Code Documentation:

To use the program just use the command line as instructed in the writeup. To change the max depth of the decision trees and the number of stumps in the adaboost ensemble simply change the constants found at the top of the program labeled MAX_TREE_DEPTH and NUM_STUMPS respectively.

Features:

To choose the 19 features, I looked up the most common 3-5 letters from each language. I chose 3+ as 1-2 letters can be found coincidentally in various words from both languages. The features are: the, for, his, that, ther, here, with, from, this, able, aa, het, tot, van, dat, niet, zijn, heeft, and voor. I tried to get a proper variety of both English and Dutch words.

Decision Tree Learning:

The decision tree loops through each feature, splitting the example data set off of whether the example has the feature or not, the entropy is then calculated and when the loop is done, the feature index with the best entropy is added to the tree and the function is recursive for each child until the max depth is reached, or all of the features have been tested. For the max depth I kept running the program with different depths until the error percent began to decrease. The depth I ended with was 11 for all 11 features.

Boosting:

For my implementation of boosting, I used the pseudo code from the book. After testing, I found that there was oddly no change in error rate between the number of stumps, so I just chose to do 10.