**HW 6 Results**

| **Enron 2** | **Enron 3** |
| --- | --- |
| **No Improvements:**  Accuracy: 90.71%  Precision: 95.68%  Recall: 66.64% | **No Improvements:**  Accuracy: 95.01%  Precision: 98.88%  Recall: 82.60% |
| **Stop words:**  Accuracy: 85.93%  Precision: 64.92%  Recall: 97.73% | **Stop words:**  Accuracy: 91.36%  Precision: 76.07%  Recall: 99.60% |
| **Lemmatization:**  Accuracy: 87.01%  Precision: 66.82%  Recall: 97.59% | **Lemmatization:**  Accuracy: 91.62%  Precision: 76.70%  Recall: 99.40% |
| **Lemmatization and Stop words:**  Accuracy: 85.15%  Precision: 63.60%  Recall: 97.86% | **Lemmatization and Stop words:**  Accuracy: 90.51%  Precision: 74.24%  Recall: 99.73% |

**Analysis**

All in all, I am pleased with the results of this algorithm. The Naive Bayes algorithm came with high accuracy and precision results. In terms of the improvements added onto the base algorithm, stop words and/or lemmatization highly improved the recall score. Individually, the stop words seemed to have slightly better recall than the lemmatization and no improvements. The lemmatization, though, provided slightly better accuracy and precision. Stop words and lemmatization individually, had a drop in accuracy and precision when compared to the algorithm with no improvements. Lemmatization and stop words together had the best recall, in slight margin from the others and the worst accuracy and precision from their counterparts.

Though I do want to state, I might have come to different conclusions if I did more training, and testing of this model, but given the results here are my conclusions. If catching correct spam was a priority for me, I would use Lemmatization and Stop words. If catching spam and accuracy of spam and correct spam/ham results were important to me, I would use the algorithm with no improvements. Though, if I wanted the best of both worlds, I would use just Lemmatization by itself.