**Description**

Carbon emissions significantly contribute to climate change, making monitoring essential for mitigating environmental impact. However, existing high-accuracy reference meters are prohibitively expensive. Chemotronix, a company developing low-cost sensors, aims to build machine-learning models that accurately map sensor readings to CO₂ levels measured by reference meters. This will enable affordable and scalable solutions for tracking carbon emissions globally.

**Overview**

This project focuses on developing a Ridge Regression model to predict CO₂ levels using data from Chemotronix's low-cost sensors. By achieving this, we can bridge the gap between affordability and precision in carbon emission tracking, promoting widespread adoption of low-cost monitoring technologies.

**Objective**

The primary goal of this project is to build an accurate prediction model for CO₂ levels using sensor data from Chemotronix's prototype devices. This breakthrough has the potential to:

* Democratize access to environmental monitoring tools.
* Assist governments and organizations in implementing data-driven policies to curb carbon emissions.
* Promote sustainability by making emission tracking affordable for communities and industries worldwide.

**Data**

The dataset used in this project was sourced from an experiment conducted by Chemotronix. The experiment deployed three prototype devices—Alpha, Beta, and Charlie—to collect sensor readings alongside a high-accuracy reference meter. The dataset consists of multiple sensor readings and environmental parameters, ensuring data diversity and robustness. Data preprocessing included handling missing values, feature selection, and scaling numerical variables for model optimization.

**Model and Approach**

The project follows these key steps:

* Data Preprocessing: Handling missing values, feature engineering, and normalizing sensor readings.
* Exploratory Data Analysis (EDA): Identifying patterns and trends in CO₂ concentration levels.
* Model Selection: The Ridge Regression model was chosen because it can handle multicollinearity and provide stable predictions.
* Evaluation: This project's evaluation metric is Root Mean Squared Error (RMSE), which measures the difference between predicted and actual CO₂ levels.

**Results**

The Ridge Regression model performed well in predicting CO₂ levels. Key findings include:

* Certain sensor readings exhibit strong correlations with CO₂ levels.
* The model reasonably balanced bias and variance, ensuring reliable predictions.
* The test set evaluation metrics indicate the model's generalization ability across different environmental conditions.

**Recommendations**

* Governments and environmental agencies can leverage this technology for large-scale CO₂ monitoring.
* Industries can adopt low-cost sensors for sustainable emission tracking.
* Further research into optimizing sensor calibration and feature selection techniques is encouraged.