Need to parse raw data

Parsing raw data...

There are non-native 1 files. Total size of samples: floor(11044/1)=11044

non-native files parsed: tofel.txt

There are native 1 files. Total size of samples: floor(11044/1)=11044

native files parsed: reddit.US.txt.tok.clean

all\_data size is 22088

saving to file: ../parsedData/alldata45\_USonly.txt

file ../parsedData/alldata45\_USonly.txt with 22088 lines was created

Finished Parsing raw data

Reading file(311 lines) ../parsedData/functionWords.txt

Reading file(22088 lines) ../parsedData/alldata45\_USonly.txt

Starting feature vector classification

Reading parsed data

There are 11044 Native samples and 11044 Non-Native

Train set size 17670 - native=8811, non-native=8859

Test set size 4418 - native=2233, non-native=2185

Train data 'all feature vector' size is 311x17670

Train data size is 17670

Test data 'all feature vector' size is 311x4418

Test data size is 4418

-----------FUNC\_WORDS\_START---------------------------

Running SVM...

non-native native

precision[0.82814677 0.82251656] - tp/(tp+fp)

recall [0.81601831 0.83430363] - tp/(tp+fn)

fscore [0.82203781 0.82836816] - harmonicAvg(prec + recall)

fscore weighted =0.8252373730803513 - (fscore1\*realPortion+fscore2\*realPortion)

Total accuracy =82.52602987777274%

Running Decision Tree...

non-native native

precision[0.70596591 0.69904597] - tp/(tp+fp)

recall [0.68237986 0.72189879] - tp/(tp+fn)

fscore [0.69397254 0.71028861] - harmonicAvg(prec + recall)

fscore weighted =0.7022192085591099 - (fscore1\*realPortion+fscore2\*realPortion)

Total accuracy =70.23540063377094%

Running NB...

non-native native

precision[0.79220779 0.80961098] - tp/(tp+fp)

recall [0.80961098 0.79220779] - tp/(tp+fn)

fscore [0.80081485 0.80081485] - harmonicAvg(prec + recall)

fscore weighted =0.8008148483476685 - (fscore1\*realPortion+fscore2\*realPortion)

Total accuracy =80.08148483476685%

-------------FUNC\_WORDS\_END-------------------------

Starting top x words classification

top\_words\_list size is 348

Reading parsed data

There are 11044 Native samples and 11044 Non-Native

Train set size 17670 - native=8811, non-native=8859

Test set size 4418 - native=2233, non-native=2185

Train data 'all feature vector' size is 348x17670

Train data size is 17670

Test data 'all feature vector' size is 348x4418

Test data size is 4418

------------------TOP\_WORDS\_START--------------------

Running SVM...

non-native native

precision[0.96723716 0.91276865] - tp/(tp+fp)

recall [0.90526316 0.96999552] - tp/(tp+fn)

fscore [0.93522459 0.94051238] - harmonicAvg(prec + recall)

fscore weighted =0.9378972056991378 - (fscore1\*realPortion+fscore2\*realPortion)

Total accuracy =93.79809868718877%

Running Decision Tree...

non-native native

precision[0.88253968 0.89200181] - tp/(tp+fp)

recall [0.89061785 0.88401254] - tp/(tp+fn)

fscore [0.88656036 0.8879892 ] - harmonicAvg(prec + recall)

fscore weighted =0.8872825460373788 - (fscore1\*realPortion+fscore2\*realPortion)

Total accuracy =88.72793119058397%

Running NB...

non-native native

precision[0.93384685 0.94391678] - tp/(tp+fp)

recall [0.94324943 0.93461711] - tp/(tp+fn)

fscore [0.93852459 0.93924392] - harmonicAvg(prec + recall)

fscore weighted =0.9388881649335703 - (fscore1\*realPortion+fscore2\*realPortion)

Total accuracy =93.88863739248528%

---------------TOP\_WORDS\_END-----------------------

duration(formatted HH:MM:SS): 00:05:04