Global Pollution Analysis and Energy Recovery

Objective
The goal is to classify countries into different pollution severity categories (Low, Medium, High) based on pollution levels, energy consumption, and other environmental factors.

Phase 1 - Data Preprocessing

- **Data Import and Cleaning**

  - Load the dataset (`Global_Pollution_Analysis.csv`).
  - Handle missing data and outliers using appropriate imputation methods.
  - Normalize/standardize features such as CO2 emissions and industrial waste.
  - Encode categorical features like country and year using LabelEncoder.
- **Feature Engineering**

  - Create new features like energy consumption per capita and yearly pollution trends.
  - Apply feature scaling to pollution indices (air, water, soil pollution).

Phase 2 - Classification using Naive Bayes, K-Nearest Neighbors, and Decision Tree

1. **Naive Bayes Classifier**

   - Implement **Multinomial Naive Bayes** to classify countries based on pollution severity (Low, Medium, High).
   - **Evaluation Metrics:** Accuracy, Confusion Matrix, Precision, Recall, F1-score.
2. **K-Nearest Neighbors (KNN)**

   - Apply the **KNN classifier** for categorizing pollution levels based on pollution indices.
   - **Hyperparameter Tuning:** Determine the optimal number of neighbors (K) for better classification accuracy.
   - **Evaluation Metrics:** Accuracy, Confusion Matrix, Precision, Recall, F1-score.
3. **Decision Tree**

   - Build a **Decision Tree classifier** to classify countries into pollution severity categories.
   - **Hyperparameter Tuning:** Tune the tree using parameters like `max_depth` and `min_samples_split` to improve performance.
   - **Evaluation Metrics:** Accuracy, Confusion Matrix, Precision, Recall, F1-score.

Phase 3 - Reporting and Insights

- **Model Comparison**

  - Compare the performance of the three classifiers based on accuracy, confusion matrix, and other relevant metrics.
  - Visualize the results using confusion matrices and classification reports.
- **Actionable Insights**

  - Highlight key findings about pollution levels in different countries and how they impact energy recovery.
  - Recommend policies for reducing pollution based on model insights.

Final Deliverables

1. Jupyter Notebook (.ipynb) containing the entire code and analysis.
2. Data Visualizations in image format or embedded in the notebook.
1. Final Report summarizing key findings, model evaluations, and actionable recommendations.