Packages that were used

Sklearn

Numpy

Pandas

Matlabplot

Purpose of this project:

Get familiar with the usage of ensemble methods and parameters that generate the highest accuracy.

Questions to Answer:

1.Write your observations about the Classifier’s behavior with respect to the

number of estimators.

Based on what I observed, the accuracy value is lower when the n\_estimator value is too high or low.

2.

Is there an optimal value of the estimator within the given range?

The optimal value usually lies in the middle around 200 for each of these classifiers. I looked for the optimal value by looking at the mean of the highest point for each classifier and got 181 n\_estimator.

Task 1:

Bin and categorize census data similarly to what I did in CA03.

Task 2:

Find the optimal value of a parameter. Optimal value is commented in the ipynb 8 n\_estimators

Task 3:

Find the accuracy for random forest with n estimator values

[0,100,150,200,250,300,350,400,450,500]

Task 4:

Repeat the same process for Random forest, Gradient boost, and XGB

Task 5:

Input the accuracy and AUC value into a model based on the classifier