

# Capstone Semester Plan

**Submission Due: January 26, 11:59 PM**

## Project Information

- **Project Title:** Climate TRACE - Estimating Global Greenhouse Gas Emissions from Buildings
- **Partner:** Nicholas Institute for Energy, Environment & Sustainability, Duke University
- **Team Members:** Barbara Flores, Meixiang Du, Jiechen Li, Yulei (Alicia) Xia

## Project Goal(s)

The primary objective of this project is to develop a machine learning model to estimate direct greenhouse gas (GHG) emissions from global building energy use. The model will predict Energy Use Intensity (EUI) using climatic, geographical, and socioeconomic variables, and these predictions will be combined with building floor area and emissions intensity factors to calculate GHG emissions for both residential and non-residential buildings. By refining and expanding the model, we aim to provide a scalable and accurate method for estimating building-related emissions at a global scale, offering valuable insights for policymakers and sustainability efforts.

The deliverables for this project are as follows:

## Main Deliverable

- **Refined EUI Estimation Technique:** The updated and improved methods for estimating EUI, incorporating additional features and refinements from previous iterations to enhance accuracy.
- **Global GHG Emissions Estimation (Reaching Goal):** The global estimation of GHG emissions based on EUI and building floor area.

## Deliverable Structure

- **GitHub Repository with Documentation, Code, and Results Analysis:** A repository containing the source code, documentation on model development, and a detailed analysis of the results, including explanations and discussions of model behavior.

- **Feature Selection and Data Pipeline:** Description of the selected features for EUI estimation, including climatic, geographical, and socioeconomic variables, along with the rationale for their selection. A Python-based data pipeline for processing, cleaning, and preparing the data for modeling will be included in the repository.
- **Model Development, Experimentation, and Evaluation:** This section outlines the testing of various models, including Linear Regression (baseline), Random Forest, K-Nearest Neighbors, and Ensemble models (XGBoost and CatBoost), to determine the best approach for EUI prediction. To address regional variations in global data, we consider three evaluation strategies: Within-Domain, Cross-Domain, and All-Domain. The models are validated across five regions, using key metrics such as MAPE,  $R^2$ , MSE, MAE, RMSE, and WAPE. Additionally, error distribution and actual vs. predicted EUI plots are generated for model diagnostics. Once the best model for predicting EUI is selected, this information will be utilized to calculate the global GHG emissions. The results and key findings are documented in Jupyter Notebooks, with analysis summarized in a slide deck and README file.
- **GHG Emissions Estimation Process (Reaching Goal):** This section outlines the data engineering and feature engineering steps involved in combining EUI predictions with building floor area and emissions intensity factors. A Python-based pipeline is utilized to ensure accurate global emissions estimates for both residential and non-residential buildings.
- **Final Report and Presentation:** A comprehensive slide deck and report summarizing the project's methodology and key results. It will synthesize the analysis, evaluations, and model performance in a structured format.

## Biweekly Breakdown

### By Jan 31

- **Goals:**
  0. Define the Capstone Semester Plan.
  1. Initialize the investigation to find new EUI data.
  2. Integrate new EUI data points into the existing pipeline.
- **Tasks:**
  0. Collect and write down the key information provided by the client in the first meeting.  
(Assigned to: [Yulei Xia])

1. Discuss and revise the Capstone Semester Plan.  
(Assigned to: [Yulei Xia, Jiechen, Meixiang, Barbara])
2. Literature review and resources search for the new EUI data.  
(Assigned to: [Yulei Xia, Jiechen, Meixiang, Barbara])
3. Integrate new EUI data points to the existing pipeline.  
(Assigned to: [Yulei Xia, Jiechen, Meixiang, Barbara])

## **Feb 1-14**

- **Goals:**

0. Complete feature engineering for fully integrating previously downloaded features into the model
1. Create and agree upon the final experimental design for the model evaluation process.
2. Imagery data exploration (Tentative)

- **Tasks:**

0. Improve Temperature Data Representation.  
(Assigned to: [Yulei Xia])
1. Improve Geographic Data Representation.
  1. Add K=1 model in KNN;
  2. Sin Cos transformation of longitude, latitude.  
(Assigned to: [Barbara][Meixiang])
2. Literature review and explore imagery data integration techniques.  
(Assigned to: [Meixiang])

## **Feb 15-28**

- **Goals:**

0. Implement the experiments discussed in the previous period and provide a full assessment of results through validation metrics.
1. Imagery data integration (Tentative)

- **Tasks:**

0. Feature-engineering opportunity chase.  
(Assigned to: [Meixiang])
1. Design and conduct experiments, perform comparative analyses.  
(Assigned to: [Barbara, Meixiang, Alicia])
2. Hyper-parameter tuning for existing models.  
(Assigned to: [Barbara, Meixiang])
3. Integrate imagery data into the model.  
(Assigned to: [Meixiang(Tentative)])

### **Mar 1-14**

- **Goals:**

0. Incorporate building area data into the pipeline.
1. Estimate GHG emissions using the refined model.

- **Tasks:**

0. Identify building area data for GHG estimation.  
(Assigned to: [Alicia, Jiechen, Barbara, Meixiang])
1. Integrate building area data into the existing pipeline.  
(Assigned to: [Alicia, Jiechen, Barbara, Meixiang])
2. Estimate GHG emissions based on EUI predictions and integrated data.  
(Assigned to: [Alicia, Jiechen, Barbara, Meixiang])

### **Mar 15-28**

- **Goals:**

0. Wrap up and document project results.
1. Finalize GitHub repository, report, and presentation.

- **Tasks:**

0. Finalize and clean the code, ensuring all elements are properly documented and ready for delivery.

*(Assigned to: [Alicia, Jiechen, Barbara, Meixiang])*

1. Create a detailed report of results, methodology, and lessons learned.

*(Assigned to: [Alicia, Jiechen, Barbara, Meixiang])*

2. Develop and finalize the presentation slides, summarizing the key results and methodology.

*(Assigned to: [Alicia, Jiechen, Barbara, Meixiang])*

3. Prepare and practice the presentation, ensuring smooth delivery and clear communication of findings.

*(Assigned to: [Alicia, Jiechen, Barbara, Meixiang])*

- 
- ~~Check here to confirm that you have discussed the plan with your client.~~