Smart Factory Energy Prediction Report

1. Introduction

Objective: Develop a machine learning model to predict equipment energy consumption based on sensor data.

Goals:

- Analyze sensor data to identify patterns and relationships.
- Build a robust regression model for energy consumption prediction.
- Evaluate model performance and provide actionable insights.

2. Data Understanding and Preprocessing

Dataset Overview:

- Shape: 16,857 rows and 29 columns.
- Missing Values: Initial missing values were handled using interpolation and mean imputation.
- Duplicate Values: 68 duplicates were identified and removed.

Feature Engineering:

- New Features: Created average temperature and humidity features.
- Multicollinearity Check: Used Variance Inflation Factor (VIF) to identify and remove highly correlated features.

3. Exploratory Data Analysis (EDA)

Energy Consumption Patterns:

- Distribution: Equipment energy consumption showed positive skewness.
- Monthly Trends: High energy usage in October, low in November.
- Weekly Trends: High energy usage on Wednesday and Thursday, low on Saturday.
- Hourly Trends: Peak consumption from 17:00 to 19:00.

Environmental Factors:

- Temperature and Humidity: Indoor and outdoor conditions significantly impacted energy consumption.
- Correlation Analysis: Identified strong relationships between temperature, humidity, and energy consumption.

4. Model Building and Evaluation

Models Implemented:

- Linear Regression, Ridge, Lasso: Poor performance with Test R^2 Scores around 0.146 and high Test MSE (~1813).
- Decision Tree: Moderate performance with Test R^2 Score of 0.395.
- Random Forest: Best performance with Test R^2 Score of 0.568 and Test MSE of 917.06.
- Extra Trees, XGBoost, LightGBM: Decent performance but not better than Random Forest.
- Stacking Regressor: Good performance but not superior to Random Forest.

Hyperparameter Tuning:

- Used GridSearchCV for optimizing model parameters.
- Best parameters identified for each model.

5. Key Insights and Recommendations

Optimize Energy Usage:

- Implement energy-saving measures during peak consumption hours (17:00 to 19:00).
- Schedule high-energy tasks during low-consumption periods (weekends, specific months).

Temperature and Humidity Management:

- Maintain optimal indoor climate conditions to reduce energy consumption.
- Utilize energy-efficient HVAC systems.

Model Deployment:

- Deploy the Random Forest Regressor model for real-time energy consumption prediction.
- Continuously monitor and update the model with new data.

6. Conclusion

Best Model: Random Forest Regressor with Test R^2 Score of 0.568 and Test MSE of 917.06.

Future Work: Explore advanced techniques such as deep learning for potentially better predictive performance.