GLOBAL POLLUTION ANALYSIS AND ENERGY RECOVERY MODEL TRAINNING

By : Anusruta Mohanty

October 22, 2025

Project Link : https://github.com/anusruta/Global-Air-Pollution-Analysis-and-Model-Training/blob/main/naivebayes_knn_decisiontree.ipynb

OVERVIEW

1. Project background and description

Environmental pollution varies widely across countries due to differences in industrial activity, energy consumption, population density, and regulatory frameworks. The dataset collected captures key indicators such as **CO_2 emissions, air, water, and soil pollution indices, industrial waste, energy consumption per capita, and other environmental factors** for multiple countries and years.

Analyzing this data allows us to identify countries with **low, medium, or high pollution severity**, providing a clear understanding of global environmental challenges. Traditional assessments often rely on isolated metrics, but this dataset enables a **comprehensive, multi-dimensional view** of pollution severity.

Machine learning techniques such as **Naive Bayes, K-Nearest Neighbors (KNN), and Decision Trees** can be applied to classify countries accurately based on these indicators. Feature engineering, including metrics like energy consumption per capita and pollution trends over time, enhances predictive accuracy.

2. Project scope

Policy Guidance:
Governments can use the classification to focus interventions on high-risk countries or regions.

Resource Allocation:
International agencies can prioritize funding countries facing severe pollution challenges.

Environmental Awareness:
Data-driven insights help inform public campaigns, regulatory measures, and sustainability initiatives.

Early Warning:
Countries at risk of increasing pollution severity can be identified for proactive measures.

3. Deliverables

Using the given dataset, this project aims to:

1. **Preprocess Data:**
    a. Handle missing values and outliers.
    b. Normalize pollution indices and energy consumption.
    c. Encode categorical features like country and year.
2. **Feature Engineering:**
    a. Compute derived metrics such as energy consumption per capita.
    b. Identify yearly trends in pollution indices.
3. **Model Development:**
    a. Apply **Multinomial Naive Bayes**, **KNN**, and **Decision Tree** classifiers.
    b. Tune hyperparameters like K in KNN and max_depth/min_samples_split in Decision Trees to improve performance.
4. **Evaluation:**
    a. Measure performance using **Accuracy, Precision, Recall, F1-score, and Confusion Matrix**.

4. Affected parties

The project will affect the residents in high-pollution areas, governments and policymakers, industries and businesses, and international Organizations.

5. Key Findings

1. Pollution Levels Across Countries

- Countries with **high industrial activity** and **energy consumption per capita** tend to fall into the **High Pollution Severity** category.
- **Air pollution** is typically higher in countries with dense urbanization and heavy vehicular traffic, while **water and soil pollution** are more influenced by industrial waste management practices.
- **Low Pollution Severity** countries generally have **stricter environmental regulations**, higher adoption of **renewable energy**, and lower industrial emissions.
- **Medium Pollution Severity** countries show a transitional pattern where certain pollution indices are high but overall environmental management is improving.

2. Impact on Energy Recovery

- High pollution levels **reduce efficiency in energy recovery systems** such as waste-to-energy plants. Contaminants in air, water, and industrial byproducts can:
    - Reduce the quality of biomass or waste fuels.
    - Increase maintenance costs due to equipment corrosion or clogging.
    - Limit adoption of renewable energy technologies if environmental conditions are poor.
- Countries with **lower pollution severity** demonstrate **better integration of clean energy and recovery systems**, indicating that effective pollution control directly supports sustainable energy utilization.

6. Policy Recommendations Based on Model Insights

1. Targeted Industrial Regulation

- Enforce **emission limits** for industries in high-severity countries.
- Encourage adoption of **clean technologies and pollution control devices** (scrubbers, filters, wastewater treatment).

2. Promote Renewable and Clean Energy

- Incentivize **solar, wind, and biomass energy** to reduce reliance on fossil fuels.
- Support **energy recovery from waste** by ensuring waste streams are clean and standardized.

3. Pollution Monitoring and Reporting

- Implement **continuous environmental monitoring** using sensors and IoT-based solutions.
- Publicly report pollution levels to improve transparency and accountability.

4. Waste Management Improvements

- Encourage **recycling, composting, and proper industrial waste disposal**.
- Reduce **soil and water contamination** to protect energy recovery efficiency.

5. Urban and Transport Planning

- Promote **green urban planning**: expand green spaces, reduce traffic congestion, and develop efficient public transport.
- Incentivize **electric and hybrid vehicles** in highly polluted cities.

6. International Collaboration

- High-pollution countries can **learn from low-pollution countries** through technology transfer and policy frameworks.
- Global organizations can provide **technical and financial support** to implement sustainable pollution reduction strategies.

7. Model Evaluation

The decision tree classification yielded the best results with an accuracy score of 0.875, followed by KNN and Naive Bayes classification with accuracy score of 0.675.

Credits :- https://www.kaggle.com/datasets/hasibalmuzdadid/global-air-pollution-dataset