

Michael Miller  
Akshay Shetty  
CS 498 – HW3  
2/22/16

**Problem a.**

The performance of the classifiers is summarized in the table below.

Classifier	Training Set	Validation Set 1	Validation Set 2	Validation Set 3
Linear SVM	.522	.5	.502	.497
Naive Bayes	.6	.502	.499	.501
Random Forest	.657	.636	.642	.639

**Problem b.**

Matcher accuracy:

Training Set	Validation Set 1	Validation Set 2	Validation Set 3
1.0	1.0	.995	.995

**Problem c.**

Since the random forest algorithm produced the best results in part a, it was selected to use for part c. The following optimizations were made to the algorithm:

1. Data was scaled with mean 0 and standard deviation 1
2. A larger sample size was used to create the random forest

Training Set	Validation Set 1	Validation Set 2	Validation Set 3
.742	.715	.716	.715