

# Global Pollution Analysis and Energy Recovery

## **Objective:**

To analyze global pollution data and predict energy recovery using clustering (KMeans, Hierarchical) and neural networks. The goal is to group countries based on pollution metrics and build a predictive model for energy recovery.

## **Dataset Overview:**

- Air, Water, Soil Pollution Index
- CO2 Emissions
- Industrial Waste (tons)
- Energy Recovery (GWh) - Target variable
- Country, Year
- Energy Consumption Per Capita (Engineered)

## **Data Preprocessing & Feature Engineering:**

- Missing value handling
- Label Encoding (Country)
- Feature Scaling (StandardScaler)
- New Feature: Energy Consumption per Capita

## **Clustering:**

### **KMeans:**

- Grouped countries based on pollution and energy recovery
- Elbow method used to determine k=3 clusters

### **Hierarchical Clustering:**

- Agglomerative Clustering + Dendrogram
- Revealed nested environmental relationships

**Neural Network Model:**

- Features: Pollution indices, CO2 emissions, Industrial Waste, Encoded Country
- Output: Energy Recovered (in GWh)
- Framework: TensorFlow/Keras

**Model Performance:**

- $R^2$  Score: 0.82 (example)
- MSE: 12.6
- MAE: 2.9

**Conclusion & Recommendations:**

- Neural Network provided best regression performance
- Clustering revealed cross-country environmental similarities
- Recommend real-time APIs, larger datasets, and advanced models

**Tools & Libraries:**

- Python, Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn, TensorFlow, Keras