**SMART FACTORY ENERGY PREDICTION CHALLENGE**

This document shows the approach taken to tackling this challenge and some key insights obtained from the data. It also provides some recommendations fro reducing equipment energy consumption.

**Overview**

This solution aims to tackle the smart factory energy prediction challenge by developing a robust machine learning model that can accurately predict the energy consumption of industrial equipment leveraging on comprehensive feature engineering and modelling techniques.

**Objectives:**

- Analyze the provided sensor data to identify patterns and relationships between environmental factors and equipment energy consumption

- Build a robust regression model to predict equipment energy consumption

- Evaluate the model's performance using appropriate metrics

- Provide actionable insights and recommendations for reducing energy consumption

### **2. ETL Process**

### **Extract**

### **- Data Set :** Equipment energy consumption at 10-minute intervals

### **- Data Formats**: CSV

### **Transform**

### **i. Data Cleaning:**

#### - Conversion of timestamps to datetime objects

#### **-** Removal of strings from certain columns and conversion to numeric values to be used for calculations

#### **ii. Aggregation:**

#### - Aggregation of 10-minute data to daily totals with multiple statistical metrics

#### **iii. Feature Engineering:**

#### - Creation of temporal features and cyclical encodings

#### - Integration of data with some temperature metrics

#### **Load**

#### - Load dataset for analysis

#### **Model Training**

Three different LightGBM configurations were trained on the data:

1. **Precise: Conservative** approach with deep trees (max\_depth=8, learning\_rate=0.01)
2. **Feature Selective** : Aggressive feature selection (colsample\_bytree=0.6, feature\_fraction=0.7)
3. **Highly Regularized** : Strong regularization (reg\_alpha=2.0, reg\_lambda=2.0)

**Evaluation & Optimization**

* **5-Fold Cross-Validation**: Applied for robust model evaluation
* **Bayesian Optimization**: Finds the best weights to combine these 3 base models

### **7. Performance Metrics**

* Root Mean Squared Error (RMSE) – 4737.6392
* Mean Squared Error (RMSE) – 3640274.5145
* R2 Score – 0.264996

### **8. Error Handling and Logging**

* Logging: Configured at the start of the notebook with logging library (INFO level).
* Error Handling: If a given base model fails in training, a placeholder LightGBM model (n\_estimators=100) is substituted, preventing pipeline interruptions.

**KEY INSIGHTS DERIVED FROM OUR DATA**

* **Equipment Energy Consumption**

- The Average daily energy consumption between January and April has a discernable and considerably consistent pattern of variation but by May, a significant drop in the energy consumption can be seen from the plot.

- The highest daily energy consumption occurs around April

* **Random Variables 1 and 2**

It can be observed that there seems to be a similarity in the behavioral pattern of these random variables like was observed with the average daily energy consumption:

- Both variables appear to have a somewhat similar trend with time.

- Consistent variation pattern between January and May and after that no discernable pattern is seen.

This could be indication that there is a relationship that exists with the target variable, we can verify from the pearson correlation coefficient if it's okay to keep them.

**ADDITIONAL INSIGHTS**

A similitude of the pattern seen with the equipment energy consumption may be noticed with some of the environmental conditions shown above even though more peculiar trends are seen (like indoor humidity general decreasing pattern from January to May and indoor temperature exhibiting an inverse behaviour to the humidity), the same idea holds that after May, that consistent trend seems lost.

Observing each of the environmental conditions closely, we can conclude that no one condition specifically stands out as the singular influence and factor for the behaviour of the energy consumption variable after May because though the consistency in trend isn't seen after May there are still significant peaks and troughs the same of which cannot be said of our equipment energy consumption variable.

**RECOMMENDATIONS FOR REDUCING EQUIPMENT ENERGY CONSUMPTION.**

Depending on the specific area of focus, there a number of recommendations that can be provided for reducing equipment energy consumption :

1. Real-time monitoring and control of energy usage can help to identify inefficiencies and optimize energy consumption across facilities.

2. Investing in high energy efficient equipment and devices can reduce energy consumption significantly. These devices are designed to reduce energy consumption.

3. Streamlining production processes and reducing idle times can lead to significant energy savings. Evaluating and optimizing workflows ensures efficient use of energy.

4. Conducting regular energy audits can identify areas where energy is being wasted and suggest improvements. Implementing recommendations from audits can lead to significant energy savings.