

Question 5 : Gradient Tree Boosting

Solution:

5.1 Abalone Data

5.1a Abalone Raw

The accuracy of Gradient Tree Boosting on training Abalone Raw with default parameters is 25.45%, better than Random Forest. After tuning with GridSearchCV accuracy comes up to 26.60 % with best parameters as 'max_depth': 5, 'max_leaf_nodes': 4, 'n_estimators': 7 and runtime as 0.556 seconds, which is more than the run time of random forest. **Abalone Raw accuracy 26.60 %, 'max_depth': 5, 'max_leaf_nodes': 4, 'n_estimators': 7.** For Abalone Raw, random forest has a better performance and runtime than gradient tree boosting.

5.1b Abalone PCA

The accuracy of Gradient Tree Boosting on training Abalone PCA with default parameters is 23.83%, better than Random Forest. After tuning with GridSearchCV accuracy comes up to 23.83 % with best parameters as 'max_depth': 4, 'max_leaf_nodes': 4, 'n_estimators': 5 and runtime as 0.392 seconds, which is more than the run time of random forest. **Abalone Raw accuracy 23.72 %, 'max_depth': 4, 'max_leaf_nodes': 4, 'n_estimators': 5.** For Abalone PCA, random forest has a better performance and runtime than gradient tree boosting.

5.1c Abalone LDA

The accuracy of Gradient Tree Boosting on training Abalone LDA with default parameters is 23.25%, less than Random Forest. After tuning with GridSearchCV accuracy comes up to 26.70 % with best parameters as 'max_depth': 2, 'max_leaf_nodes': 16, 'n_estimators': 10 and runtime as 0.815 seconds, which is more than the run time of random forest. **Abalone Raw accuracy 26.70 %, 'max_depth': 2, 'max_leaf_nodes': 16, 'n_estimators': 10.** For Abalone LDA, random forest has a better runtime than gradient tree boosting however the accuracy of gradient is more .

5.2 Wine Data

5.2a Wine Raw

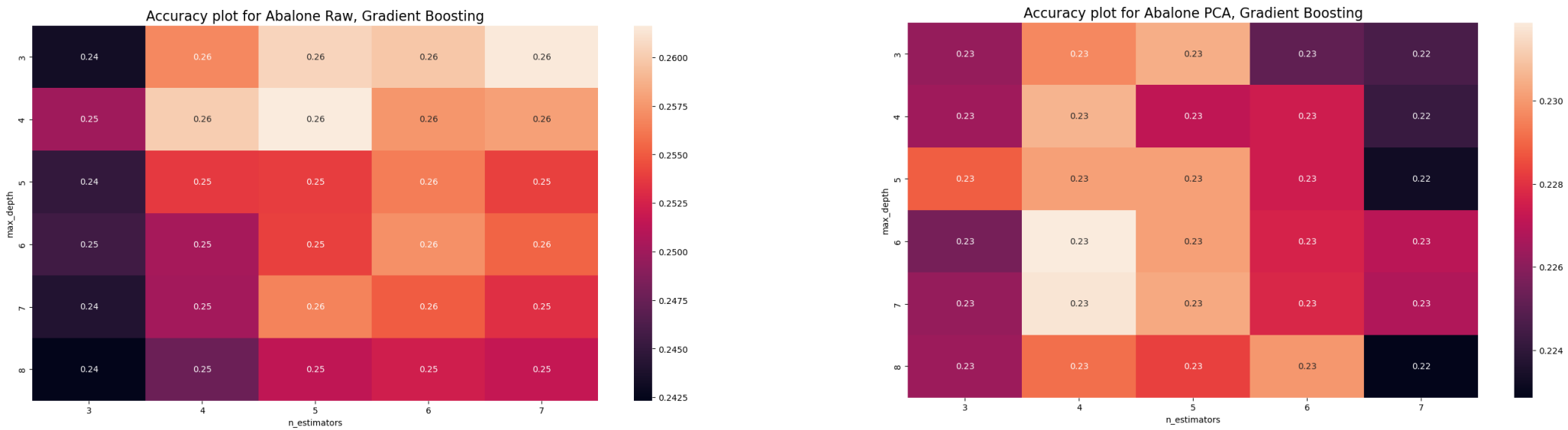
The accuracy of Gradient Tree Boosting on training Wine Raw with default parameters is 59.45%, better than Random Forest. After tuning with GridSearchCV accuracy comes up to 59.57% with best parameters as 'max_depth': 6, 'max_leaf_nodes': 32, 'n_estimators': 10 and runtime as 0.956 seconds, which is more than the run time of random forest. **Abalone Raw accuracy 59.57%, 'max_depth': 6, 'max_leaf_nodes': 32, 'n_estimators': 10.** For Wine Raw, random forest has a better performance and runtime than gradient tree boosting.

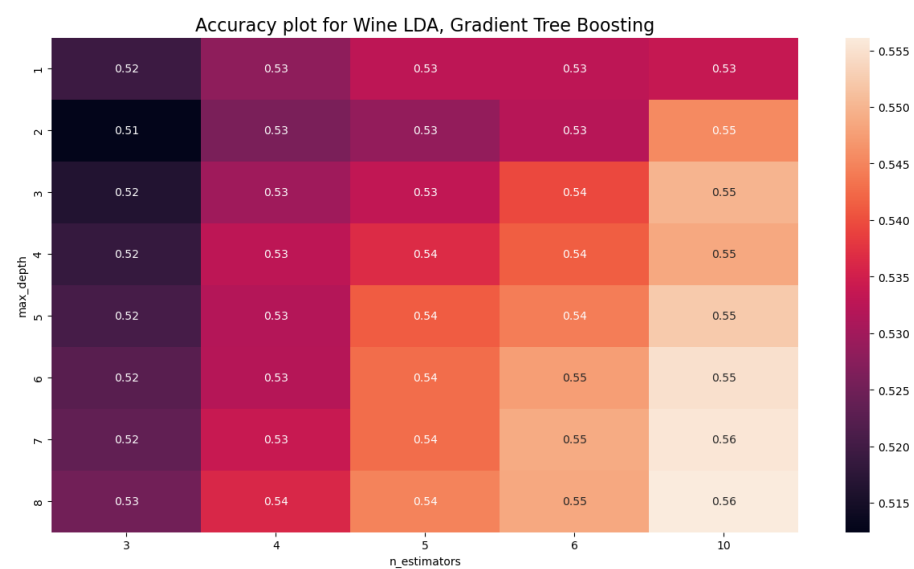
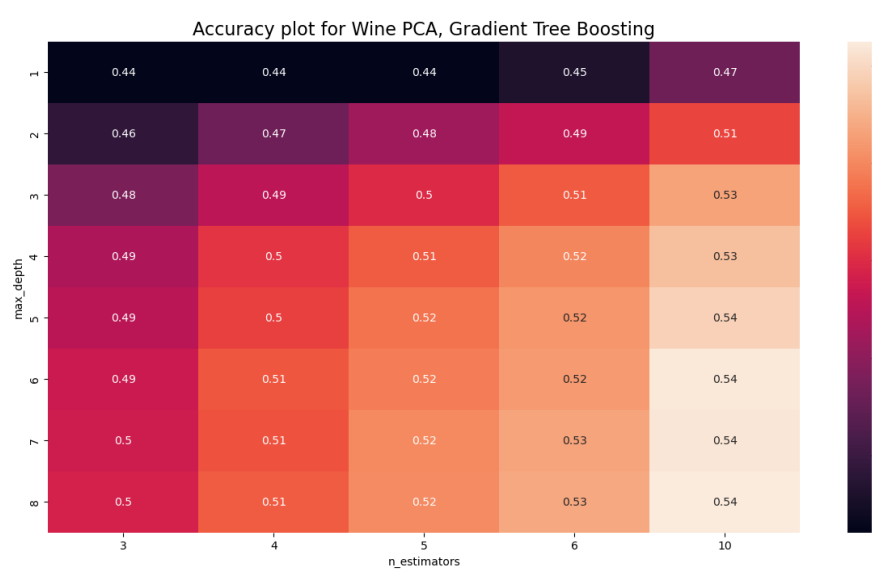
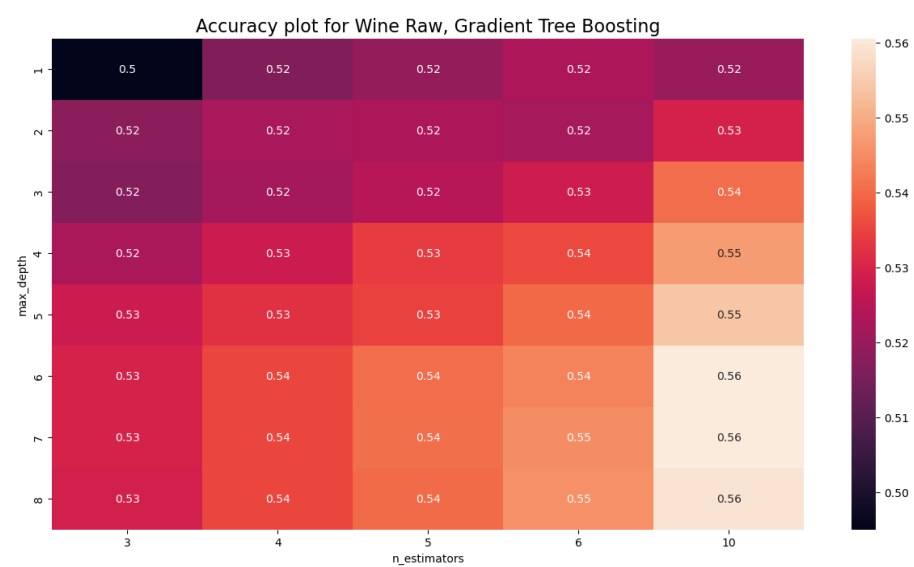
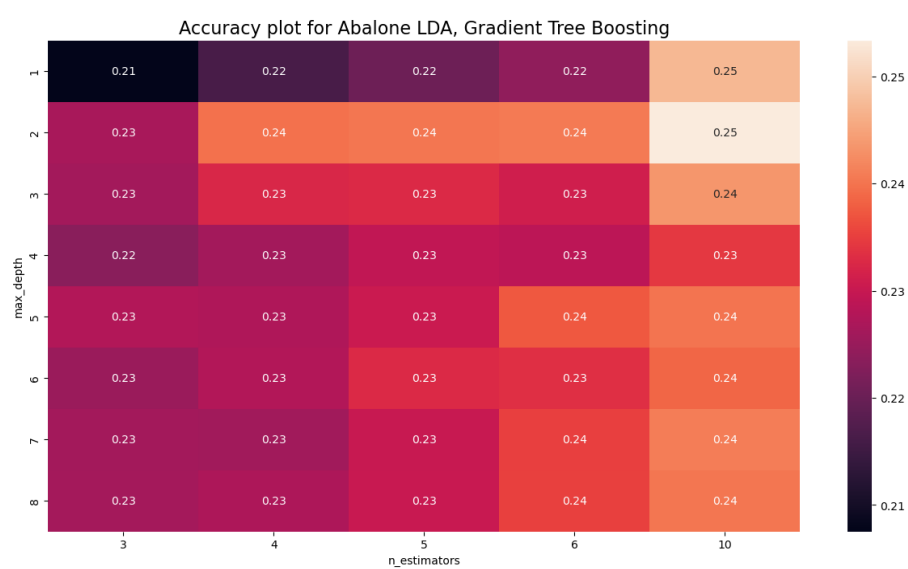
5.2b Wine PCA

The accuracy of Gradient Tree Boosting on training Wine PCA with default parameters is 57.48%, better than Random Forest. After tuning with GridSearchCV accuracy comes up to 56.55% with best parameters as 'max_depth': 7, 'max_leaf_nodes': 32, 'n_estimators': 10 and runtime as 1.604 seconds, which is more than the run time of random forest. **Abalone Raw accuracy 56.55 %, 'max_depth': 7, 'max_leaf_nodes': 32, 'n_estimators': 10.** For Wine PCA, random forest has a better performance and runtime than gradient tree boosting.

5.2c Wine LDA

The accuracy of Gradient Tree Boosting on training Wine LDA with default parameters is 57.05%, less than Random Forest. After tuning with GridSearchCV accuracy comes up to 57.97% with best parameters as 'max_depth': 8, 'max_leaf_nodes': 32, 'n_estimators': 10 and runtime as 0.723 seconds, which is more than the run time of random forest. **Abalone Raw accuracy 57.97%, 'max_depth': 8, 'max_leaf_nodes': 32, 'n_estimators': 10.** For Wine LDA, random forest has a better performance and runtime than gradient tree boosting.





Overall, Random Forest has better performance and less runtime than Gradient Tree Boosting in both the datasets and over both the dimensionality reduction techniques applied on the datasets.