Mentox AI Bootcamp

Capstone Project: Electricity Cost Prediction (Regression Task)

Problem Statement

Modern infrastructure depends heavily on efficient energy consumption. Predicting electricity costs can help both residential and commercial sites optimize resource allocation, plan budgets, and identify energy efficiency opportunities.

In this capstone project, you are tasked with building a **regression-based machine learning model** that can accurately predict the monthly electricity cost of a property based on a variety of operational, environmental, and structural features.

This is a real-world scenario where data from multiple sources affect utility costs. Your goal is to perform data analysis, develop predictive models, and deploy your best model using FastAPI on AWS.

About the Dataset

You will be using the following dataset:\ 🔗 [Electricity Cost Prediction Dataset - Kaggle](#)

Key Columns:

- `Site Area (sq meters)`
- `Structure Type` : {Residential, Commercial, Mixed-use, Industrial}
- `Water Consumption (liters/day)`
- `Recycling Rate (%)`
- `Utilisation Rate (%)`
- `Air Quality Index (AQI)`
- `Issue Resolution Time (hours)`
- `Resident Count`
- `Electricity Cost (USD/month)` – **Target variable**

These features provide a mix of direct and indirect influences on energy expenditure. You are expected to analyze and model these relationships to predict electricity cost as accurately as possible.

What You Need To Do

- 🔹 **1. Exploratory Data Analysis (EDA)**

    - Analyze the distribution and correlations of features

• Identify and visualize missing or skewed data
• Understand feature-target relationships

◆ **2. Data Preprocessing**

• Handle missing values
• Encode categorical variables (`Structure Type`)
• Normalize or standardize numerical data
• Optional: Feature engineering

◆ **3. Model Building (Regression)**

• Use algorithm based on your understanding of Problem statement
• Evaluate models using RMSE, MAE, R^2-score
• Tune hyperparameters to improve accuracy

◆ **4. Model Deployment**

• Build a FastAPI application to serve the model
• Deploy the app on AWS (EC2 or Lambda)
• The API should accept input features and return predicted electricity cost

---

Objectives Recap

• Apply EDA and preprocessing techniques to structured data
• Build, train, and evaluate regression models
• Understand feature importance and model interpretability
• Create and deploy a FastAPI Endpoints on AWS
• Build a real-world predictive system with a clean deployment pipeline

---

Submission Guidelines

• **Deadline**: 25th July 2025
• **Mode of Submission**: Upload a **ZIP file** containing:
• Your entire **codebase** (well-organized)

• A detailed **documentation/report** including:

  ◦ EDA insights

  ◦ Feature preprocessing details

  ◦ Model selection reasoning

  ◦ Model performance metrics

- API Endpoints

Evaluation Criteria

Component	Weightage
EDA & Preprocessing	30%
Model Performance & Selection	25%
API Development & Deployment	25%
Documentation & Explanation	20%