Energy Consumption Prediction Report

1. Problem Statement

The goal of this project was to predict equipment energy consumption in a facility using environmental and sensor data. Accurate predictions will help in identifying inefficiencies and guide energy-saving decisions.

2. Approach

- **Data Preprocessing:**
  - Handled missing values.
  - Converted `timestamp` to datetime format.
  - Removed or imputed incomplete rows/columns.
- **Exploratory Data Analysis (EDA):**
  - Analyzed distributions, correlations, and outliers.
  - Found strong inter-correlations among zone-wise temperature and humidity.
  - Identified that `random_variable1` and `random_variable2` had weak correlations with most features but showed mutual correlation.
- **Feature Engineering:**
  - Created interaction terms using polynomial features (2nd-degree, interaction-only).
  - Added domain-driven features like humidity-temperature ratios and temperature differences across zones.
  - Scaled numerical features using `StandardScaler`.
- **Modeling:**
  - Trained and evaluated several models:
    - Ridge Regression
    - Lasso Regression
    - Random Forest Regressor
    - Gradient Boosting Regressor
  - Performed feature selection using `SelectKBest` and regularization.

3. Key Insights

- Random variables had weak interaction with most features, indicating low explanatory power, but were retained for completeness.
- Dew point and atmospheric pressure combinations affected energy consumption patterns.
- High correlation between wind speed and visibility suggested meteorological interplay.
- Equipment energy consumption varied most with zone-wise temperatures and humidities, especially from zone 1 and zone 5.

4. Model Evaluation

Model	RMSE	MAE	R^2
Ridge Regression	0.9189	0.4246	0.1717
Lasso Regression	0.9175	0.4097	0.1744
Random Forest	**0.9004**	**0.4059**	**0.2048**
Gradient Boosting	0.9062	0.4041	0.1945

5. Recommendations

- **Focus on Climate Control in Zones 1 and 5:** These zones showed significant influence on energy consumption. Optimizing air conditioning or humidity control in these zones could yield energy savings.
- **Implement Feature-Based Monitoring:** Use derived features like `humidity_temp_ratio` and `temp_diff` as triggers for operational efficiency checks.
- **Explore Wind and Visibility Impact:** Since wind speed and visibility are strongly correlated and potentially affect HVAC efficiency, adaptive control strategies during low visibility or high wind should be considered.
- **Optimize Equipment During High Usage Times:** The `is_high_energy` feature can help flag peak usage periods for possible load balancing or scheduling adjustments.

-Submitted by

Satwik Kumar