

Bayes Testing

	Normal	Bayes		Best	Bayes	
	Recall	Precision	F-measure	Recall	Precision	F-measure
Trial 1	0.888645	0.889984	0.888549	0.896703	0.902879	0.896306
Trial 2	0.885165	0.886577	0.885060	0.882784	0.891705	0.882113
Trial 3	0.885897	0.888129	0.885733	0.892491	0.899876	0.891992
Trial 4	0.882967	0.885447	0.882778	0.893956	0.898547	0.893650
Trial 5	0.883700	0.885791	0.883542	0.890476	0.896776	0.890040
Average	0.8852748	0.8871856	0.8851324	0.891282	0.8979566	0.8908202

The normal naive Bayes testing produced an average recall, precision, and f-measure of around 88%. This system performed rather decently - this was most likely due to adjustments to testing, such as making sure that the number and proportion of positive/negative files tested/trained were equal. The best Bayes testing produced an average of around 89% for its metrics, which improved upon normal naive Bayes by around 1%. This system used bigrams on top of unigrams to classify the reviews. However, this system didn't seem to increase the system's performance dramatically. This could be because important bigrams like 'not X' didn't appear often enough to significantly affect classification. In order to further improve the system, I could try accounting for other factors in addition to bigrams, such as stems, length, and amount of punctuation in positive and negative reviews.

*Metrics used macro averaging to combine results for positive and negative classes.