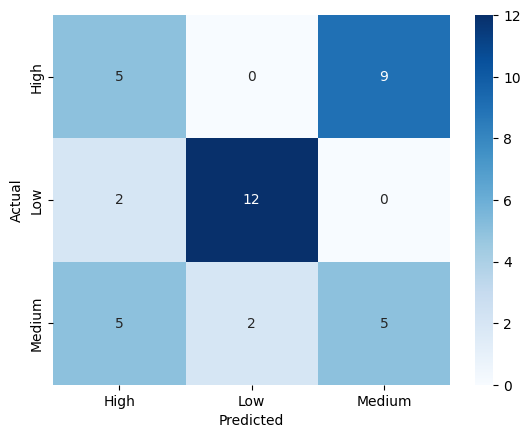
**Reporting and Insights for Global Pollution**

**Model Performance Comparison**

1. **Multinomial Naive Bayes**

**Accuracy**: 55%

* Performed well in classifying **Low pollution severity** (Precision = 0.86, Recall = 0.86, F1 = 0.86).
* Struggled with distinguishing **High** and **Medium** categories, leading to significant misclassifications.
* Indicates that Naive Bayes, with its assumption of feature independence, was not able to capture complex relationships in the data.

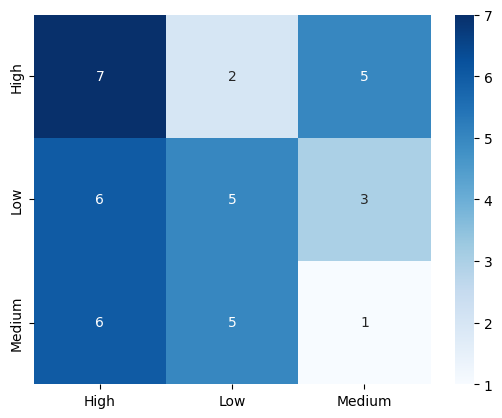


1. **K-Nearest Neighbors (KNN)**

**Best K**: 9

**Accuracy**: 32.5% (lowest among the three models)

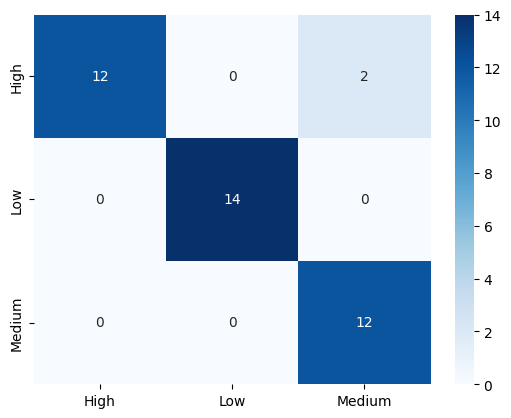
* Poor generalization across all classes, particularly **Medium**, which had very low precision (0.11) and recall (0.08).
* Indicates sensitivity to data distribution and possibly overlapping feature spaces.
* Suggests that KNN is not suitable for this dataset without significant feature scaling or dimensionality reduction.



1. **Decision Tree**

**Accuracy**: 95% (highest among the models)

* Achieved near-perfect classification across all categories.
* **Low** category was classified with 100% precision and recall.
* Slight misclassifications between **High** and **Medium**, but overall performance was balanced and robust.
* Hyperparameter tuning (max\_depth and min\_samples\_split) effectively reduced overfitting while maintaining strong predictive power.



Key Insights and Recommendations

The analysis reveals that Naive Bayes, while computationally efficient, tends to oversimplify relationships and is therefore better suited for datasets with stronger feature independence. KNN, on the other hand, performs poorly on this dataset due to class overlap and unclear neighbourhood boundaries, making it unsuitable in its current form. The Decision Tree stands out as the most accurate and interpretable model, achieving 95% accuracy and offering clear insights into feature importance, which makes it the most appropriate choice for this classification task. Across models, Low pollution severity is consistently predicted with high accuracy, while differentiating between High and Medium severity proves more challenging.

For practical deployment, the Decision Tree should be prioritised due to its strong accuracy, interpretability, and balanced performance across classes. Naive Bayes may serve as a lightweight baseline, but it is unlikely to meet real-world requirements. KNN is not recommended unless further preprocessing techniques such as PCA, feature engineering, or feature scaling significantly improve its performance. The model’s reliability in identifying low-severity regions provides an opportunity to shift focus toward High- and medium-severity zones, where misclassifications are more frequent. Policymakers and organizations can leverage these insights to design targeted interventions in areas classified with high pollution severity, thereby improving environmental management and resource allocation.