EECS 348 - Kevin Cheng (klc954)

Bayes Positive

Trial #	Precision	Recall	F-Measure
1	0.91214879	0.84432234	0.876926
2	0.91048292	0.84945055	0.87890847
3	0.91038858	0.84102564	0.87433359
4	0.91988613	0.82857143	0.87184429
5	0.91276511	0.83553114	0.87244215
6	0.90598619	0.86483516	0.88493253
7	0.91952191	0.84542125	0.88091603
8	0.92174264	0.83699634	0.8773277
9	0.91326733	0.84468864	0.87764034
10	0.91583566	0.84102564	0.87683788
average	**0.91420253**	**0.84318681**	**0.8772109**

Bayes Negative

Trial #	Precision	Recall	F-Measure
1	0.85509717	0.91868132	0.8857496
2	0.85890834	0.91648352	0.88676236
3	0.85228046	0.91721612	0.88355681
4	0.84405198	0.92783883	0.8839644
5	0.84836204	0.92014652	0.8827974
6	0.87070778	0.91025641	0.89004298
7	0.85694915	0.92600733	0.89014085
8	0.85072123	0.92893773	0.88811066
9	0.85553663	0.91978022	0.88649603
10	0.85303082	0.92271062	0.88650361
average	**0.85456456**	**0.92080586**	**0.88641247**

Bayes Best Positive

Trial #	Precision	Recall	F-Measure
1	0.92991386	0.86996337	0.8989402
2	0.93357343	0.85457876	0.89233123
3	0.93118536	0.85750916	0.8928299
4	0.94771511	0.84322344	0.89242101
5	0.94286889	0.84029304	0.88863064
6	0.94157663	0.86190476	0.89998088
7	0.93230404	0.86263736	0.89611872
8	0.93881686	0.84871795	0.89149673
9	0.94065758	0.85934066	0.89816233
10	0.93253968	0.86080586	0.8952381
average	**0.93711514**	**0.85589744**	**0.89461497**
improvement	**0.02291262**	**0.01271062**	**0.01740407**

Bayes Best Negative

Trial #	Precision	Recall	F-Measure
1	0.87783895	0.93443223	0.90525195
2	0.86592367	0.93919414	0.90107187
3	0.86795655	0.93663004	0.90098661
4	0.85879248	0.95347985	0.90366256
5	0.855963	0.94908425	0.90012159
6	0.87267815	0.94652015	0.90810051
7	0.87218814	0.93736264	0.9036017
8	0.86196524	0.94468865	0.90143307
9	0.8705327	0.94578755	0.90660112
10	0.8707483	0.93772894	0.90299824
average	**0.86745872**	**0.94249084**	**0.90338292**
improvement	**0.01289416**	**0.02168498**	**0.01697045**

Evaluation:

These sample statistics were gotten through 10 trials of 10-fold cross validation (using microaveraging for each trial), with each fold having an equal number of random positive reviews and negative reviews. An equal number of positive and negative reviews were used for each fold because I noticed that an unequal amount creates a bias by making one classification's dictionary much larger than the other's. In fact, using an unequal amount of positive and negative reviews yielded really bad results.

It was surprising to see the Bayes (non-best) Classifier perform so well using only unigrams. Even so, after adding in bigrams to Bayes, the results improved slightly. This is as expected, since bigrams take into account phrases such a "not good" and "not bad." For my bigram implementation, I decided to store only the lower case versions of each bigram and get rid of punctuation marks to reduce the number of distinct bigrams. This helped to improve my results too. To improve on this classifier further, I could add trigrams and implement different forms of smoothing.