**Task 5 - Removing Stop Words and Do Stemming**

**Task 5.1 - Remove Stop Words (5 point)**

Describe if removing the English Stop words (most common words like ”a, the, is, are, i, you, ...”) would change the final kNN results. Does your result change significantly after removing the stop words? Why?

Provide reasons.

You do not need to implement this task.

Yes. If removing the English Stop words, it will change the final results. Because the stop words in English are usually used more than other words and appear in every sentences, the top frequency words are usually stop words, and k-nearest neighbors will show that most of the article are similar due to their stop words. If we remove the stop words, the neighbors of articles will depend on their core words but not the stops words. Thus, removing stop words will not only change the results but also make the prediction more accurate.

**Task 5.2 – Considering English word stemming**

We can stem the words [”game”,”gaming”,”gamed”,”games”] to their root

Word ”game”. Does stemming change your result significantly? Why? Provide reasons.

You can learn more about stemming at:

https://en.wikipedia.org/wiki/Stemming

You do not need to implement this task.

No. It will not change the results a lot. The words stemming is just another version of their root words, so it will not change the similarity of two articles, that is the neighbors of a single articles will not change. Thus it will not change the result significantly.