Tejashree Khot and Sanjana Agarwal

ASSIGNMENT 3 - MACHINE LEARNING IN COMPUTATIONAL LINGUISTICS

We have used the majority class as the baseline.
Since some data points had more than one class, we used the last class of every data point.
The following tables depict the parameters used and the accuracy associated with it.

1) ARM.N
Baseline: 0.82

PARAMETERS	ACCURACY
timbl -dID -mJ -k5 -f ~/arm.n.train -t ~/arm.n.test -dID: Inverse Distance -mJ: Jeffrey Divergence -k5: '5' nearest neighbors	0.863636
timbl -dIL -mJ -k13 -f ~/arm.n.train -t ~/arm.n.test -dIL: Inverse Linear -mJ: Jeffrey Divergence -k13: '13' nearest neighbors	0.871212
timbl -dIL -mJ -k11 -f ~/arm.n.train -t ~/arm.n.test -dIL: Inverse Linear -mJ: Jeffrey Divergence -k11: '11' nearest neighbors	0.878788

2) DIFFICULTY.N
Baseline: 0.35

PARAMETERS	ACCURACY
timbl -dID -mJ -k7 -f ~/difficulty.n.train -t ~/difficulty.n.test -dID: Inverse Distance -mJ: Jeffrey Divergence -k7: '7' nearest neighbors	0.388889
timbl -dED -mJ -k7 -f ~/difficulty.n.train -t ~/difficulty.n.test -dED: Exponential Decay -mJ: Jeffrey Divergence -k7: '7' nearest neighbors	0.444444
timbl -dIL -mJ -k1 -f ~/difficulty.n.train -t ~/difficulty.n.test -dIL: Inverse Linear -mJ: Jeffrey Divergence -k1: '1' nearest neighbor	0.500000

3) INTEREST.N
Baseline: 0.42

PARAMETERS	ACCURACY
timbl -dIL -mJ -k11 -f ~/interest.n.train -t ~/interest.n.test -dIL: Inverse Linear -mJ: Jeffrey Divergence -k11: '11' nearest neighbors	0.622222
timbl -mJ -k9 -f ~/interest.n.train -t ~/interest.n.test -mJ: Jeffrey Divergence -k9: '9' nearest neighbors	0.633333
timbl -dID -mJ -k11 -f ~/interest.n.train -t ~/interest.n.test -dID: Inverse Distance -mJ: Jeffrey Divergence -k11: '11' nearest neighbors	0.644444

Thus, as can be seen above, the best parameters for each word is different.