# UNIVERSITY OF CALGARY

## ENSF444: Machine Learning Systems

### Project Proposal: Predictive Modeling for Household Energy Consumption Optimization

| Team members | Names |
|---|---|
| Team Member 1 | S M Wahid Chowdhury |
| Team Member 2 | Himel Paul |
| Team Member 3 | Kewei Shu |

## Client: Green Leaf Energy

## Executive Summary

Green Leaf Energy aims to give homes smart tools to use less electricity. This plan is to build a strong computer model that guesses home energy use. Using past energy use and electrical data, this model will help people make smart choices to save energy, lower costs, and help the environment. This project will use proven computer learning methods, careful testing, and a clear plan to put it to use, making sure Green Leaf Energy's customers get real benefits.

## Introduction

As we know the increasing cost of energy and the growing concern for environmental health illustrate the importance of managing our energy use wisely. Green Leaf Energy's dedication to providing smart solutions directly addresses this need. This project aims to create a prediction tool that gives homeowners clear, practical information about their energy use, enabling them to take control and improve their energy efficiency.

## Problem Statement

Households currently lack the granular, predictive insights necessary to optimize their energy usage. This results in:

- **Increased energy costs:** Without accurate forecasts, users cannot anticipate and manage peak consumption periods.
- **Environmental impact:** Inefficient energy usage contributes to a larger carbon footprint.
- **Limited control:** Users lack the tools to make informed decisions about their energy consumption patterns.

Green Leaf Energy needs a solution that gives correct, on-time guesses about how much energy homes use. This will let people plan their energy use ahead of time and get big savings.

## Proposed Solution

This project will develop a machine learning-based predictive model using the publicly available Residential Energy Consumption Survey (RECS). The model will be designed to:

- Accurately forecast household energy consumption based on historical data.
- Identify key factors influencing energy usage.
- Provide actionable insights for users to optimize their energy consumption.

## Project Methodology

### Data Acquisition and Preprocessing

- Utilize the Residential Energy Consumption Survey (RECS), a comprehensive dataset containing household-level energy consumption, fuel usage, and appliance-specific data across the U.S.
- Implement rigorous data cleaning procedures to address missing values and inconsistencies.
- Normalize numerical features to ensure consistent scaling.
- Encode categorical variables as necessary.
- **Dataset Link:** **Residential Energy Consumption Survey (RECS) 2020 dataset**

## Feature Engineering

### Extract Relevant Features

- Total energy consumption
- End-use consumption
- Household characteristics
- Building details
- Geographical factors

### Create Derived Features

- Energy usage per household member.
- Energy efficiency based on home size.
- Seasonal variations in energy consumption.
- Proportion of energy used by different appliances.

### Feature Importance Analysis

- Identify key factors influencing high energy consumption.
- Analyze the relationship between household income and energy use.
- Assess the impact of fuel type on total consumption.

### Model Development and Training

- Implement and compare the following machine learning models:
    - **Linear Regression:** To establish a baseline and understand linear relationships.
    - **Decision Tree:** To capture non-linear relationships and interactions.
    - **XGBoost:** A gradient boosting algorithm for high predictive accuracy.
- Employ cross-validation techniques to ensure model robustness and generalization.
- Hyperparameter tuning of all models.

## Frameworks

- Python (NumPy, Pandas, Scikit-learn, XGBoost, Matplotlib, Seaborn)

## Components

- **Data Processing:** Pandas (handling missing values), Scikit-learn (categorical encoding, scaling).
- **Feature Selection:** Feature importance ranking, correlation analysis.
- **Model Training:** Scikit-learn for Linear Regression & Decision Tree, XGBoost for boosting.
- **Hyperparameter Tuning:** Hyperparameter tuning, cross-validation.
- **Model Evaluation:** Evaluating model performance using established metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared ($R^2$). Visualize model performance and residuals to identify areas for improvement. Compare the performance of all three models to select the optimal solution.

## Expected Outcomes

- A high-accuracy predictive model for household energy consumption.
- Detailed insights into key factors influencing energy usage.
- A comprehensive report documenting the project methodology, results, and recommendations.
- Improved ability for Green Leaf Energy's end users to manage their energy consumption.

## Potential Challenges and Mitigation Strategies

- **Data Quality Issues:** Implement robust data cleaning and validation procedures.
- **Feature Selection:** Employ feature importance analysis and domain expertise.
- **Model Performance:** Utilize cross-validation and hyperparameter tuning to optimize model performance.
- **Data Shift:** Household energy consumption patterns may change over time due to new appliances, economic factors, or policy changes. Future RECS datasets (e.g., 2025, 2030) can be incorporated to retrain the model and maintain accuracy.

## Conclusion

This project gives Green Leaf Energy a strong tool for their smart energy plans. By guessing how much energy homes will use, people can make better choices, save money, and help the environment. We think this will be very useful for Green Leaf Energy and their customers.