

CSCI/ARTI 8950 Machine Learning

Assignment Number 5: Due 4/11/2019 (in class)

1. **[100 points]** Obtain at least **100** text documents belonging to at least two classes and apply the Naive Bayes algorithm or another algorithm for text classification to the documents. You should use ten-fold cross-validation and report the results. To obtain the documents you may use old news group data from Tom Mitchell's website below, some web queries or any method of your choice as long as you clearly describe it. You can use the code provided by Tom Mitchell which can be downloaded from **<http://www.cs.cmu.edu/afs/cs/project/theo-11/www/naive-bayes.html>** or any code or package that supports text classification (WEKA does) or write your own code. If you use or write a code other than Tom Mitchell's, please provide me with the code and/or documentation.