Brief project report

Energy Consumption Prediction Report

*** Problem Statement**

SmartManufacture Inc. aims to accurately predict equipment energy consumptionin a manufacturing facility using historical sensor data. The goal is to support operational efficiency and identify areas for energy savings.

*** Approach**

1. Exploratory Data Analysis (EDA):
   - Inspected data shape, types, and missing values.
   - Converted timestamps to datetime.
   - Explored distribution of energy consumption and time-based trends (hour/day).
   - Used correlation heatmaps and scatter plots for feature relationships.

2. Data Preprocessing:
   - Handled missing and invalid values.
   - Removed outliers using Z-score.
   - Dropped unrealistic values (e.g., negative or very low energy readings).
   - Engineered new features like `hour`, `weekday`.

3. Feature Engineering & Selection:
   - Scaled temperature, humidity, and other numerical features using StandardScaler.
   - Used correlation analysis to select top features related to energy consumption.
   - Applied log transformation to the target variable to stabilize variance.

4. Model Development & Training:

- Built and evaluated three models:
  - Linear Regression
  - Random Forest Regressor
  - XGBoost Regressor
- Performed hyperparameter tuning using RandomizedSearchCV for tree-based models.

5. Model Evaluation:
- Compared models using R^2 and RMSE on the test set.
- Best-performing model: XGBoost

*** Model Performance**

Model	R^2	RMSE
XGBoost	0.999998	0.000873
Random Forest	0.999996	0.001150
Linear Regression	0.847977	0.217117

*** Key Insights**

- Random variable 1 and 2 both were not related to any variable or feature present in the data, so it was not used while training the model.
- XGBoost outperformed all other models, capturing almost all variance in the data.
- Linear Regression underperformed, likely due to nonlinear patterns in sensor readings.

*** Recommendations**

1. Optimize temperature and humidity control during peak working hours.
2. Schedule maintenance during off-peak hours (identified through hourly trends).
3. Deploy XGBoost model in production for real-time energy prediction and alerting.