

Global Pollution Analysis and Energy Recovery

1. Introduction

This project focuses on analyzing global pollution data to understand environmental trends and predict energy recovery using machine learning techniques.

2. Dataset Description

The dataset contains pollution indicators such as air, water, and soil pollution indices, CO2 emissions, industrial waste, population statistics, and energy recovery data across countries and years.

3. Data Preprocessing

Data preprocessing included handling missing values, encoding categorical variables, and scaling numerical features to improve model performance.

4. Exploratory Data Analysis

EDA was performed using visualizations and correlation analysis to identify patterns and relationships between pollution indicators and energy recovery.

5. Linear Regression Model

A Linear Regression model was built to predict energy recovered (GWh) using pollution and energy-related features. The model was evaluated using MSE, MAE, and R^2 score.

6. Logistic Regression Model

Logistic Regression was applied to classify pollution severity into low, medium, and high categories. Model performance was evaluated using accuracy and confusion matrix.

7. Results and Insights

The analysis revealed that industrial waste and CO2 emissions significantly influence energy recovery. The classification model helps identify regions with severe pollution levels.

8. Conclusion

This project demonstrates how data science and machine learning can be used to analyze environmental data and support sustainable development and informed decision-making.