Lab-3   Write-up   AK4579

(1) Features chosen: I chose 10 features for initialization of features.

The features chosen are of two-types: a, Common words
                                      b, Sub-words that
                                      distinguish b/w English & Dutch.

ⓐ If ij exist together by side.
ⓑ Common words in dutch.
ⓒ If en exists atleast once.
ⓓ If ending words of english exists.
ⓔ If dutch numbers exists,...
     and many more.

(2) The tree keeps building until no info gain is obtained. If no info is gained, leaf node will be generated.

My own testing results have been included along with the code.

(3) The weight is being assigned in accordance with the length of lines, which is inversely proportional.

0.5 is considered as the threshold for the trees, to be considered useful or not.

At each iteration, weight of current tree is being calculated and it is being added to the Present weight of the tree.

A random number is being generated and that no. is used to find index of tree which is then used to calculate the weight of tree.

While the random number is not less than the wt. of tree, index of tree will keep being incremented and wt. will be calculated.

Every decision tree returns a decision, and based on that, weight will get increased (or) decreased.

And based on that, final decision will be cumulative, representing either "en" (or) "nl".