Results and Discussion

1. What is the effect of removing stop words in terms of precision, recall, and accuracy? Show a plot or a table of these results.

   ▪ There are changes in terms of precision, recall, and accuracy. If we remove the stop words, the accuracy and precision values lessen. The value for recall increases if the stop words were not utilized.

   ```
   Accuracy:   0.8718072872533592 Accuracy:   0.812394362821654
   Recall:     0.949879339149805  Recall:     0.954203849283472
   Precision:  0.7346733668341708 Precision:  0.634292742911332
   ```

2. Experiment on the number of words used for training. Filter the dictionary to include only words occurring more than k times (1000 words, then k > 100, and k = 50 times). For example, the word "offer" appears 150 times, that means that it will be included in the dictionary.

   ▪ Given that the occurrence limit increases, the result for this process will have larger accuracy and precision values and lesser recall value.

   ```
   Accuracy:  0.892803099086011
   Recall:    0.910335934237980
   Precision: 0.792556634304207
   ```

3. Discuss the results of the different parameters used for Lambda smoothing. Test it on 5 varying values of the λ (e.g. λ = 2.0, 1.0, 0.5, 0.1, 0.005), Evaluate performance metrics for each.

   ▪ As the value of λ decreases, the better it performs.

   ```
   lambda = 2.0                     lambda = 0.05
   Accuracy:  0.867865141335270     Accuracy:  0.871315295684280
   Recall:    0.946352329682569     Recall:    0.949879339149805
   Precision: 0.729118993135011     Precision: 0.733830489029112
   ```

4. What are your recommendations to further improve the model?

   ▪ Some steps in the process require a long time to process. In order to have easy modeling, since this has a lot of data, features such as matrices should be replaced or skipped. I encountered a performance warning while doing the step. Maybe another approach to prevent this process.