Bayes Classifier

Rosalie Chan (rmc982) and Karina Sirota (kss118)

File	Total Recall	Precision	FMeasure
Bayes.py	0.999901126032	0.803333913255	0.890901318475
Bayesbest.py	0.999901272458	0.821369667348	0.901875788899

Overall, both files performed decently. Both had precision and f-measures of over 80 percent, and recalls were close to 100. bayesbest.py performed slightly better than bayes.py. Bayesbest.py performed better because we created a dictionary of both unigrams and bigrams when training, and when classifying, the test can look at both unigrams and bigrams for reference, as opposed to bayes.py, which only uses unigrams. This allows for more nuanced language processing since individual words might be more prevalent in positive files, but when put together, the phrases might be prevalent in negative files. To improve future performance, we can take into account other factors like length of the review, punctuation and caps lock.