Task5:

1:

I think that preventing these stops will have some impact on my results because I'm using a small training data set. However, as the amount of training increases, that is, the size of the training data set gets larger and larger, preventing the use of these words will have less and less impact on the results. In the real world, as the data set gets larger, the frequency of these stops will be close to each other. Because too much samples will rule out individual impact on affection for certain stop words, such as Tom liked to use 'the', and Jack liked to use "I", when the sample through the hours, such as only one person to the corpus, so will be for some special people like to use the word statistics on the high side, and the sample is large enough time can ignore these problems. Again, these words are not helpful because they have no explicit meaning and are used only for structural pauses or transitions. Therefore, when the data set is small, stop words may appear multiple times in a particular category, and the results may sometimes change after deleting them.

2:

In my opinion, rooting these words will not have a big impact on my results, because no matter what tense or form they are used in, their meaning is the same. It just expresses different tenses, the essential content has not changed, so when they change the tense, their ideology will not change and therefore will not change the result.