Donovan Moini, Ian Lizarda, Tyler Edmiston

Within our preprocessing, we began by replacing the missing data, denoted as "?" within the training and test file, with the most frequent value within each respective feature via SimpleImputer. For the continuous data, we chose to sort the data within bins. We created 3 bins for ages, 5 bins for the number of years in education, 3 bins for capital gain, 3 bins for capital loss, and 3 bins for hours per week worked. With our Naive Bayes Classifier, we did not manually set any hyperparameters. After running naive_bayes.py ten times, the average classification accuracy is 80.48%. After running d_tree.py ten times, the average classification accuracy is 82.62%. After running random_tree.py ten times, the average classification accuracy is 83.65%.