Summary:

* How each algorithm behaves under a variety of circumstances
* Compare performance
* Meaning in practice

Algorithm:

1. Decision Tree with some pruning
   1. Use pruning
   2. What is used to split attributes
   3. What is the pruning algorithm
2. Neural networks
3. Boosting
   1. Boosted version of decision trees
   2. Because of boosting, can go more aggressive on pruning
4. Support vector machines
   1. Swap out kernel functions (at least 2)
5. K-nearest neighbors
   1. Different k

Data:

* 2 classification problems
* Explain why they are interesting

Analysis:

* Description of the two classification problems and why they are interesting
  + Non-trivial
  + Different algorithms would produce different results for comparison
* Training and testing error rates for all algorithm on the datasets
  + Performance vs training set size
  + Graphs that show performance on both training and test data (performance vs algorithm)
  + Iterative algorithms – training data performance vs training time, test data performance vs training time.
  + **Training time means all variables: training data size graph, epochs, iterations**
* Analysis of results
  + Why did I get the results
  + Compare and contrast different algorithms
  + What sort of changes might I make to each of the algorithms to improve performance
  + Time taken, iterations
  + Does cross validation help (what is negative about cross validation)
  + How did the data contribute to the performance (biased data or unevenly distributed data)
  + How does the parameters contribute to performance (learning rate, stopping criteria, regularization(make sure everything stays small), pruning methods …)
  + Which algorithm performed best (test error rate + time)

Submission:

* Named jwang775.zip
* Same name folder
  + README.txt – instructions for running everything and recreate the results
  + Code/ no code with weka
  + jwang775-analysis.pdf
  + supporting files (datasets) (if too big, provide instruction)