Michael Erwin

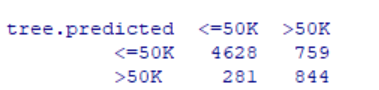
Data Analytics

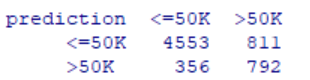
Exercise 11

Classification

The models I used for this exercise were the decision tree and the naïve Bayesian model. I used an 80/20 split of the data for the training and test set, respectively. It appeared that for this data set, the decision tree classifier is the better model.

The decision tree classifier had an accuracy of about 84-85%, whereas the naïve Bayesian model only had an accuracy of about 83%.





Confusion matrix for decision tree classifier Confusion matrix for naïve Bayesian model

Above is the confusion matrices for the decision tree classifier and the naïve Bayesian model. The decision tree had more True Positives and True Negatives, which we expected since we determined that the decision tree was the better classifier for this data set. The decision tree also had less false positives and false negatives, which was also expected since the decision tree had a better accuracy.