

Model Comparison Notes

This document summarizes the performance of ten classification models based on various metrics and provides guidance on when to use each model based on dataset characteristics.

Model Performance Summary

Model	Test Score	Accuracy	Precision (False)	Precision (True)	Recall (False)	Recall (True)	F1-Score (False)	F1-Score (True)
OVO	0.9586	0.96	0.94	0.98	0.98	0.94	0.96	0.96
Random Forest (RF)	0.9579	0.96	0.94	0.98	0.98	0.94	0.96	0.96
KNN	0.9158	0.92	0.97	0.88	0.86	0.97	0.91	0.92
XGBoost	0.9415	0.94	0.93	0.95	0.95	0.93	0.94	0.94
Bagging	0.9436	0.94	0.92	0.97	0.97	0.91	0.95	0.94
Decision Tree (dt)	0.9143	0.91	0.93	0.90	0.90	0.93	0.91	0.92
SGD	0.8665	0.87	0.87	0.87	0.87	0.87	0.87	0.87
Logistic Regression (Lgr)	0.8494	0.85	0.87	0.83	0.83	0.87	0.85	0.85
SVC	0.8951	0.90	0.86	0.94	0.94	0.85	0.90	0.89
Bagging Classifier (baf_clf)	0.8915	0.89	0.83	0.97	0.98	0.80	0.90	0.88

Model Recommendations

1. OVO Model

Use when you need high accuracy and balanced performance across classes, especially in binary classification tasks.

2. Random Forest Model

Ideal for datasets with a mix of categorical and numerical features; provides robust performance with interpretability.

3. KNN Model

Consider for smaller datasets with low dimensionality where interpretability is important; ensure proper parameter tuning.

4. XGBoost Model

Best for structured data and when you require high performance with flexibility in hyperparameter tuning.

5. Bagging Model

Use for models prone to overfitting; good for increasing accuracy on complex datasets.

6. Decision Tree Model

Useful for interpretability and simplicity; can be used for both classification and regression tasks.

7. SGD Model

Consider for large-scale datasets with many features where computational efficiency is essential.

8. Logistic Regression Model

Ideal for linearly separable datasets or when you need a baseline model for binary classification.

9. SVC Model

Use when you have a smaller dataset and require high accuracy, especially with a clear margin of separation.

10. Bagging Classifier Model

Good for high-variance models; can help in improving stability and accuracy.