My job was to use machine learning algorithms to determine which computer brand our existing customers prefer. I had 10,000 completed surveys informing us of brand preference, with another 5000 surveys that did not include brand preference. Using existing completed survey data to tune the classifiers, my goal was to predict what these customers would have likely chosen given their other attributes (e.g. income, brand of vehicle, etc.).

To that end, I first used all of the K-Nearest Neighbor (KNN) classifiers in WEKA and tuned them on the data. Here were my best results corresponding to each algorithm:

* LinearNNsearch: 24NN = 82% correctly classified
* KDTree: 24NN = 81.24% correctly classified
* Covertree: 24NN = 82.3% correctly classified (Note: The most computationally demanding!)
* Balltree: 22NN = 82.06% correctly classified

So for the KNN classifier, the Covertree algorithm was most accurate. Next, I tuned the Decision Tree classifier on the data with the following result:

* J48 Decision Tree (Minimum number of objects: 24, 4 folds, reduced error pruning): 86.28% correctly classified instances

It's clear that the decision tree was the most accurate classifier for our purposes. It is also less computationally demanding, with is a nice bonus if more data or a larger data set should need to be predicted upon in the future.