---

1. Introduction

Project Overview

In this project, we aimed to develop a machine learning model capable of predicting the energy consumption of industrial equipment. By utilizing sensor data collected from the factory's sensor network, the model will help facility managers optimize operations to improve energy efficiency and reduce operational costs.

Objective

The main goal of this project was to predict the energy consumption of industrial equipment (equipment_energy_consumption) based on environmental factors such as temperature, humidity, and energy usage in different zones. We sought to identify patterns and relationships within the data and use them to build a reliable predictive model.

---

2. Data Analysis

Data Overview

The dataset provided consisted of multiple features related to environmental conditions and energy usage across various zones in the factory. It included the following key columns:

- **equipment_energy_consumption**: The target variable we aimed to predict.
- **lighting_energy**: Energy consumed by lighting.
- **temperature and humidity**: Features describing the environmental conditions in different zones (zone1, zone2, etc.).
- **outdoor and atmospheric conditions**: Features such as outdoor temperature, wind speed, atmospheric pressure, and visibility index.

Exploratory Data Analysis (EDA)

Through initial analysis, it was observed that environmental factors like **temperature** and **humidity** were highly correlated with energy consumption. Additionally, the data revealed some missing values across several columns, particularly in the environmental features and random variables.

Data Preprocessing

- **Missing Data**: Several columns exhibited missing data, with the highest being in features like **equipment_energy_consumption**, **zone1_temperature**, **zone3_humidity**, and others. The missing values were replaced with the mean of each respective column to ensure no data loss.

- **Feature Engineering**: Initially, the **timestamp** column was not altered, but later it was converted to **hour** and **day of the week** to better capture time-based patterns that may influence energy consumption.

---

3. Model Development

Approach

We chose the **GradientBoostingRegressor** for predicting the energy consumption due to its effectiveness in handling complex relationships and interactions between features. This model is suitable for regression tasks with non-linear relationships and has proven to be robust in similar real-world applications.

Feature Selection

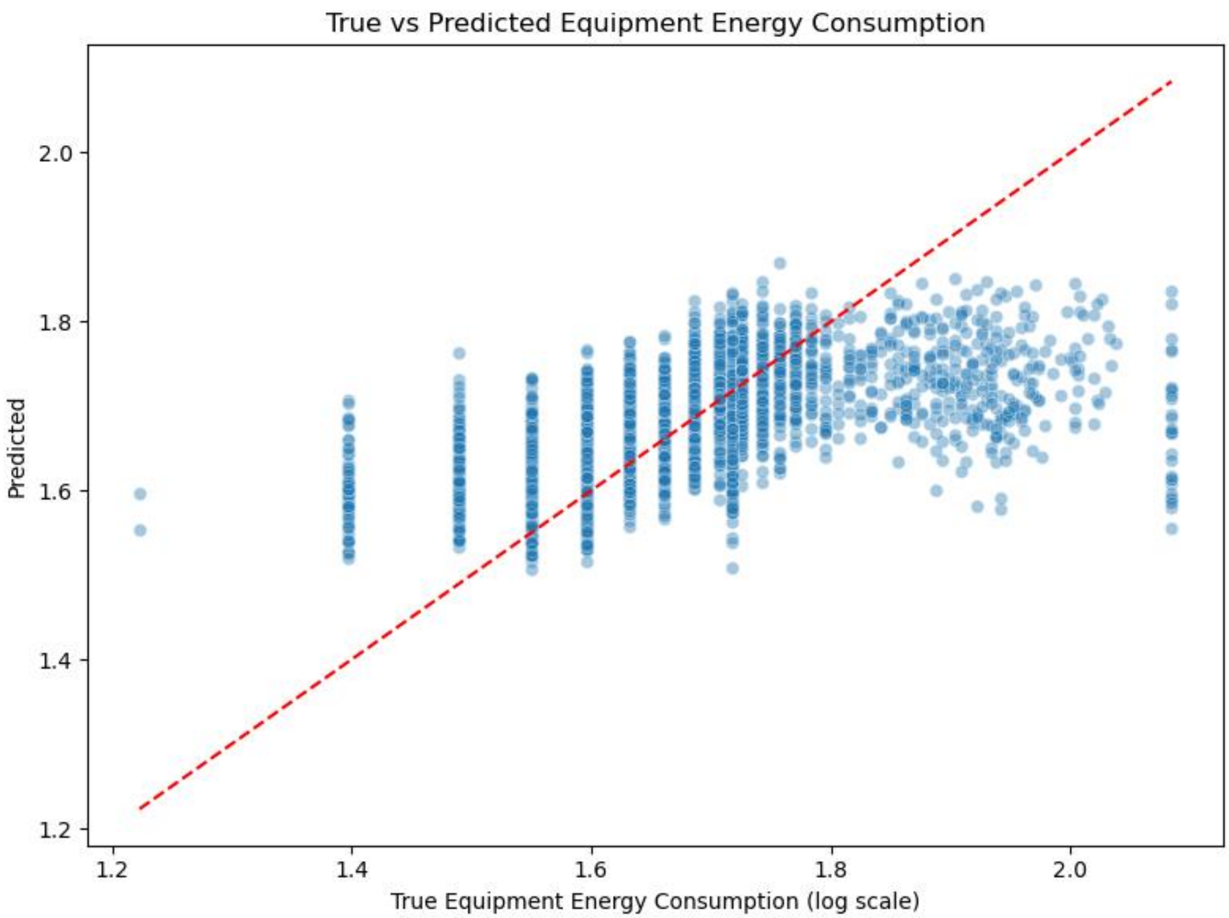
- **Hour** was extracted from the timestamp and became the most important feature, indicating that energy consumption is heavily influenced by the time of day.
- Features like **lighting_energy**, **zone8_temperature**, and **zone3_humidity** also played significant roles in predicting energy consumption.
- **Random variables** showed minimal importance in the model, indicating that they do not contribute substantially to the predictions.

Model Training and Testing

The dataset was split into training (80%) and testing (20%) sets. We used this division to train the model and evaluate its performance. Cross-validation techniques were not explicitly mentioned, but the model was trained on the training set and tested on the holdout testing set.

4. Model Evaluation

Evaluation Metrics



We evaluated the model using the following metrics:

- **Root Mean Squared Error (RMSE)**: 0.097
- **Mean Absolute Error (MAE)**: 0.067
- **R^2 Score**: 0.358

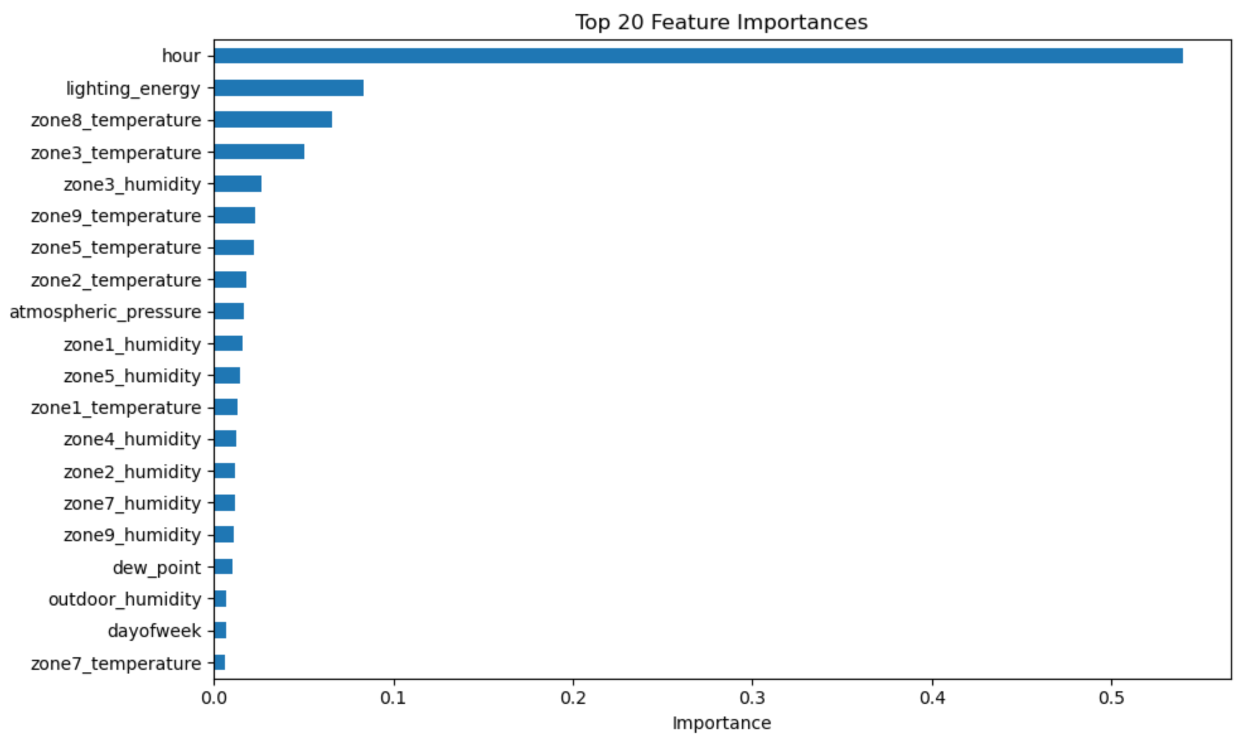
These metrics suggest that while the model is not perfect, it has a reasonable ability to predict energy consumption, with the R^2 score indicating that around 35.8% of the variance in energy consumption is explained by the model.

Feature Importance

The top 5 most important features were:

1. **hour**: 0.5395
2. **lighting_energy**: 0.0831
3. **zone8_temperature**: 0.0659
4. **zone3_temperature**: 0.0507
5. **zone3_humidity**: 0.0268

Interestingly, **random_variable1** and **random_variable2** had very low importance, suggesting they were not useful predictors for the model.



---

5. Insights & Recommendations

Key Insights

- The **time of day**, represented by the **hour**, was found to be the most important feature for predicting energy consumption. This suggests that energy usage patterns may fluctuate based on the time of day.
- **Lighting energy** and **temperature** in specific zones (**zone8_temperature**, **zone3_temperature**) also have notable influences on the energy consumption.

Actionable Recommendations

- **Optimizing Lighting Usage**: Since **lighting_energy** is a significant feature, the facility could consider optimizing lighting based on time-of-day schedules, turning off lights in unused areas during off-hours.
- **Temperature Control**: Temperature fluctuations in **zone8** and **zone3** strongly influence energy consumption, suggesting that energy-efficient HVAC systems and insulation improvements in these zones could lead to energy savings.
- **Time-based Adjustments**: Given the importance of **hour**, implementing energy-saving protocols that account for peak hours of energy usage could improve overall energy efficiency.

---

6. Conclusion

Summary

In this project, we successfully developed a machine learning model to predict energy consumption in industrial equipment based on environmental data. The model achieved a reasonable performance with an R^2 score of 0.358. We identified key factors such as the time of day, lighting energy, and zone temperatures as influential features for energy consumption.

Future Work

To improve the model's accuracy, we could:

- Incorporate additional data, such as maintenance schedules or equipment usage logs, which may further explain energy consumption patterns.
- Experiment with more advanced machine learning models, like neural networks or ensemble techniques, to capture more complex relationships.
