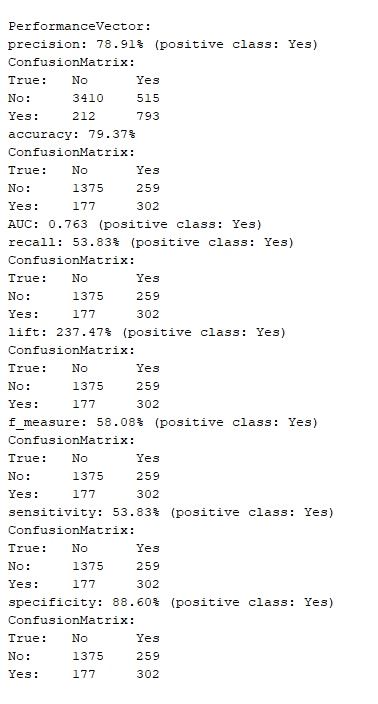
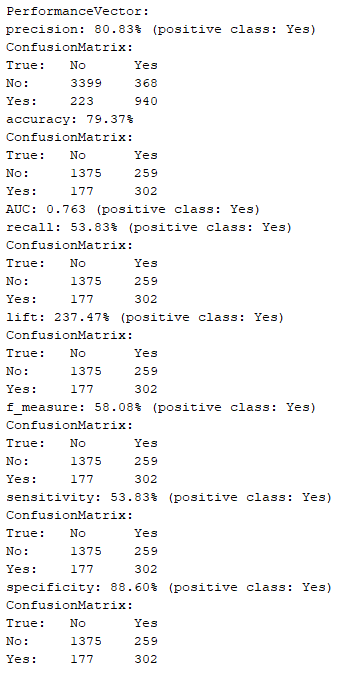
Nick Iudiciani

MIST4060 – Data Mining for Business Intel

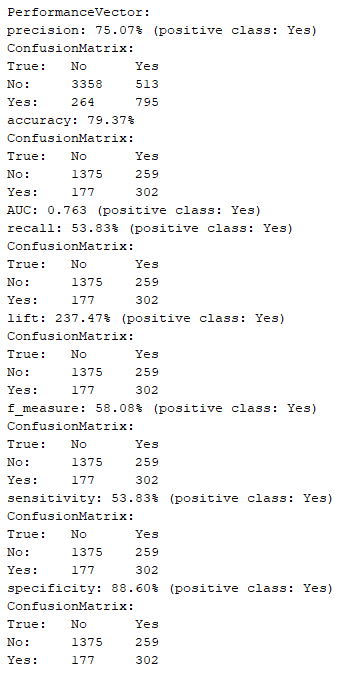
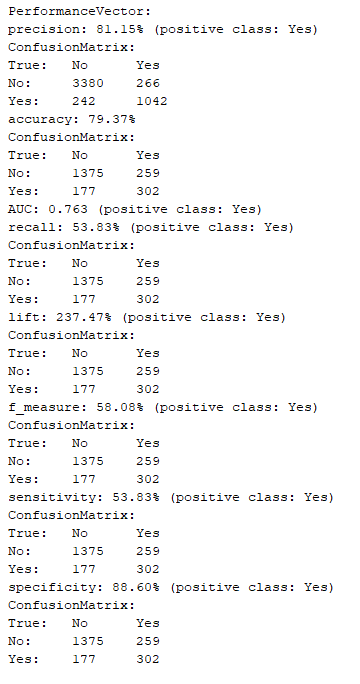
Prof. Amit Deokar

Homework Assignment 3

1a) Random Forest Performance Vectors 1b)Boosting Performance Vectors



1c) Bagging Performance Vectors 1d) Voting Performance Vectors



2a) Accuracy – Random Forest, Boosting, Bagging, Voting

2b) F measure – Random Forest, Boosting, Bagging, Voting

2c) Precision – Bagging, Boosting, Random Forest, Voting

2d) Recall/sensitivity – Random Forest, Boosting, Bagging, Voting

2e) AUC – Random Forest, Boosting, Bagging, Voting

3) For accuracy, f measure, recall or sensitivity, and AUC, the models show the same values, therefore any model would be suitable when focusing on those metrics. Precision however, is highest when using the bagging model as opposed to the others. This means that bagging shows the highest rate of true positive predictions over all positive predictions for this data set.

4) For this particular data set, the given metrics show that accuracy, precision, AUC, and specificity should be used (all above 75%). The ones that should not be used would be recall, f measure, and sensitivity (the three lowest metrics, below 60%). Lift should also be used as a performance metric because a lift of 237% shows that the probability of choosing a positive example from all positive predictions is almost 2.5 times the probability of choosing such an example from all of the examples.