5.2 Base on our twitter corpus regarding Trump, we sort all the nouns according to the number of appearances they have. We choose 5 most regular words as our key words: Trump, politics, president, candidate and opinion.

5.4

6.1)

a) S1, S2 on S3:

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

1 1 0.78 1 0.876 0.5 pos

0 0 0 0 0 0.5 neg

Weighted Avg. 0.78 0.78 0.608 0.78 0.684 0.5

B) S2, S3 on S1:

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

1 1 0.85 1 0.919 0.5 pos

0 0 0 0 0 0.5 neg

Weighted Avg. 0.85 0.85 0.723 0.85 0.781 0.5

c) S1, S3 on S2:

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

1 1 0.87 1 0.93 0.5 pos

0 0 0 0 0 0.5 neg

Weighted Avg. 0.87 0.87 0.757 0.87 0.81 0.5

d) Average:

Precision: 0.83

Recall: 1

F-Measure: 0.91

5.5

7.1) Random set:

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

1 1 0.83 1 0.907 0.5 pos

0 0 0 0 0 0.5 neg

Weighted Avg. 0.83 0.83 0.689 0.83 0.753 0.5

7.2) Word vector representation (word2vec – deep learning):

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

1 0.464 0.93 1 0.964 0.768 pos

0.536 0 1 0.536 0.698 0.768 neg

Weighted Avg. 0.935 0.399 0.94 0.935 0.926 0.768

7.3) We can further increase the model performance by including some information from WordNet or Wikipedia as these knowledge bases provide valuable tags about the emotion expression level of words. A possible algorithm to include these tags is to assign weights to words in our corpus, with the weights be determined according to tags.