Project Proposal: Predicting Energy Efficiency of Buildings

By Team Nwapa, AI Saturdays Lagos Cohort 9 - Flipped

Problem Statement

Nigeria and many developing countries face growing challenges in energy consumption, with households spending a large portion of their income on electricity for heating and cooling. Poorly designed buildings often lead to excessive energy usage, higher bills, and pressure on the already unstable national grid.

Builders and homeowners lack simple, data-driven tools to estimate how efficient a house will be before construction or renovation.

Our project aims to build a predictive model to estimate the heating and cooling loads of buildings based on features such as relative compactness, surface area, wall area, roof area, glazing area, and orientation.

This solution matters because it will help guide cost-effective construction decisions, promote energy conservation, and support sustainable development in Nigeria and beyond.

Existing Solutions

Currently, building energy efficiency assessments in Nigeria are mostly done through manual inspections, expert judgment, or compliance with basic building codes. These methods are often subjective, inconsistent, and not easily accessible to the average homeowner.

In contrast, our project applies machine learning techniques to predict energy efficiency in a transparent, reproducible, and interactive way.

Unlike existing rule-based approaches, our model will:

i. leverage real-world datasets,

ii. perform statistical analysis and feature interpretation,

iii. provide a deployable tool that anyone can use to assess efficiency instantly.

Objectives

i. Clean and preprocess the energy efficiency dataset for accurate analysis.

ii. Explore data trends and perform Exploratory Data Analysis (EDA) to uncover patterns in building features.

iii. Build regression models to predict exact heating load and cooling load values.

iv. Build classification models to categorize buildings into Low / Medium / High energy-consuming groups (by creating labeled categories from the regression targets).

v. Analyze feature importance to identify which building parameters (roof area, wall area, glazing, orientation, etc.) influence energy consumption the most.

vi. Evaluate model performance using metrics such as RMSE, MAE, Accuracy, Precision, Recall, and F1-score.

vii. Deploy the final model on a web interface for easy public access.

Proposed Dataset

We will use the Energy Efficiency Dataset from the UCI Machine Learning Repository (also available on Kaggle): [URL]

Why We Chose This Dataset:

This dataset contains detailed building characteristics (e.g., relative compactness, surface area, wall area, roof area, orientation, glazing area) and their corresponding heating and cooling loads.

It is well-structured, clean, widely used in ML studies, and provides an excellent basis for predictive modeling that aligns with real-world energy efficiency challenges.

Dataset Details:

768 building samples

8 input features (X1–X8)

2 numeric targets (Heating Load, Cooling Load)

Proposed Methodology

1. Data Sourcing & Cleaning

Load the dataset using Pandas

Handle missing values, duplicates, and outliers

Standardize feature formats and ensure data quality

2. Exploratory Data Analysis (EDA)

Visualize distributions of heating and cooling loads

Explore feature relationships and correlations using heatmaps, scatter plots, and pair plots

Detect patterns between building shape and energy performance

3. Feature Engineering

Encode categorical features (Orientation, Glazing Area Distribution)

Create new engineered features like wall-to-roof ratio or surface-to-volume ratio

Normalize numerical features for better model performance

Create new classification labels (Low / Medium / High) from the numeric targets based on chosen thresholds (like quantiles)

4. Modeling & Evaluation

Regression: Train Linear Regression, Random Forest, and Gradient Boosting (XGBoost/LightGBM) to predict heating and cooling loads

Classification: Train Logistic Regression, Random Forest Classifier, and SVM to classify buildings as Low, Medium, or High energy-consuming

Evaluate using RMSE, MAE (for regression) and Accuracy, Precision, Recall, F1-score (for classification)

Use cross-validation and GridSearchCV for hyperparameter tuning

Perform feature importance analysis (using Random Forest, SHAP, or permutation importance) to understand which design features most affect energy loads

5. Deployment

Build a Streamlit web app where users can input building features (roof area, glazing, etc.) and get predicted heating/cooling loads and efficiency categories

Optionally host on Hugging Face Spaces or GitHub Pages for wider access.

Modeling Plan

Linear Regression: Baseline for interpretability

Random Forest Regression: Handles non-linear relationships

Gradient Boosting (XGBoost/LightGBM): High accuracy and robustness

Classification models (Logistic Regression, Random Forest Classifier, SVM): For Low/Medium/High efficiency categories

Models will be compared using cross-validation and metrics like RMSE, MAE, Accuracy, and F1-score to choose the best approach.

Expected Outcomes

A robust predictive model for building energy efficiency

A simple classification system (Low/Medium/High) for buildings

Clear insights showing which features drive energy consumption the most

An interactive web tool to help homeowners, builders, and policymakers make informed choices

Community Impact

This project will empower individuals, builders, and organizations with reliable tools to design and select energy-efficient buildings.

It can help:

reduce electricity bills

ease pressure on the national grid

promote sustainability in Nigeria's housing and construction sector

By providing data-driven insights, the project contributes to both economic savings and environmental protection.

Team Members

- Chitom Uzokwe (team lead)
- Precious Ita
- Joy Eromosele

Acknowledgement

Our proposal drew inspiration from the AI Saturdays Lagos learning resources and global research on sustainable housing design.

References

UCI Machine Learning Repository – Energy Efficiency Dataset

Kaggle – Energy Efficiency Dataset

Nigerian Building and Energy Reports

Building Energy Efficiency Guideline for Nigeria (2016) - **By:** Federal Ministry of Power, Works & Housing with GIZ support

Transition towards Energy Efficiency: Developing the Nigerian Building Energy Efficiency Code