Predictive Modeling for Household Energy Consumption Optimization

Smart energy optimization: Building a predictive model that forecasts household electricity consumption to enhance efficiency.

Client: Green Leaf Energy: A fictional energy-tech firm focused on enabling residential energy efficiency through data insights.

Team Members:
S. M. Wahid Chowdhury
Himel Paul
Kewei Shu

Problem Statement

Why Optimizing Household Energy Use Matters

Rising energy costs
Households face escalating electricity bills due to inefficient usage and inability to forecast demand.

Carbon footprint concerns
Poor energy habits contribute to higher greenhouse gas emissions, increasing environmental impact.

Lack of control and feedback
Consumers have limited access to predictive insights needed to adjust usage patterns proactively.

Proposed Solution

Using Machine Learning to Forecast and Optimize Energy Use

ML-powered forecasting
Develop a predictive model trained on historical residential data to forecast household energy usage.

Key factor identification
Leverage feature importance analysis to discover which home attributes most affect energy consumption.

User-focused energy insights
Deliver actionable outputs for users to reduce costs and emissions via smarter decisions.

Dataset Overview

Residential Energy Consumption Survey (RECS 2020)

- **Comprehensive national data:** RECS 2020 provides detailed household-level data on energy consumption across the United States.
- **Diverse energy variables:** Includes electricity, natural gas, fuel oil usage; building characteristics, and appliance-level metrics.
- **Geographical and climate details:** Regional distinctions allow for climate-aware energy consumption modeling.

Photo by NASA on Unsplash

Preprocessing & Feature Engineering

Enhancing Data Quality and Predictive Power

- **Data cleaning and normalization:** Handled missing values, normalized numerical features, and encoded categoricals for consistent modeling.
- **Custom feature generation:** Created new variables like per capita energy use, appliance usage ratios, and seasonal adjustments.
- **Correlation analysis:** Evaluated feature interdependence via heatmaps and importance metrics for model-ready input.

Photo by Skye Studios on Unsplash

Models Used

Linear Regression, Decision Tree, XGBoost

- **Linear Regression:** Baseline model to capture linear dependencies and benchmark against more complex methods.
- **Decision Tree:** Captures non-linear patterns and feature interactions with interpretable structure.
- **XGBoost:** Gradient boosting ensemble model offering superior accuracy via regularized training.

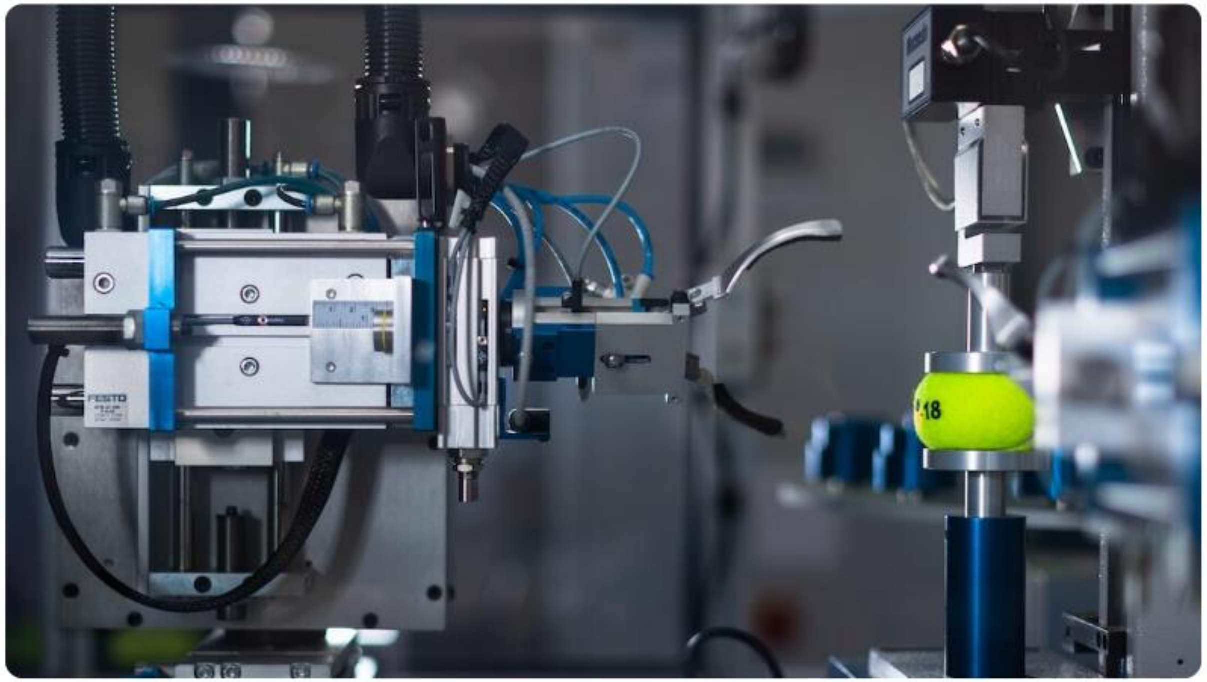
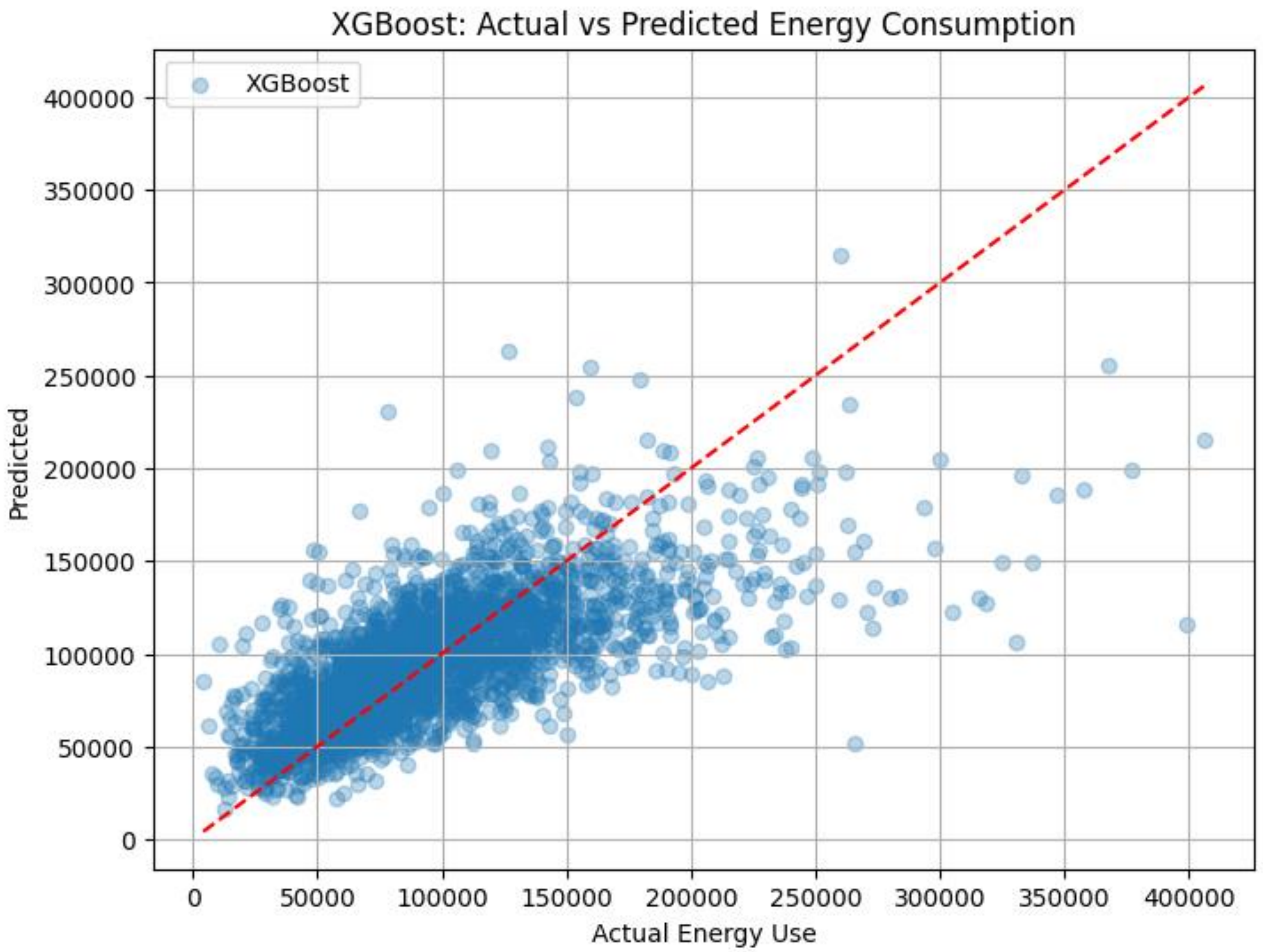


Photo by ThisisEngineering RAEng on Unsplash

Results Summary

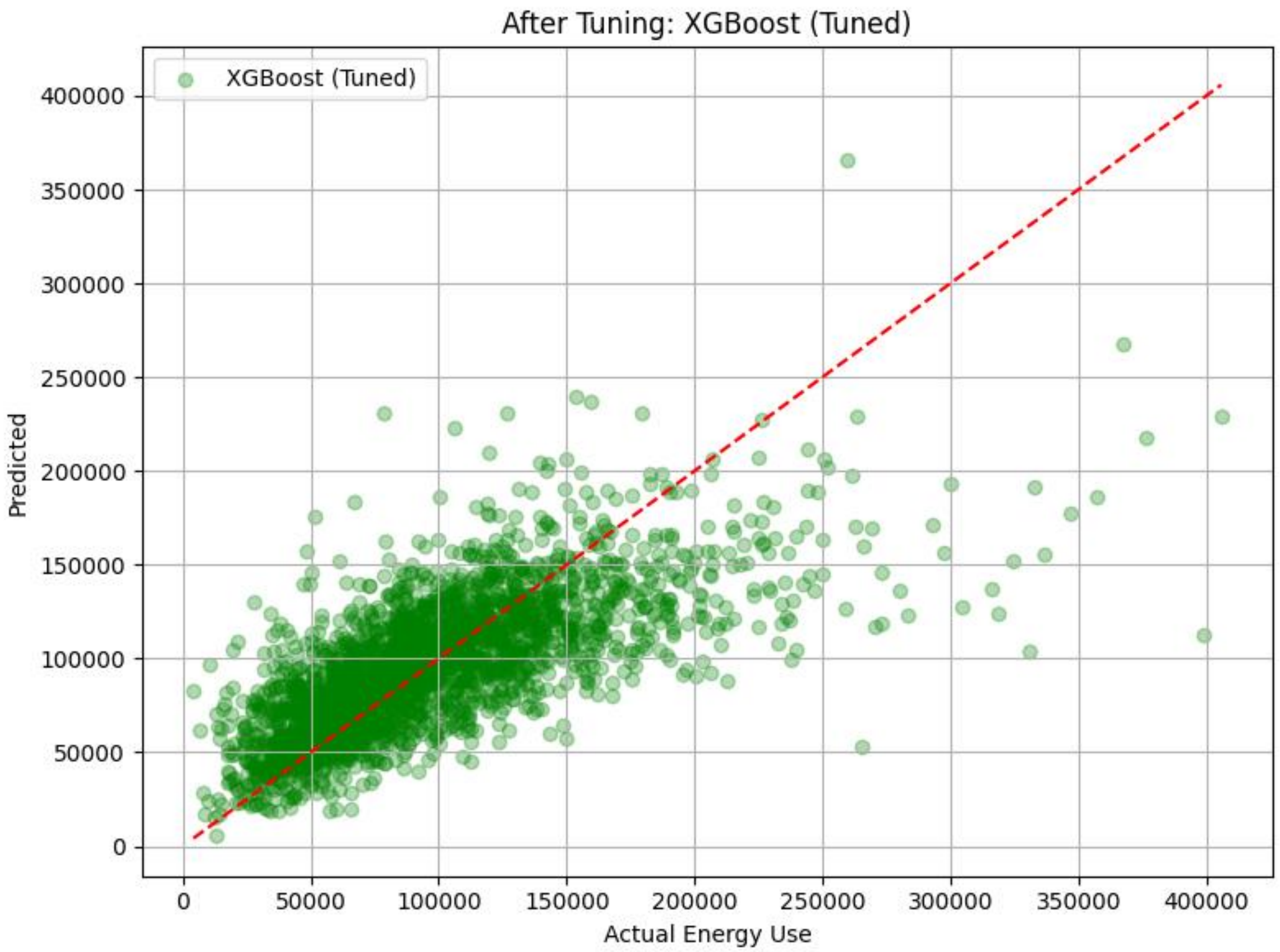
Model Comparison and Performance Metrics

- **XGBoost outperformed others:** Achieved lowest RMSE and highest R^2, demonstrating strong predictive capabilities.
- **Impact of feature engineering:** Enhanced features significantly improved model accuracy and reduced prediction error.
- **Model benchmarking:** Comparative metrics showed Decision Tree and Linear Regression lagged behind XGBoost.

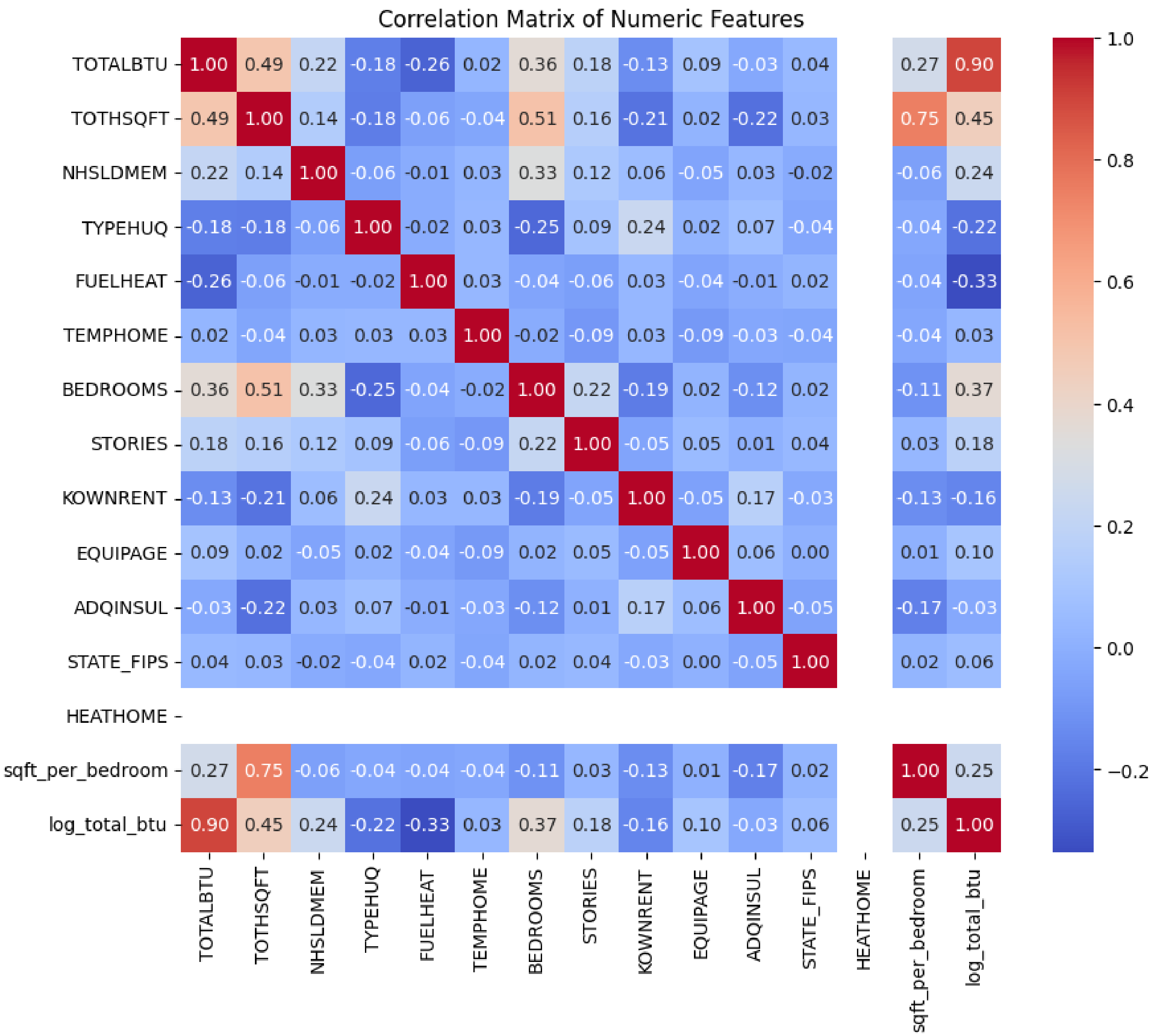


Hyperparameter Tuning

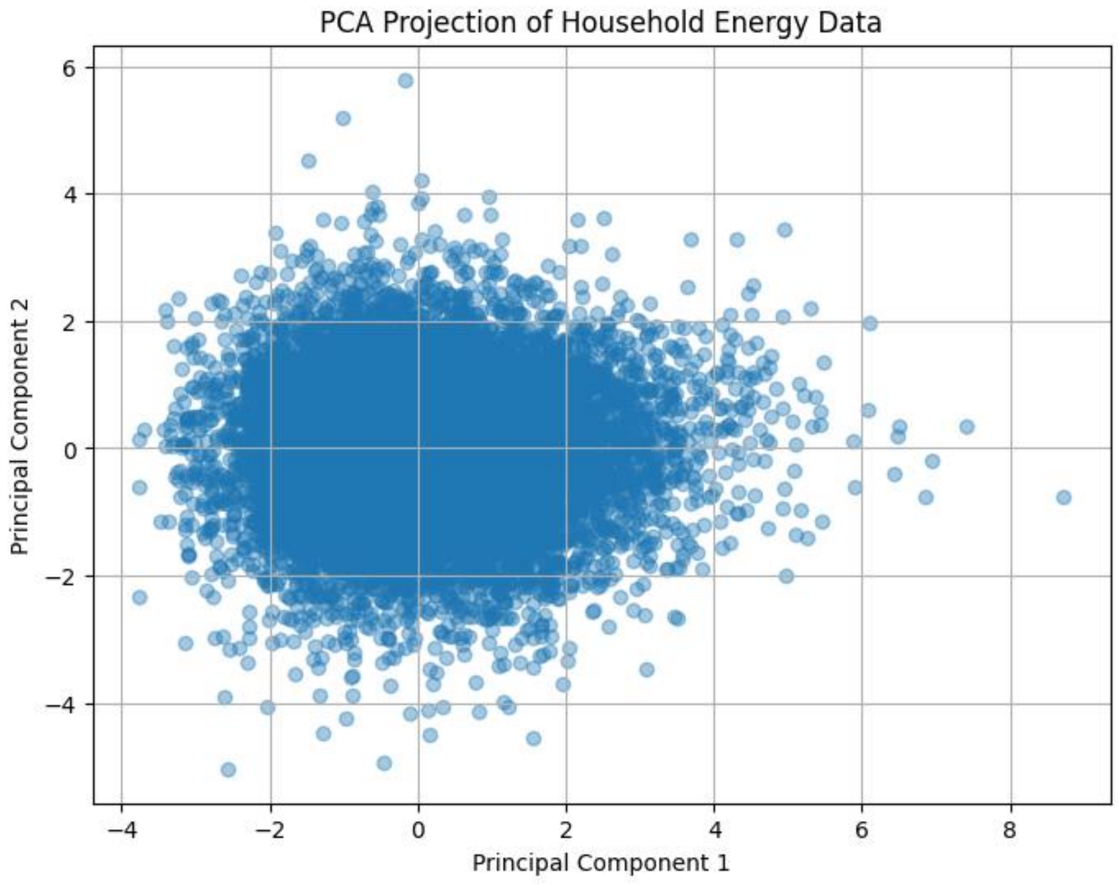
- To further optimize performance, we applied GridSearchCV on XGBoost. We tested multiple combinations of hyperparameters across 3-fold cross-validation. The best model achieved a slightly improved R-squared of 0.4985 with reduced overfitting.

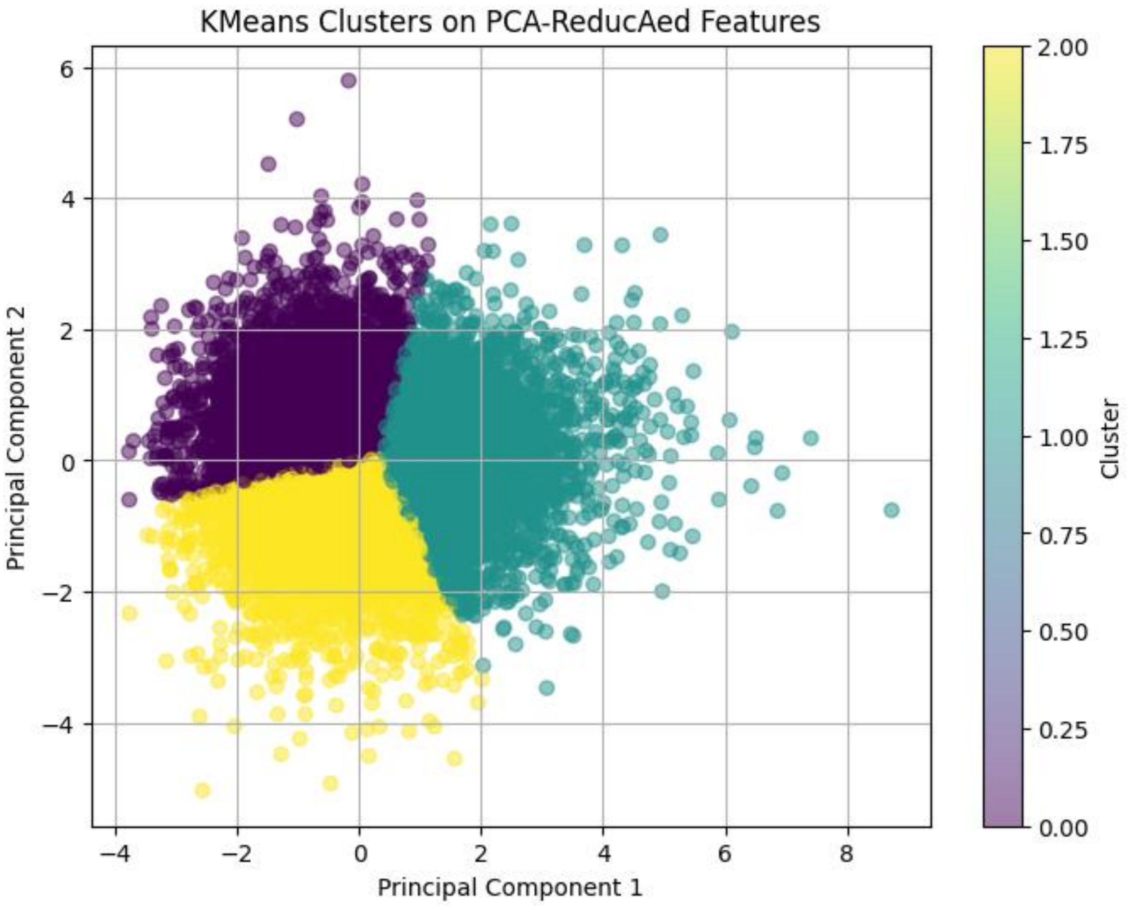


Feature Correlation & Engineering



PCA + KMeans Visualization





Cluster Insights

cluster	TOTALBTU	TOTHSQFT	NHSLDMEM	TYPEHUQ	FUELHEAT	TEMPHOME	BEDROOMS	STORIES
0	79531.593244	1598.520214	2.391686	2.141686	2.746186	71.469298	2.801678	1.229977
1	129237.853445	3077.098284	3.498700	2.043162	2.099064	69.469579	4.151846	2.082163
2	88395.023713	1765.300992	1.993385	2.167144	2.298567	67.667475	2.871224	1.653583

- Cluster 0 represented moderate households with fair insulation and warmth.
- Cluster 1 had the largest, most energy-intensive homes with older equipment.
- Cluster 2 had smaller families, newer equipment, and the lowest indoor temperatures. Clustering helps design targeted energy-saving strategies.

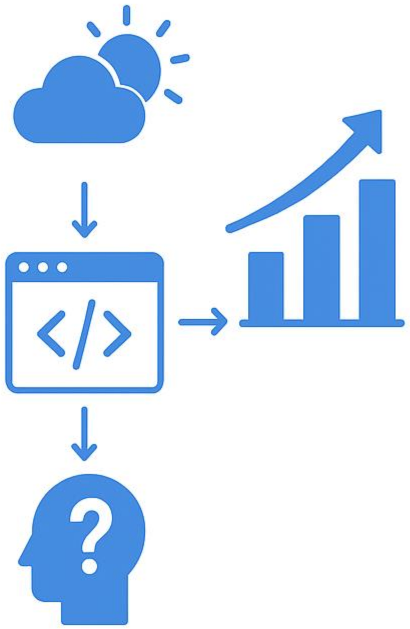
Model Limitations

While our model performed well given the features, the R-squared score suggests that a large portion of energy use variation remains unexplained. This may be due to missing context like occupant behavior, local weather, or seasonality.

Photo by Adeolu Eletu on Unsplash

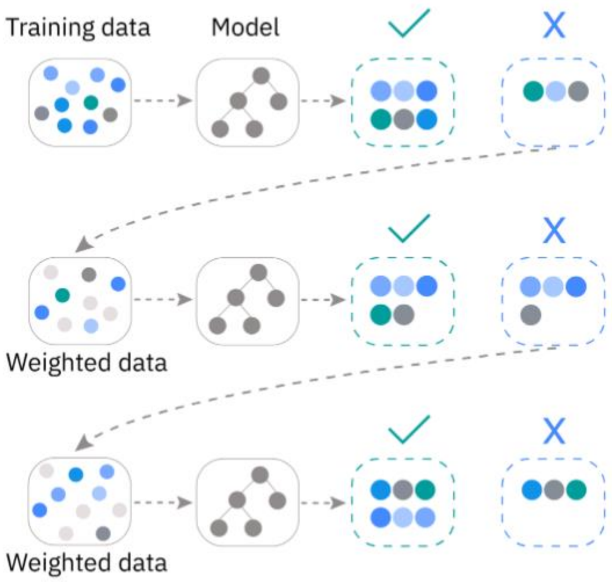
Recommendations for Future Work

To improve
we suggest adding external data such as climate records or utility rates, testing more advanced models like LightGBM, and exploring SHAP for explainability. Better features and deeper models could enhance predictive power.



Conclusion

- Successfully built predictive models, with XGBoost performing best
- Clustering analysis revealed household energy patterns
- Provided data-driven efficiency recommendations for Green Leaf Energy



Green Leaf ENERGY