

John Lopez

Classification Report

For all of the following methods, I applied very similar preprocessing methods. First, I ran each of the features (except for age) through an ordinal encoder to ensure they all had an ordinal encoding. Next, I removed any data points that were marked with unknown data. Finally, I classified all of the “age” categories into bins of 3, from the ages of 0-21, 21-65, and 65 onward.

When I applied the Naive Bayes Classifier to my data, I received a 77.6% accuracy on my test data. I applied no extra parameters or preprocessing methods.

When I applied a Decision Tree Classifier to my data, I received a 79.5% accuracy on my test data. I set the seed of the random state generator to 5.

When I applied the Random Forest Classifier to my data, I received an 85.3% accuracy on my test data. I believe this was due to the fact that I had 69 estimators, and I increased the max depth of the classifier to be 10. By increasing both values, I improved my accuracy