Energy Consumption Prediction Using Machine Learning

Project Report

Energy Prediction

Name: Pulkit Mangal

Institution: IIT Madras

Submission Date: May 9, 2025

Energy Consumption Prediction Using Machine Learning

Objective

The objective of this project is to develop machine learning models that can accurately predict equipment energy consumption based on a variety of environmental and operational parameters. The approach involves data preprocessing, feature engineering, model training, and evaluation to identify the most effective predictive techniques.

1. Dataset Overview

- Source: CSV file named data.csv
- Type: Time-series data
- Target Variable: equipment_energy_consumption
- Features: 34 variables, including environmental conditions, outdoor climate metrics, and engineered variables
- Number of records: ~16,000
- Missing values: Predominantly in humidity and temperature fields

2. Exploratory Data Analysis (EDA)

- Computed descriptive statistics for all features
- Conducted correlation analysis on engineered variables and the target
- Weak linear relationships observed
- Visualizations used to explore feature distributions and interactions

3. Data Preprocessing

Missing Value Imputation:

- Used KNNImputer for missing values (excluding timestamp and target)

Energy Consumption Prediction Using Machine Learning

Feature Engineering:

- Extracted time features such as hour, weekday, month

Feature Scaling & Encoding:

- Applied StandardScaler and RobustScaler

- Used OrdinalEncoder for categorical features

4. Feature Selection

- Used Recursive Feature Elimination (RFE) to identify relevant predictors

- Improved efficiency and performance by reducing dimensionality

5. Model Development

Algorithms:

- LinearRegression, Ridge, Lasso, SGDRegressor, RandomForestRegressor, XGBRegressor

Training:

- Split into training/testing sets

- Used pipelines for preprocessing and training

- Computed feature importances for tree-based models

6. Model Evaluation

Metric:

- Coefficient of Determination (R^2) used for evaluation

Findings:

- All models evaluated on standardized data
- Feature selection improved performance
- Tree-based models provided interpretability
- Best-performing model had highest R^2 score

7. Results Summary

- Robust preprocessing enabled effective training despite missing values
- Compared linear and ensemble models
- Feature selection improved generalization
- Workflow serves as a replicable pipeline

8. Conclusion

This project demonstrates a comprehensive ML workflow for predicting energy consumption. It highlights the importance of systematic data preprocessing, model selection, and evaluation in building effective predictive systems for energy management.

Note

For detailed R^2 scores, feature importance charts, or visual outputs, refer to the project notebook.

References

[1] Energy Prediction Colab Notebook:

https://colab.research.google.com/drive/1RxHeVDvjYceDkUDocn_9aBPB7XuBrgpj?usp=sharing