Naive Bayes

Training Set	Recall	Accuracy	Cross Validation
10	100	87.5	86.1111
20	100	94.25	92.8125
30	100	95	92.8571
40	100	96.5	94.1667
50	100	97	94
60	100	97.75	94.375
70	99.6249	97.5	92.5
80	99.375	98	92.5
90	99.4444	98.5	90
100	99.5	99.5	100

SVM :

Number of words / Training Set	100	200	400	1000	4000
10	68	64	67	56	43.75
20	85.25	88.5	82.25	81.25	69.75
30	89.25	91.75	86.25	87.75	78.25
40	93.25	94.75	92.75	92.25	87.5

50	93.75	96.25	94.75	94.75	92.25
60	95	97	96	96	94.75
70	96.75	97.75	97.5	97.25	96
80	97.75	98	98.5	98.25	97.5
90	98.5	98.5	99	98.75	98.25
100	98.75	99.5	99.25	99.25	99.25

Features \ Training Set	10	20	30	40	50
Chi-square test+tfidf(2)	64	88.5	91.75	94.75	96.25
Chi-square test +tfidf(1)	44.75	69.25	75.5	81.75	87.5
Chi-square test	65.25	82.25	82.25	89.5	91.5
Random+tfidf(2)	39.25	49	50.75	56.5	59.75
Random+tfidf(1)	37.25	47.5	52	57	59.5
Random	33	34.5	37	38.5	41.5
Random 10000+tfidf(2)	43.5	67	78.25	85.5	90.75
Random 10000+tfidf(1)	51	66.5	73.5	86	91
Random 10000	28	28.75	30	33	37

Training Set	60	70	80	90	100

Features					
Chi-square test+tfidf(2)	97	97.75	98	98.5	99.5
Chi-square test +tfidf(1)	93	95.75	96.75	98.25	99.75
Chi-square test	94	95.25	95.75	96.75	97.75
Random+tfidf(2)	62.75	66.5	67.25	69.5	70
Random+tfidf(1)	62.5	66.25	66.75	69.25	70.25
Random	43	44.75	45.5	48	50.25
Random 10000+tfidf(2)	93.75	95.25	96	97.25	98.5
Random 10000+tfidf(1)	92.5	94.75	96.75	97.5	99.25
Random 10000	39.5	44.25	50.5	55.25	59.5

- Naive Bayes

   -bayes.cpp

-SVM(libSVM + feature extraction)

   -v0.cpp chi-square test + tf-idf(2)

   -v1.cpp chi-square test + tf-idf(1)

   -v2.cpp chi-square test + frequency

   -v3.cpp random selection +tf-idf(2)

   -v4.cpp random selection +tf-idf(1)

   -v5.cpp random selectin + frequency