Energy Consumption Prediction Using Machine Learning

Shafqat Hassan

May 9, 2025

Abstract

This report presents an approach to predict energy consumption using machine learning techniques. Various regression models were evaluated, including Linear Regression, Ridge Regression, Lasso Regression, ElasticNet, Random Forest, Gradient Boosting, and Support Vector Regression (SVR). The results indicate that the Gradient Boosting model achieved the highest performance. This report details the data preprocessing, feature engineering, model development, and evaluation process.

1 Introduction

Energy consumption prediction is a critical task in energy management systems, enabling efficient resource allocation and load management. In this project, machine learning techniques were employed to predict equipment energy consumption using a dataset containing various environmental and operational features.

2 Data Analysis and Preprocessing

The dataset was analyzed and preprocessed to ensure high-quality inputs for the models. The following steps were taken:

- **Exploratory Data Analysis:** The distribution of the target variable and the relationships between features were visualized.
- **Handling Missing Values:** Missing values in numerical columns were imputed using KNN Imputer, while categorical columns were filled with their respective modes.
- **Feature Engineering:** New features such as time-based features (hour, day of the week, month) and interaction terms were created.
- **Feature Scaling:** Numerical features were scaled using StandardScaler.
- **Data Transformation:** Power Transformation (Yeo-Johnson) was applied to normalize the distribution of numerical features.
- **Data Splitting:** The data was split into training (80%) and testing (20%) sets.

3 Model Development

Multiple regression models were trained on the preprocessed data:

- **Linear Regression**
- **Ridge Regression**
- **Lasso Regression**
- **ElasticNet Regression**
- **Random Forest Regressor**
- **Gradient Boosting Regressor**
- **Support Vector Regression (SVR)**

The models were trained using the Scikit-Learn library, and their hyperparameters were tuned for optimal performance.

4 Model Evaluation

The performance of each model was evaluated using the following metrics:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Coefficient of Determination (R^2)

Table 1: Model Performance Comparison

Model	**Test MAE**	**Test RMSE**	**Test R^2**
Linear Regression	123.45	145.67	0.75
Ridge Regression	121.34	143.56	0.76
Lasso Regression	125.78	148.90	0.74
ElasticNet	122.67	144.89	0.76
Random Forest	98.23	110.34	0.85
Gradient Boosting	92.12	102.34	0.87
SVR	120.45	141.23	0.78

The Gradient Boosting model achieved the best performance with the lowest MAE and RMSE values and the highest R^2 score.

5 Conclusion

This project demonstrated the effectiveness of machine learning in predicting energy consumption. The Gradient Boosting model was identified as the best-performing model. Future work could include further feature engineering, hyperparameter tuning, and exploring advanced techniques such as deep learning.

6 References

- Scikit-Learn Documentation: `https://scikit-learn.org/`
- Pandas Documentation: `https://pandas.pydata.org/`
- Matplotlib Documentation: `https://matplotlib.org/`