***Obtaining an estimated roc – auc score.***

This Kaggle competition uses a method called “Receiver Operator Characteristic – Area Under Curve” to determine the precision and accuracy (or sensitivity and specificity) of the method used in prediction.

Scikit-Learn includes an roc and roc-auc method based on the true and predicted dependent variables, which is great. We have written a method called return\_roc\_auc() that reads in and applies a transform to the directory passed and then divides that data (either using a straight cross-validation, or you can use a 5-fold) and trains the specified classifier on some data and predict on the rest. Then we can see where the pipeline predicted correctly and where it didn’t.