Pollution Severity Classification Report

Objective

To classify countries into different pollution severity categories — Low, Medium, High — based on factors such as pollution levels, energy consumption, and other environmental parameters.

Models Evaluated

Three models were implemented and compared:

1. Naive Bayes Classifier

2. K-Nearest Neighbors (KNN)

3. Decision Tree Classifier

Model Performance Summary

Model	Accuracy	Precision	Recall	F1 Score	Key Issue
Naive Bayes	75%	0.90 (weighted)	0.75	0.81	Poor "Medium" class detection
KNN (Best K=2)	93.3%	0.87	93.3%	0.90	Does not detect "Medium" class
Decision Tree	93.3%	0.87	93.3%	0.90	Same issue as KNN

Detailed Model Analysis

Naive Bayes

Accuracy: 75%

Confusion Matrix:
[[43 13]
 [2 2]]

High Precision (0.96) but Low Recall (0.77) for the "High" class.
Poor performance on the "Medium" class (Precision: 0.13, Recall: 0.50).

Insight: The model struggles with class imbalance and misclassifies many "Medium" countries as "High".

⬜ K-Nearest Neighbors (K=2)
Accuracy: 93.33%

Confusion Matrix:
[[56 0]
 [4 0]]

Insight: Very high precision/recall for "High", but fails to generalize or detect any variation.

⬜ Decision Tree
Accuracy: 93.33%

Confusion Matrix:
[[56 0]
 [4 0]]

Insight: Shows high performance but fails in multi-class discrimination.

⬜ Actionable Insights

1. Class Imbalance:
   - All models perform poorly on the "Medium" category due to fewer samples.
   - Oversampling techniques like SMOTE or class weighting should be considered.

2. Dominance of High Pollution Countries:
   - A majority of data points belong to "High" pollution severity, indicating a global crisis needing attention.

3. Energy Consumption Link:
   - High pollution countries likely show high energy usage from non-renewable sources.
   - Promoting renewable energy and clean technologies can reduce pollution.

▯ Policy Recommendations

1. For High Pollution Countries:
   - Enforce stricter emission norms and invest in pollution control tech.
   - Incentivize industries to adopt clean energy.

2. For Medium Pollution Countries:
   - Focus on early intervention to prevent them from sliding into the "High" category.
   - Introduce targeted subsidies for eco-friendly practices.

3. Global Strategy:
   - Promote data-driven energy policies.
   - Create international platforms to share green innovations and monitoring techniques.

✅Conclusion

While KNN and Decision Tree showed high overall accuracy, they fail to differentiate between all classes.
Naive Bayes shows better multiclass recognition but needs improvement through data balancing.
Future work should focus on:
   - Class rebalancing,
   - Feature engineering,
   - Ensemble techniques (like Random Forest, XGBoost) for better generalization.