Accuracy Table:

no stop word: unigram + bigram: 0.8239125 unigram: 0.807125 bigram: 0.7898875 stop word: unigram + bigram: 0.8335875 unigram: 0.8068875 bigram: 0.827125

Answers to the 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.

With stopwords, it performed better.

Stopwords are thought of as "the most common words in a language". However,when context of the text is concerned, treating negation words, which is valence of a text, differently is neccessary somehow. Due to the elimination of stopwords, some emotional words may be removed, which may increase the difficulty of classifying. Besides, removing stop words result in the decrease of the amount of the words, which means accuracy loss in some way. In addition, sometimes eliminating stopwords harm inference. Frequent words echo and reinforce patterns in contentbearing words.

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

unigrams + bigrams performed best.

Bigram has lower bias than unigram mostly, due to the more accurate feature for the text. But it has high variance since there are more kins of feature than unigram. Therefore, unigram suffers high bias but with low variance. Combining them two helps to balance the relative merits and get better results. Also, for specific bigrams and unigrams, those bigrams or unigrms who appear too few or too many can get adjustment from the other one.

Refrence:

https://aclweb.org/anthology/E17-2069

https://stats.stackexchange.com/questions/258134/naive-bayes-mix-unigrams-and-bigrams-for-text-classification