Joshua Herman  
Homework 2 Report  
CS 6001

**Comparing Before and After Feature Selection**

Feature selection  
Decision Tree Classifier:   
average accuracy: 0.5506194801690187, average f-measure: 0.5511275833483547

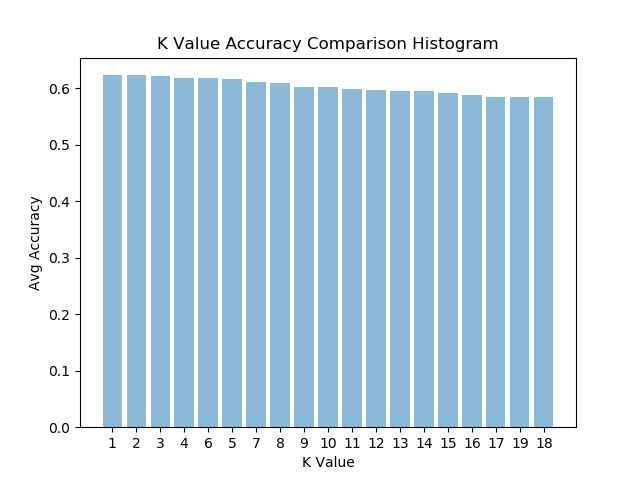
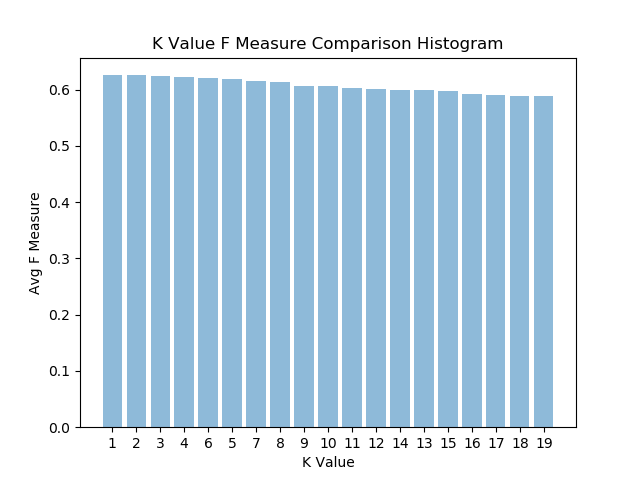
K Nearest Neighbors Classifier:   
average accuracy: 0.6229574769052798, average f-measure: 0.6258295225451558

Without Feature selection:  
Decision Tree Classifier:   
average accuracy: 0.6157083032546342, average f-measure: 0.617023470862013

K Nearest Neighbors Classifier:   
average accuracy: 0.7971364311807014, average f-measure: 0.7966237595222775

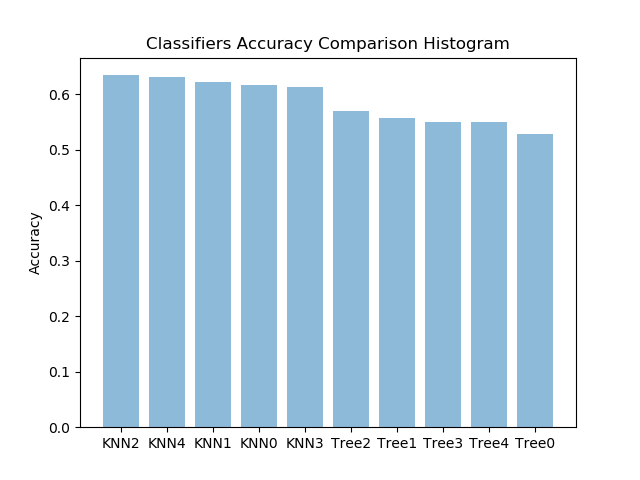
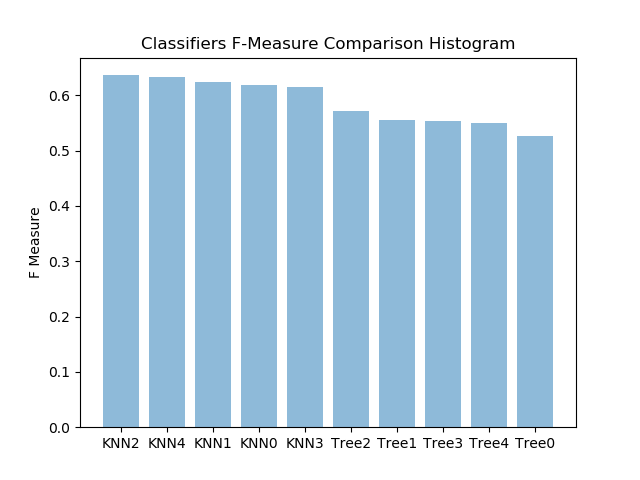
Feature selection makes the average accuracies and f-measures slightly worse since the classifiers have less features to compare when making the models.

**Comparing How K Impacts KNN**



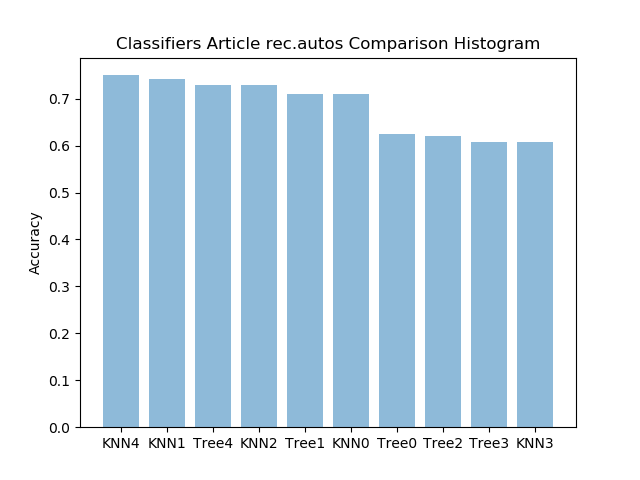
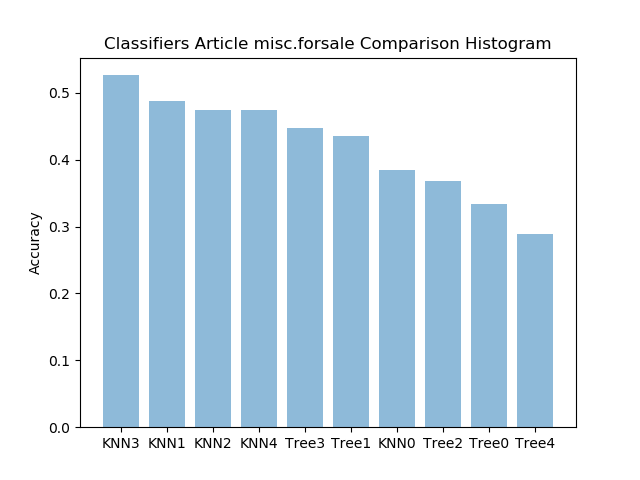
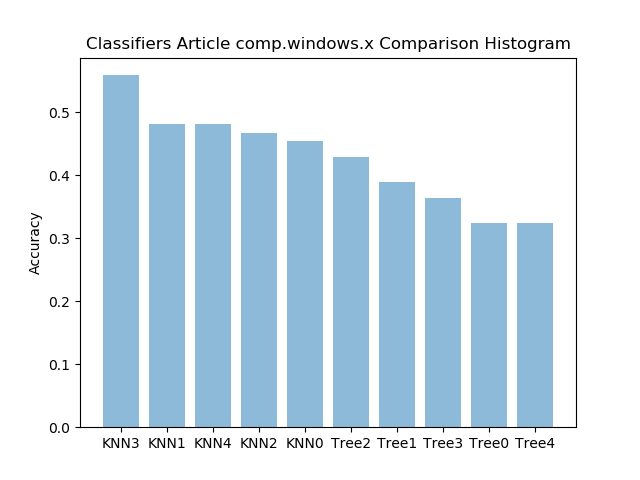
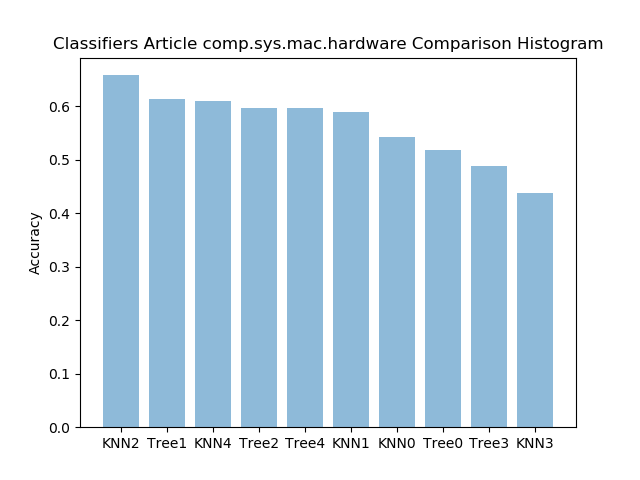
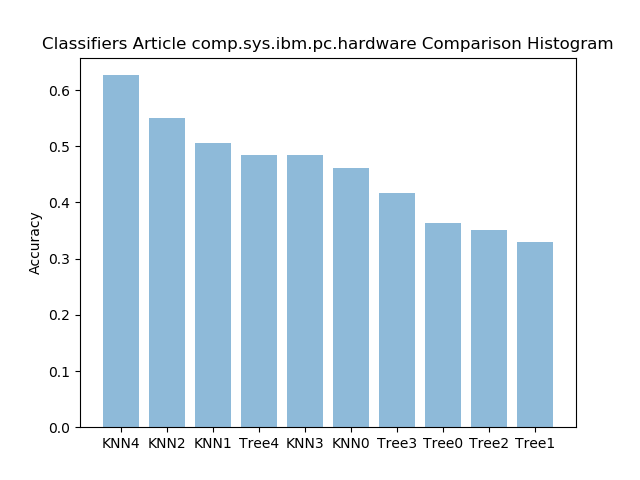
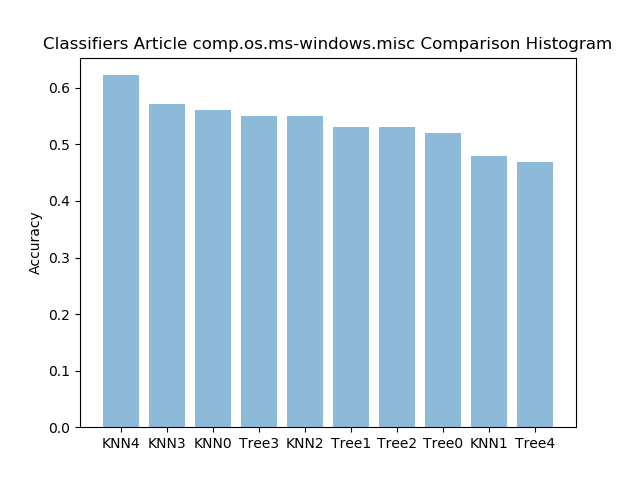
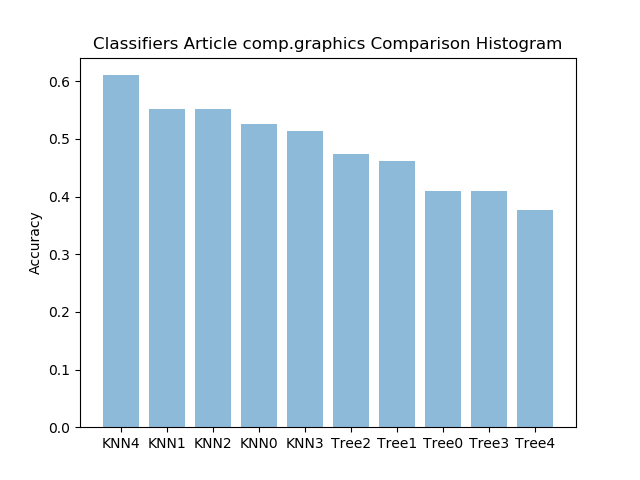
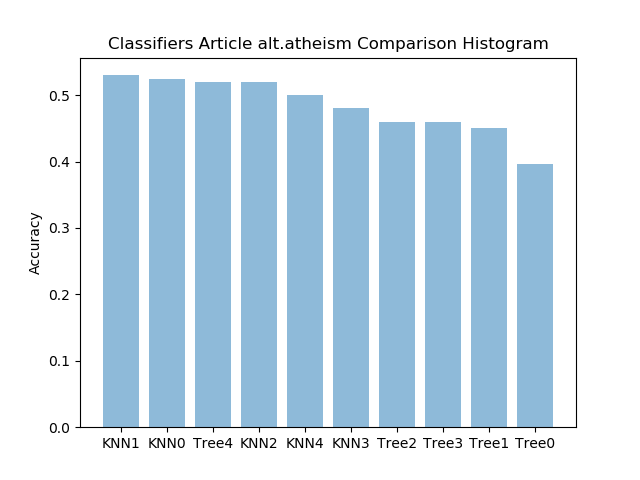
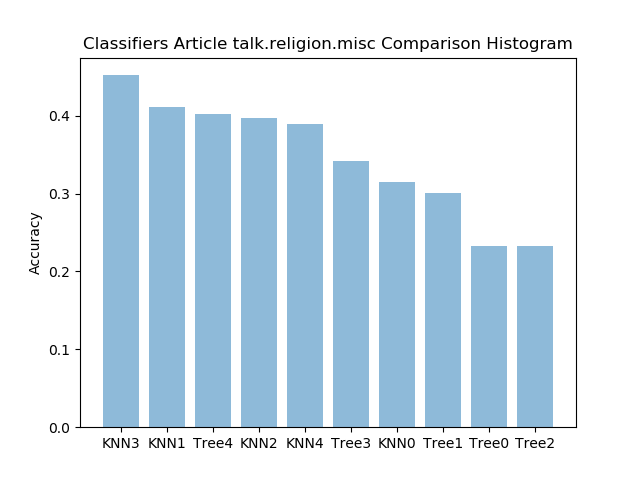
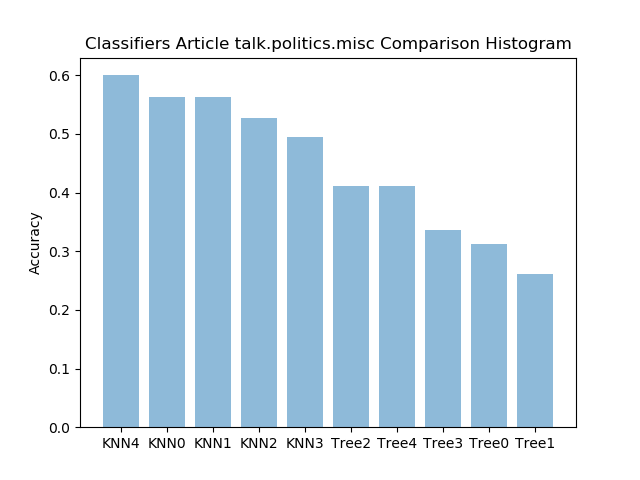
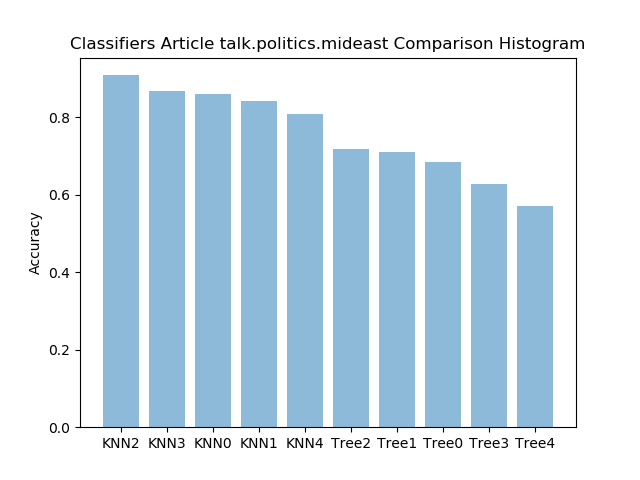
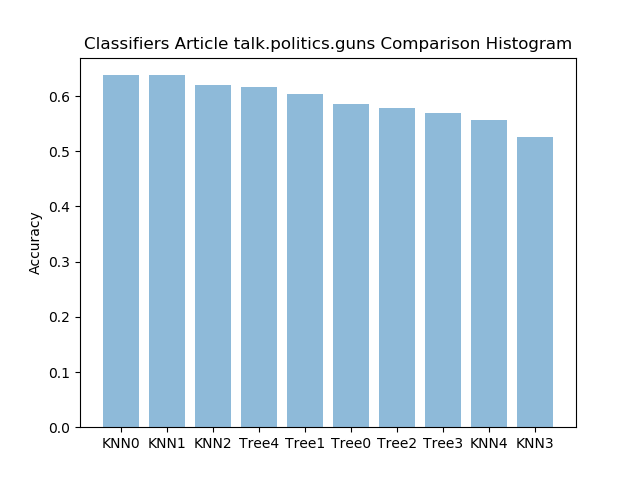
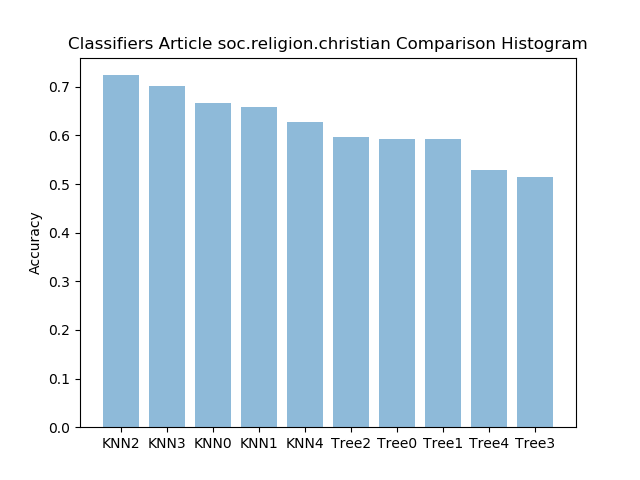
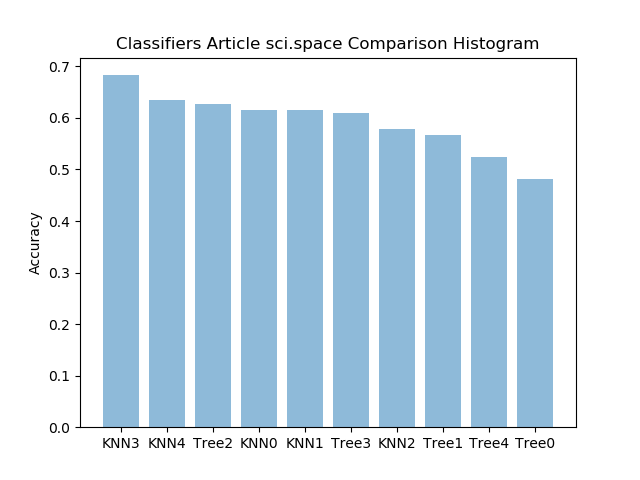
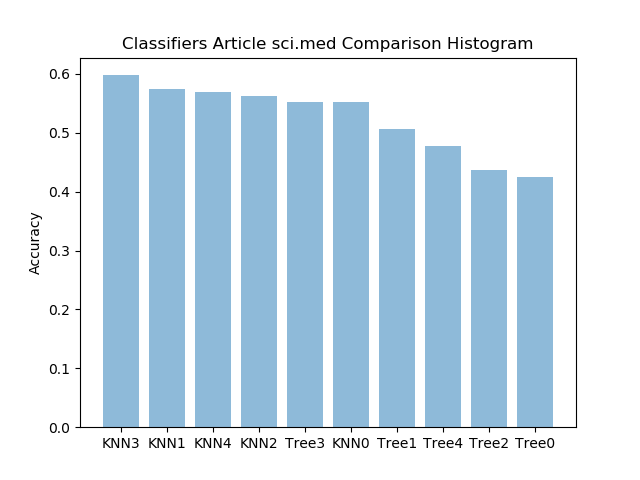
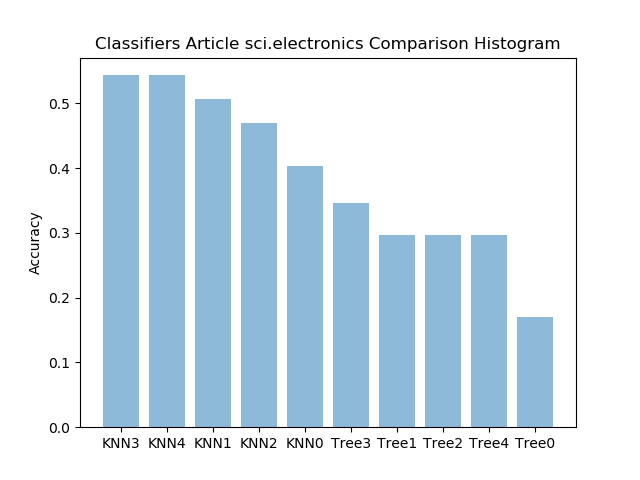
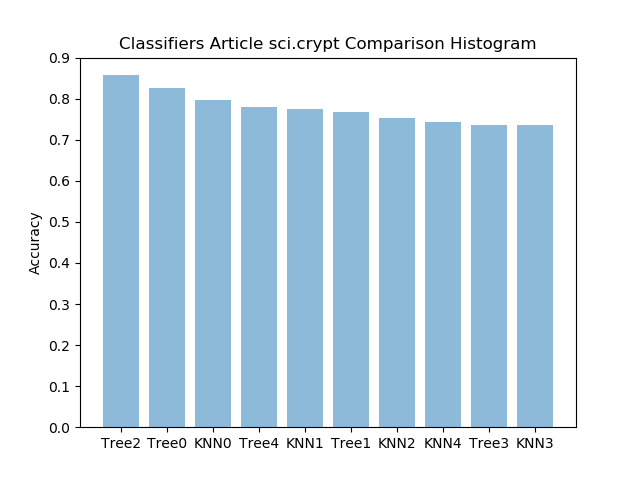
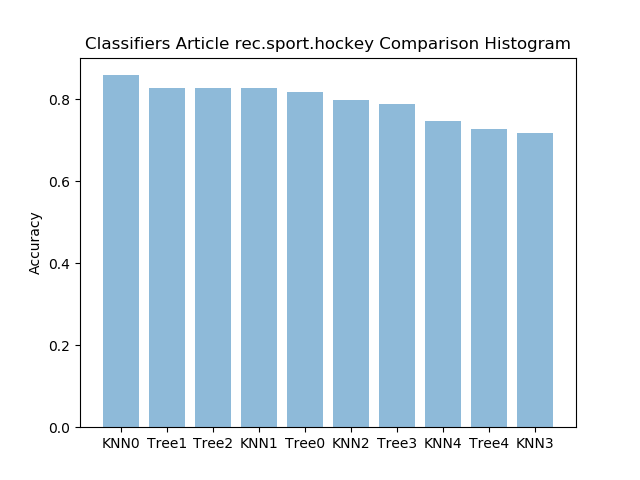
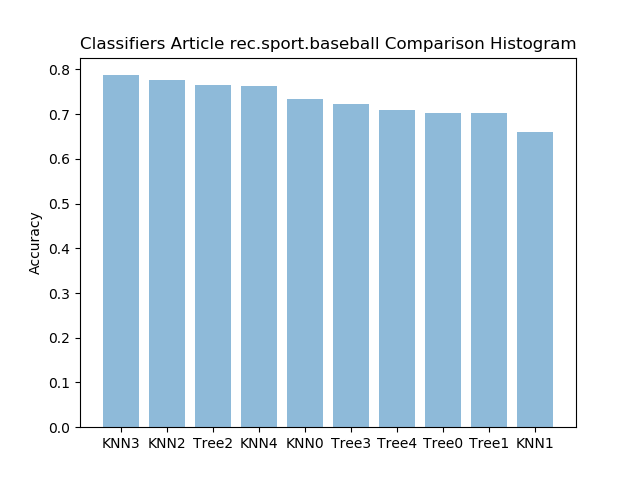
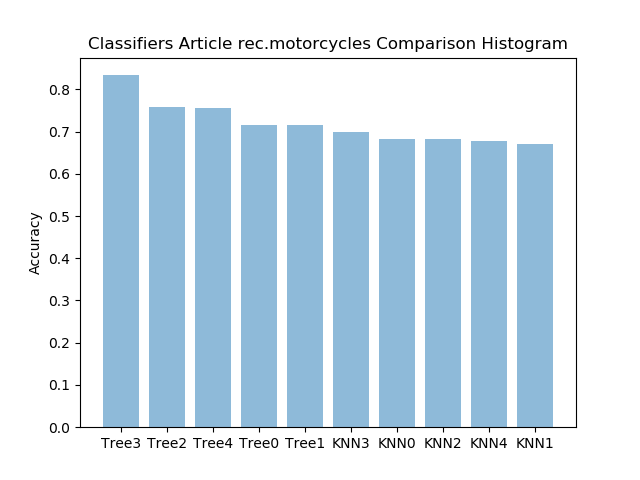
With K Nearest Neighbors Classifier it is able to perform the best with lower k values.

**Comparing KNN with Best K to Decision Tree**



With the best K value K Nearest Neighbors Classifier outperforms the Decision Tree Classifier.

**Comparing Accuracies of kNN and Decision Trees Per Article**Decision tree classifiers get better accuracy for the rec.motorcycles article category but for most of the others K nearest neighbors classifier performs better.



Overall k nearest neighbors classifier performs better than the decision tree classifier for classify the article categories.