Erikson Sodergren   Intelligent Systems Project 2

1. Features:
   a. Does a word in the sentence contain 'ijk': more common in dutch
   b. Does a word in the sentence end in 'en': more common in dutch
   c. Does a word in the sentence contain ' 'nt ':more common in english
   d. Does a word in the sentence contain ' 're': more common in english
   e. Does a word in the sentence contain 'aa': more common in dutch
   f. Does a word in the sentence contain 'q': more common in english
   g. Does a word in the sentence contain 'ij': more common in dutch
   h. Does a word in the sentence contain 'ee': more common in dutch
   i. Does the sentence contain the word "the"
   j. Does the sentence contain the word "de": this is 'the' in dutch

2. Decision tree:
   a. I modified the decision tree program ive been using in another class so it slightly differs from the kind we discussed in class: It stops recursing if the data it would be working on is at least 95% from the same class.I ended up deciding on a maximum depth of 3 for the decision tree trainer, however the deepest i saw it actually generate with my testing was 2.

3. Adaboosting:
   a. I decided on 3 decision stubs for adaboosting due to the fact that this problem is generally solved easily with very few attributes/features, and an odd number is ideal to avoid some edge cases where a sentence gets misclassified due to the different weighting of the trees when there are only 2.

4. Misc:
   a. The program is designed to read text files with an encoding of utf8 and will not handle other encodings. However, when testing on my own, i altered the code to handle the default windows encoding instead.