Parsed data exists

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

Reading file(22084 lines) ../parsedData/alldata.txt

Starting feature vector classification

Reading parsed data

There are 11040 Native samples and 11044 Non-Native

Train set size 17667 - native=8842, non-native=8825

Test set size 4417 - native=2198, non-native=2219

Train data 'all feature vector' size is 311x17667

Train data size is 17667

Test data 'all feature vector' size is 311x4417

Test data size is 4417

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

Running K-FOLD SVM with k=3

split 1:

Running SVM...

non-native native

precision[0.83198211 0.8002642 ] - tp/(tp+fp)

recall [0.79742765 0.83443526] - tp/(tp+fn)

fscore [0.81433849 0.81699258] - harmonicAvg(prec + recall)

Total accuracy=81.56750882912253%

split 2:

Running SVM...

non-native native

precision[0.82160804 0.81000265] - tp/(tp+fp)

recall [0.80387872 0.8272973 ] - tp/(tp+fn)

fscore [0.81264669 0.81855863] - harmonicAvg(prec + recall)

Total accuracy=81.56500475478875%

split 3:

Running SVM...

non-native native

precision[0.82542801 0.81305951] - tp/(tp+fp)

recall [0.80553273 0.83234501] - tp/(tp+fn)

fscore [0.81535902 0.82258924] - harmonicAvg(prec + recall)

Total accuracy=81.90463252275507%

Summary: SVM average accuracy=0.8167904870222212%

Running Decision Tree...

non-native native

precision[0.70096022 0.69237668] - tp/(tp+fp)

recall [0.69085174 0.70245678] - tp/(tp+fn)

fscore [0.69586927 0.69738031] - harmonicAvg(prec + recall)

Total accuracy=69.66266696853067%

Running NB...

non-native native

precision[0.79536335 0.799908 ] - tp/(tp+fp)

recall [0.80396575 0.79117379] - tp/(tp+fn)

fscore [0.79964142 0.79551693] - harmonicAvg(prec + recall)

Total accuracy=79.76001811184061%

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

Starting top x words classification

top\_words\_list size is 343

Reading parsed data

There are 11040 Native samples and 11044 Non-Native

Train set size 17667 - native=8842, non-native=8825

Test set size 4417 - native=2198, non-native=2219

Train data 'all feature vector' size is 343x17667

Train data size is 17667

Test data 'all feature vector' size is 343x4417

Test data size is 4417

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

Running K-FOLD SVM with k=3

split 1:

Running SVM...

non-native native

precision[0.97310728 0.89774169] - tp/(tp+fp)

recall [0.89201501 0.97465565] - tp/(tp+fn)

fscore [0.93079827 0.93461894] - harmonicAvg(prec + recall)

Total accuracy=93.27628361858191%

split 2:

Running SVM...

non-native native

precision[0.97124026 0.89584887] - tp/(tp+fp)

recall [0.8855504 0.97405405] - tp/(tp+fn)

fscore [0.92641806 0.93331607] - harmonicAvg(prec + recall)

Total accuracy=93.00366797989403%

split 3:

Running SVM...

non-native native

precision[0.96347981 0.89832206] - tp/(tp+fp)

recall [0.88879759 0.96684636] - tp/(tp+fn)

fscore [0.92463314 0.93132546] - harmonicAvg(prec + recall)

Total accuracy=92.8134764298329%

Summary: SVM average accuracy=0.9303114267610294%

Running Decision Tree...

non-native native

precision[0.87232143 0.87827285] - tp/(tp+fp)

recall [0.88057684 0.86988171] - tp/(tp+fn)

fscore [0.87642969 0.87405714] - harmonicAvg(prec + recall)

Total accuracy=87.52546977586597%

Running NB...

non-native native

precision[0.93351303 0.93564582] - tp/(tp+fp)

recall [0.93645786 0.93266606] - tp/(tp+fn)

fscore [0.93498313 0.93415357] - harmonicAvg(prec + recall)

Total accuracy=93.45709757754132%

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

duration(formatted HH:MM:SS): 00:08:00