**Accuracy Comparison**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Accuracy** | KNN | SVM | Naïve Bias | Logistic Regression | Random Forest | XG Boost |
| Dataset 1 | 74.68% | 77.27% | 80.52% | 74.68% | 73.38% | 90.07% |
| Dataset 2 | 96.45% | 96.14% | 89.98% | 96.13% | 97.14% | 97.23% |

**Dataset 1:**

* **XG Boost (90.07%)**: This algorithm performs significantly better than the others, indicating it is well-suited for the patterns and features in Dataset 1. XGBoost is known for its ability to handle non-linear relationships and interactions between features effectively.
* **Naïve Bayes (80.52%)**: The second-best algorithm, which suggests that the dataset may have independent or conditionally independent features that Naïve Bayes leverages well.
* **SVM (77.27%)**: Performs moderately well, likely due to its ability to separate classes using a hyperplane, but it might not be capturing complex patterns as effectively as XG Boost.
* **KNN and Logistic Regression (74.68%)**: Their performance indicates they struggle to generalize well, possibly due to the nature of the dataset.
* **Random Forest (73.38%)**: Surprisingly lower performance might suggest overfitting or insufficient feature importance ranking for this dataset.

**Best Algorithm for Dataset 1**: **XG Boost**  
**Worst Algorithm for Dataset 1**: **Random Forest**

### **Dataset 2:**

* **XG Boost (97.23%)**: Again the best-performing algorithm, which reinforces its ability to generalize and capture complex patterns in the data.
* **Random Forest (97.14%)**: A close second, showing its ensemble decision tree structure works effectively for this dataset.
* **KNN (96.45%)** and **SVM (96.14%)**: Both perform well, indicating that simpler methods like SVM and distance-based KNN are sufficient for this dataset.
* **Logistic Regression (96.13%)**: Almost on par with KNN and SVM, showing that the dataset is linearly separable or has a simple structure.
* **Naïve Bayes (89.98%)**: Falls significantly behind, likely due to feature dependency violating the independence assumption.

**Best Algorithm for Dataset 2**: **XG Boost**  
**Worst Algorithm for Dataset 2**: **Naïve Bayes**

### **Conclusion:**

* **XG Boost** is the best-performing algorithm across both datasets, indicating its robustness and suitability for diverse datasets. It is highly effective for datasets with complex patterns and relationships.
* **Naïve Bayes** struggles with Dataset 2, likely due to feature dependency.
* **Random Forest**, although close in Dataset 2, underperforms in Dataset 1, possibly due to its sensitivity to feature selection or overfitting.