

1. Average Accuracy

	part a	part b	part c	part d
test1	0.79085	0.73856	0.7778	0.81163
test2	0.705882	0.79085	0.7647	0.76817
test3	0.803921	0.730261	0.7516	0.7851
test4	0.771242	0.745098	0.7386	0.7651
test5	0.712418	0.720458	0.7582	0.79085
Average	0.756863	0.745045	0.75818	0.78417

2.

The support vector machine works best. The second best is the Naïve Bayes classifier using Klar and Caret. The third best is the first one. The worst one is substituting NA for zeros when using Naïve Bayes classifier. The reason why second one is worse than the first one is that the normal distribution model we choose is not accurate enough in this problem. After looking through the data, we find that there are many missing values for these four attributes. Therefore, the original highly skewed distribution is shifted a lot to the right, which causes the drop of accuracy. In theory, the first one should be better than the second one. The third one is better than the first two since caret may use better distribution model and it also uses cross-validation. The last one is the best because the Naïve Bayes classifier assumes that each class is totally conditionally dependent on each other which might not be true in this case.