	Precision	Recall	F-Measure
Naive Bayes			
Good	.843	.915	.878
Bad	.476	.308	.373
Best Bayes			
Good	.827	.581	.584
Bad	.545	.461	.210

Our Naive Bayes Classifier performs well when classifying “Good” reviews, with high precision, recall and f-measure. Initially performance was skewed towards “Good.” We removed the prior probability from our probability function and got less skewed results. However the results are still skewed towards Good (significantly less so). This makes sense because an inherent flaw of this dataset is the overlapping nature that reviews share.

We ended up adding a feature that updated good and bad lists based on misses from validation. The intuition behind this is this was a better way to eliminate words that misled our guesser. When debugging, we saw a lot of words that lead to misses were high frequency words in both lists, which intuitively shouldn’t be on this list. We added this post-training element to adjust for serially misleading words. When a word is repeatedly high scoring and wrong, it will decrement the good/bad list accordingly to reduce the chance of risk in the future.

In our best bayes model, precision and recall decreased for good and increased for bad. This wasn’t a particularly successful predictor but the one notable aspect is that it made performance less skewed towards positive without creating an unstable environment (completely shifting skew). I think if we ran multiple iterations, had exact match on (instead of “in”), and carefully calibrated a new algorithm for correction (right now we multiply the good[key] or bad[key] value by .99 to the nth power (where n is the number of times appeared in misses), we would see even more accurate results.