**Long Story Short of the SIE 533 Project**

Thus far we’ve tried all 3 methods we talked about in our proposal: support vector classifiers, random forests, and neural networks. The dataset authors note that simple accuracy isn’t a good scoring method due to real-life class imbalance and recommend the receiver operating characteristic curve (ROC). The ROC is nice because the area under the curve (AUC) is intuitively the likelihood the classifier will correctly classify a random datapoint (that is not in the training data) correctly. The classifiers have an AUC as follows:

**SVC – 0.90**

**RDC – 0.93**

**NN—0.92**