed Invalid Data

• Negative energy values (replaced with NaN and interpolated).

• Outliers capped using IQR method (upper bound = Q3 + 1.5\*IQR).

✔ Feature Engineering

• Created temperature differentials (zoneX\_temp\_diff = zoneX\_temperature - outdoor\_temperature).

• Added rolling window statistics (4-hour mean & std for energy and temperature).

✔ Feature Selection

• Removed irrelevant features (random\_variable1, random\_variable2).

• Selected best predictors using:

Correlation analysis (> 0.1 with target), Random Forest feature importance, Statistical tests (SelectKBest with f\_regression).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## 3. Exploratory Data Analysis (EDA)

Key Findings:

📌 Energy Consumption Distribution

• Initially had negative values (sensor errors) and extreme outliers (up to 12,000 kWh).

• After cleaning: bimodal distribution (peaks near 0 and 500-1000 kWh).

📌 Correlation Insights

• Zone temperatures and outdoor temperature strongly influence energy use.

• Rolling mean/std features improved model performance (captured temporal trends).

📌 Time-Series Patterns

• Missing data was best handled with time-based interpolation.

• 4-hour rolling windows helped smooth fluctuations.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## 4. Model Performance Comparison

|  |  |  |  |
| --- | --- | --- | --- |
| Model | RMSE | MAE | R² |
| Gradient Boosting | 19.48 | 12.30 | 0.77 |
| Linear Regression | 19.87 | 12.56 | 0.76 |
| Ridge Regression | 19.87 | 12.56 | 0.76 |
| Lasso Regression | 19.87 | 12.55 | 0.76 |
| Random Forest | 20.05 | 12.47 | 0.76 |
| XGBoost | 21.28 | 12.83 | 0.73 |

### Best Model: Gradient Boosting Regressor

### ✅ Hyperparameters:

• learning\_rate=0.1

• max\_depth=3

• n\_estimators=50

### ✅ Performance:

• R² = 0.77 (77% variance explained).

• RMSE = 19.48 (lower error than other models).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## 5. Recommendations for Energy Reduction

1. Optimize Temperature Control

• Reduce temperature differentials between zones and outdoors.

• Use dynamic HVAC scheduling based on weather forecasts.

2. Improve Equipment Efficiency

• Monitor high-consumption periods (rolling mean helps detect anomalies).

• Upgrade inefficient machinery causing extreme energy spikes.

3. Predictive Maintenance

• Detect sensor malfunctions (negative values indicate errors).

• Use model predictions to optimize energy usage in real-time.

4. Future Enhancements

• Add more sensor data (equipment status, production load).

• Deploy real-time monitoring with alerts for abnormal consumption.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## 6. Conclusion

The Gradient Boosting model (R²=0.77) provides reliable predictions for energy consumption. Key takeaways:

✔ Temperature management is crucial for reducing energy use.

✔ Time-based features improve prediction accuracy.

✔ Real-time monitoring can help optimize energy efficiency.