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.
   Evaluation of the effect of stop words on the performance of a spam classification model using the Naive Bayes Classifier.

	With stop words	Without stop words
Accuracy	0.5469	0.9316
Precision	0.3907	0.8610
Recall	0.6362	0.9483

   The table shows that the model with stop word removal had higher precision, recall, and accuracy compared to the model without stop word removal. This suggests that stop word removal had a positive effect on the model's performance.
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.
   By filtering the dictionary to include only words occurring more than a certain number of times (k), the model's performance will change. Specifically, the precision, recall, and accuracy of the model may increase or decrease depending on the value of k and the specific characteristics of the dataset. Smaller dictionary may increase the performance of the model however, it may risk overfitting the data.
   The statement above can be tested by comparing the precision, accuracy, and recall of different values of k.
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.
   It is the same that applies with lambda smoothing - the precision, recall, and accuracy of the model may increase or decrease depending on the value of lambda and the specific characteristics of the dataset.
   Using a smaller value for lambda will result in less smoothing during training, which means that the model will assign lower probabilities to words that do not occur frequently in the training data. This may improve the model's ability to capture subtle patterns in the training data and result in better performance. However, it may also reduce the model's ability to classify unseen emails correctly, especially if those emails contain words that have low occurrences in the training data.
4. What are your recommendations to further improve the model?
   a. Use a different and larger dataset - it can help the model learn more about the characteristics of spam and non-spam emails and improve its generalization ability.
   b. Try other classification algorithm such as support vector machines or decision trees, to see if it has a positive effect on the model's performance.
   c. Try other feature extraction methods to see if it is more effective than the Naïve Bayes classifier.