Predicting Appliance Energy Use in Residential Buildings

Collins Ayidan

Master of Science in Mathematics and Data Science

September 2024

Introduction

- **Background**: Energy use in residential buildings represents a major portion of total energy consumption, impacting both economic costs and environmental consequences.

- **Objective**: Identify the best machine learning model for predicting appliance energy consumption in residential buildings.

Literature Review

Influential Factors: Building characteristics, household composition, appliance usage patterns, and environmental conditions (e.g., temperature, humidity, wind speed)

Machine Learning Models: Random Forest, Gradient Boosting Machines (GBM), Support Vector Regression (SVR), and Neural Networks (NN) are effective in predicting energy use.

Optimization Techniques: This study uses HalvingRandomSearchCV, chosen due to the success of RandomSearchCV in the literature, for tuning hyperparameters.

Evaluation Metrics: Common metrics for assessing model performance include RMSE, MAE, R^2, and MAPE.

Methodology

DATA SOURCE: UCI APPLIANCE ENERGY PREDICTION DATASET.

DATA PREPROCESSING: FEATURE ENGINEERING INCLUDING CLASSIFICATION OF DAYS AS WEEKENDS OR WEEKDAYS, SPECIFIC DAYS OF THE WEEK, SEASONS AND MONTHS.

MODELS: 8 MACHINE LEARNING MODELS, INCLUDING LINEAR REGRESSION, SUPPORT VECTOR REGRESSION, RANDOM FOREST, DECISION TREE, GRADIENT BOOSTING, XGBOOST, EXTRA TREES, AND NEURAL NETWORKS.

Training and Evaluation Metrics

- Training and Testing Procedure: HalvingGridSearchCV

- Metrics Used:
  - Root Mean Square Error (RMSE)
  - R-Squared (R^2)
  - Mean Absolute Error (MAE)
  - Mean Absolute Percentage Error (MAPE)

Results - Model Performance

BEST MODELS:

- EXTRA TREES: TESTING RMSE = 63.28, R^2 = 0.60

- RANDOM FOREST: TESTING RMSE = 65.51, R^2 = 0.57

- XGBOOST: TESTING RMSE = 64.36, R^2 = 0.57

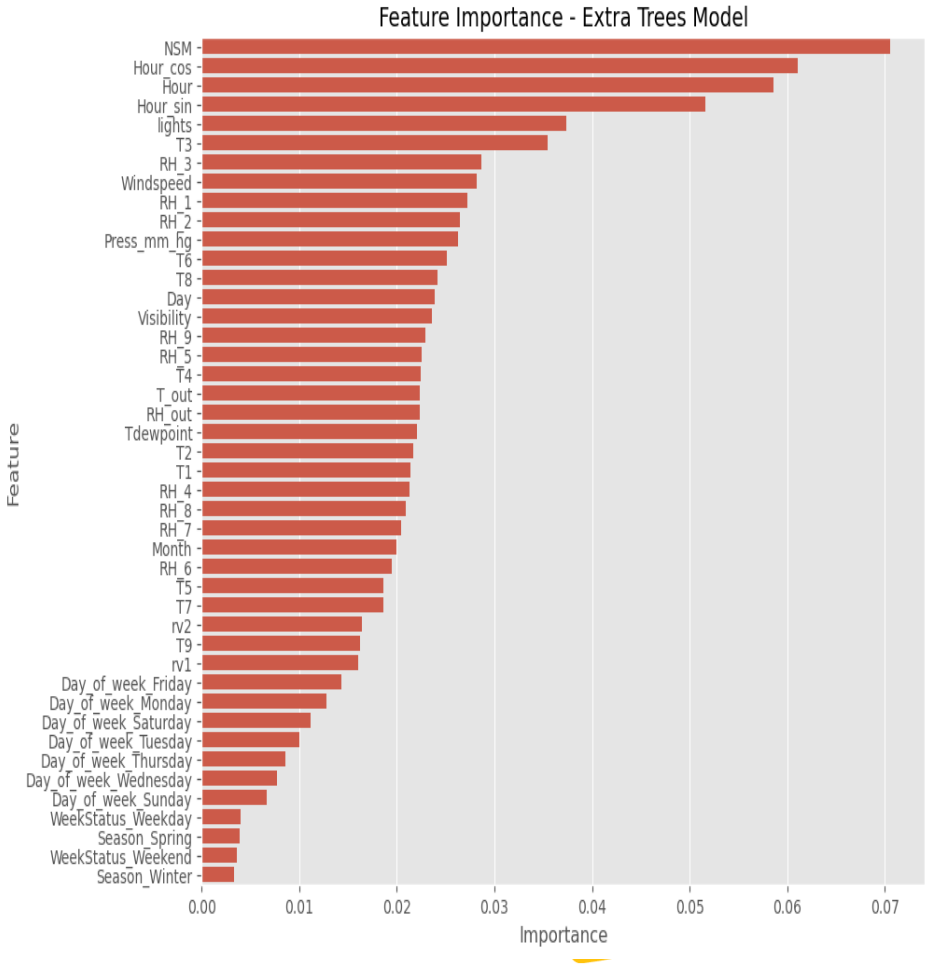
UNDERPERFORMING MODELS: DECISION TREE AND LINEAR REGRESSION.

	Testing RMSE	Testing R2	Testing MAE	Testing MAPE %
LM	89.39	0.20	51.95	61.74
SVM	75.14	0.43	33.91	30.38
GBM	83.92	0.29	46.48	52.84
RF	65.51	0.57	31.37	31.44
XGB	64.36	0.58	31.84	32.51
ET	63.28	0.60	29.86	29.68
DT	90.53	0.18	38.69	31.84
NN	78.51	0.38	43.41	47.89

Feature Importance (Extra Trees Model)

- Key Features:
- - Number of Seconds from Midnight (NSM)
- - Hour of the Day
- - Environmental conditions (e.g., temperature, humidity)

- Less Important Features: Season, Week Status



Conclusion

- Summary:
  - Extra Trees identified as the best model.
  - Feature importance showed the some of the new features including hour of the day as one of the influential factors that affects appliance energy usage.

Future Work

Suggestions:

- Further refine underperforming models (e.g., SVM, NN).

- Explore additional features or external data for better prediction accuracy.

- Incorporate socio-demographic factors and real-time data in future models.