1. Train Multinomial Naïve Bayes (MNB) classifier to classify the documents in the Amazon corpus into positive and negative classes. Conduct experiments with the following conditions and report classification accuracy in the following table:

Stopwords removed	text features	Accuracy (test set)
yes	unigrams	0.804
yes	bigrams	0.788
yes	unigrams+bigrams	0.823
no	unigrams	0.808
no	bigrams	0.823
no	unigrams+bigrams	0.832

2. Answer the following two questions:

a. Which condition performed better: with or without stopwords? Write a brief paragraph (5-6 sentences) discussing why you think there is a difference in performance.

Without stopword performed better. A stop word is a commonly used word that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query. We would not want these words taking up space in our database, or taking up valuable processing time. We should remove these tokens only if they don't add any new information for your problem. Classification problems normally don't need stop words because it's possible to talk about the general idea of a text even if you remove stop words from it.

b. Which condition performed better: unigrams, bigrams or unigrams+bigrams? Briefly (in 5-6 sentences) discuss why you think there is a difference?

Language model of unigrams+bigrams performed better. We human can understand language easily but machines cannot so we trying to teach them specific pattern of language. As specific word has meaning but when we combine the words (i.e group of words) than it will be more helpful to understand the meaning. Unigrams+bigrams gave more information to help understand the meaning, therefore, it performed better than others.