**Model Evaluation and Insights Report**

This report presents the evaluation of Linear Regression and Logistic Regression models applied on the dataset to analyze the relationship between pollution indicators and energy recovery. The study compares the predictive capabilities of these models and extracts actionable insights for environmental and energy management.



Correlation heat map

**Correlation of pollution index with different variables:-**

CO2\_Emissions: 0.0017853303399134996

Industrial Waste: 0.10797764404925328

Population: 0.06415928498641145

Renewable Energy: -0.07251559189030143

Plastic Waste Produced: 0.09253671244285885

**Linear Regression Results:**

* Features: Air Pollution Index, Industrial Waste (in tons), CO2 Emissions (in MT)
* Target: Energy Recovered (in GWh)

Performance Metrics:

* R²: 0.017
* Mean Squared Error (MSE): 22,131.47
* Root Mean Squared Error (RMSE): 148.77
* Mean Absolute Error (MAE): 131.38

**Inference:**

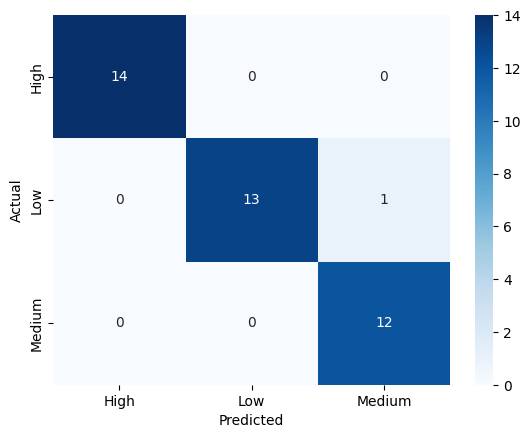
The R² score close to zero suggests that the model is unable to explain the variance in energy recovery based on the given features. Despite a moderate MAE, the large RMSE indicates high variability and low predictive accuracy. Thus, linear regression is not suitable for predicting energy recovery in this dataset.

**Logistic Regression Results:**

* Features: Air Pollution Index, Industrial Waste (in tons), CO2 Emissions (in MT)
* Target: Pollution Level (categorized as Low, Medium, High)

Performance Metrics:

* Accuracy: 0.975
* Precision: 0.977
* Recall: 0.975
* F1 Score: 0.975



**Inference:**

The logistic regression model demonstrates high performance across all metrics, with an accuracy of 97.5%. The classification report highlights that all pollution categories are well distinguished, especially the "High" category which shows perfect precision and recall. This indicates the model is highly reliable for categorizing pollution levels.

**Model Comparison:**

- Linear Regression fails to predict energy recovery effectively, indicating weak linear relationships between pollution indicators and energy recovery.

- Logistic Regression, however, shows excellent classification ability for pollution levels, making it more suitable for categorical predictions in this dataset.

**Actionable Insights:**

1. **Impact of Pollution on Energy Recovery:** Since linear regression performed poorly, it indicates that multiple complex, non-linear factors beyond pollution indicators alone influence energy recovery.

2. **Pollution Categorization:** Logistic regression results suggest pollution levels can be predicted

reliably, enabling policymakers to identify high-risk regions.

3. **Recommendations:**

* Focus on reducing industrial waste and CO2 emissions to shift countries from "High" to "Medium" or "Low" pollution levels.
* Invest in advanced waste-to-energy technologies that can enhance recovery irrespective of pollution fluctuations.
* Countries in "High" pollution categories could benefit from stricter regulations and sustainable practices.