CSCI 6350/7350 Project 3 (Due midnight, Thursday, March 30th)

In this project, we empirically study the feature selection approach, and the classify ensemble approach through a number of experiments performed using the glass data. All the experiments are to be performed using Weka. All classification is to be performed with 10 fold cross validation.

Experiments
Import the **glass.names** and **glass.data** files into Weka explorer (the data comes with the package).

1. Apply the following classification method on the glass data:
    a. K nearest neighbor (classifier→lazy → IBK)
    b. Backpropogation Neural Net (classify→function→Multilayer Perceptron)
    Vary the number of hidden layers (nodes in each layer), learning rate, batch vs. incremental to get a classification accuracy above 95%.
    c. Decision tree (classify→trees→J48)
    d. NaiveBayes (classify→bayes→NaiveBayes)

Record the classification accuracy results.

2. Apply feature ranking methods:
    a. Ranker with information gain
    b. Ranker with correlation
    c. Relief
    d. Wrapper with greedy stepwise

    on the glass data and record the rank order of the features determined by these criteria.
    1. Show the ranking of features from each criteria
    2. Do the rankings agree with each other?
    3. Choose the top 6 features based on their rankings from various feature selection methods. What method did you use to make the selection? Explain.

Modify the original data to only include these 6 features. This is the new data.

3. Repeat the classifications performed in step 1 on this new data. Record all the classification accuracy results.

4. This experimentation step evaluates the Adaboost ensemble classifier performance. Use the original data and the new data for this experiment. Choose the Meta (classifier) – Adaboost method, and each of the four classification methods used in step 1 as the base classifier. Record the classification results

Fill the results obtained from experiments 1 – 4 in the table below. Compare the results across different classification methods, as well as across different data. What conclusions can you draw based on the results?

	Original data	New data (Feature reduced)
Decision tree		
Naïve Bayes		
K Nearest Neighbor		
BackPropogation		
Ensemble (Decision tree)		
Ensemble (Naïve Bayes)		
Ensemble(K Nearest Neighbor)		
Ensemble (BackPropogation)		