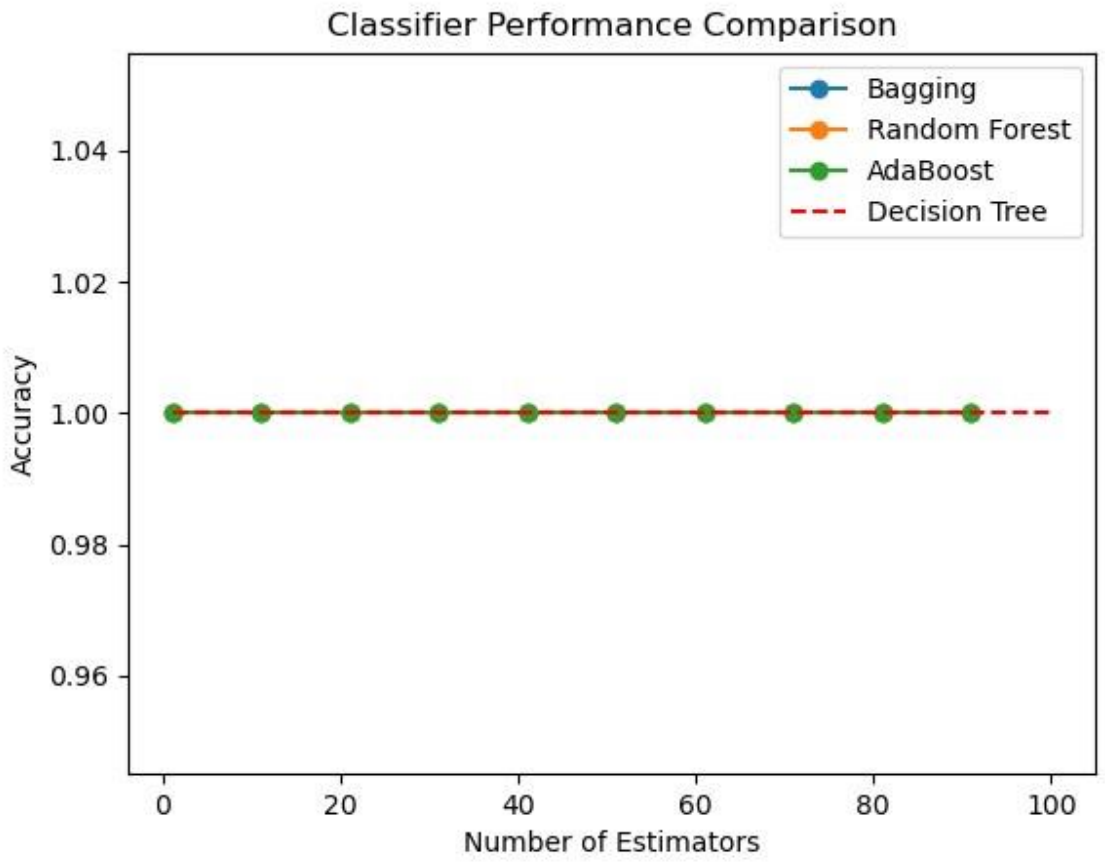
ENSEMBLE CLASSIFIERS PERFORMANCE ACROSS VARYING ESTIMATOR COUNTS ON THE IRIS DATASET

SUBMITTED BY: THANUJ KUMAR SHIVALINGAIAH (1002136007)

The graph presents a comparative analysis of four classification algorithms: Decision Tree, Bagging, Random Forest, and AdaBoost, with respect to their predictive accuracy on the Iris dataset as the number of estimators varies.



The Decision Tree, shown with a red dashed line, exhibits a constant accuracy level, which aligns with expectations since the Decision Tree algorithm does not utilize a collection of estimators but rather makes predictions based on a singular decision tree structure.

In contrast, the Bagging, Random Forest, and AdaBoost algorithms, depicted with blue, orange, and green lines respectively, incorporate multiple estimators. The Bagging algorithm constructs multiple decision trees independently and integrates their outcomes. Random Forest, a more complex variant of Bagging, incorporates randomness in the tree generation process to create a "forest" of trees, using random subsets of features. AdaBoost, short for Adaptive Boosting, focuses on difficult-to-classify instances by sequentially refining the decision boundaries.

The accuracy scores for these ensemble methods are plotted against a range of estimator values from 1 to 100. Interestingly, the accuracy for Bagging, Random Forest, and AdaBoost remains relatively stable across the number of estimators. One might anticipate increased accuracy with more estimators due to reduced variance from averaging predictions, yet the plot does not exhibit this trend.

ENSEMBLE CLASSIFIERS PERFORMANCE ACROSS VARYING ESTIMATOR COUNTS ON THE IRIS DATASET

SUBMITTED BY: THANUJ KUMAR SHIVALINGAIAH (1002136007)

Several factors could contribute to this observation:

1. The Iris dataset's low complexity and small size may not provide enough variance between subsets of data to challenge the robustness of ensemble methods.
2. The base decision trees could be highly effective for this dataset, making additional estimators superfluous.
3. A ceiling effect may be in place where the classifiers have reached maximum predictive power, beyond which further improvements are negligible.

In summary, while each algorithm typically benefits from ensemble strategies, the Iris dataset does not showcase this advantage, likely due to its simplicity. These findings suggest that for certain datasets, especially those with clear patterns and low noise, simple models may suffice, and complex ensemble strategies might not yield significant improvements in performance.