(venv2) D:\workspace\dataMining\src1>python main.py

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

-----------FEATURE\_VECTOR\_START---------------------------

Running K-FOLD SVM with k=3

split 1

Running SVM...

Accuracy=81.56750882912253%

split 2

Running SVM...

Accuracy=81.56500475478875%

split 3

Running SVM...

Accuracy=81.90463252275507%

Running K-FOLD SVM with k=4

split 1

Running SVM...

Accuracy=81.74243796413693%

split 2

Running SVM...

Accuracy=81.54319869588844%

split 3

Running SVM...

Accuracy=81.16283282014129%

split 4

Running SVM...

Accuracy=82.21336714363339%

Running K-FOLD SVM with k=5

split 1

Running SVM...

Accuracy=81.70704097803939%

split 2

Running SVM...

Accuracy=81.63912157573013%

split 3

Running SVM...

Accuracy=82.7258320126783%

split 4

Running SVM...

Accuracy=81.0731265564863%

split 5

Running SVM...

Accuracy=82.17844202898551%

Running K-FOLD SVM with k=6

split 1

Running SVM...

Accuracy=82.31458842705787%

split 2

Running SVM...

Accuracy=81.03776147785928%

split 3

Running SVM...

Accuracy=82.4232545503939%

split 4

Running SVM...

Accuracy=81.17359413202934%

split 5

Running SVM...

Accuracy=82.25543478260869%

split 6

Running SVM...

Accuracy=82.09239130434783%

Running Decision Tree...

Accuracy=69.66266696853067%

Running NB...

Accuracy=79.76001811184061%

-------------FEATURE\_VECTOR\_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...

Accuracy=93.27628361858191%

split 2

Running SVM...

Accuracy=93.00366797989403%

split 3

Running SVM...

Accuracy=92.8134764298329%

Running K-FOLD SVM with k=4

split 1

Running SVM...

Accuracy=93.49755479079876%

split 2

Running SVM...

Accuracy=92.7730483608042%

split 3

Running SVM...

Accuracy=93.24397754030068%

split 4

Running SVM...

Accuracy=93.02662561130231%

Running K-FOLD SVM with k=5

split 1

Running SVM...

Accuracy=93.79669458908762%

split 2

Running SVM...

Accuracy=92.70998415213946%

split 3

Running SVM...

Accuracy=93.45709757754132%

split 4

Running SVM...

Accuracy=92.936382159837%

split 5

Running SVM...

Accuracy=93.38768115942028%

Running K-FOLD SVM with k=6

split 1

Running SVM...

Accuracy=93.58869872317305%

split 2

Running SVM...

Accuracy=93.50719913067101%

split 3

Running SVM...

Accuracy=92.44770442814453%

split 4

Running SVM...

Accuracy=93.80603096984515%

split 5

Running SVM...

Accuracy=92.52717391304348%

split 6

Running SVM...

Accuracy=93.58695652173913%

Running Decision Tree...

Accuracy=87.52546977586597%

Running NB...

Accuracy=93.45709757754132%

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

duration(formatted HH:MM:SS): 00:45:53