5.1) A description of the dataset

The dataset used for this program was downloaded from kaggle as a .csv file and read into a python file. With a total of 5000 entries, this dataset is of a reasonable size to get accurate predictions from our model. Since there were no empty rows in the column reviews.text, all 5000 data entries were used in the model. The dataset is considered unlabelled as there is no clear indication if the customer gave a positive or negative review. Although there is a column to suggest if the customer would recommend it, this does not necessarily mean that they liked or disliked it. There is a numerical rating column but this too may be misleading as a customer may leave a neutral review and still give the product five stars. Therefore the data was treated as unlabelled.

5.2) Details of the preprocessing steps

This involved cleaning the data by dropping any missing rows from the column reviews.text as this could have affected the accuracy in our results. Then the text was passed through the nlp model. After passing through this, tokens were created by using .lower(), .strip() and .is_stop() to convert all non-stopwords to lowercase and to strip any whitespace in the words too. The tokens created were then joined together again and passed through the nlp model.

5.3) Evaluation of results

Numerical methods would be preferred for the evaluation of results. However we don’t have the data to compare our results. We shall evaluate the results using the visual inspection method. The visual inspection method allows us to see that the majority of results are predicted correctly. As expected, the positive reviews are returned positive, the negative ones are returned as negative and the model even returns neutral reviews as neutral. However, the last review returned as neutral even though, “would recommend to a friend” is generally considered to be a positive review. This might be misleading. Overall, the model seems to have good accuracy for determining sentiment on product reviews.

5.4) Insight into the model’s strengths and limitations

The greatest strength of this model is its prediction accuracy. On a quick visual inspection, the model seems to do really well with predicting sentiment of the product review. However, there are a few things that could be improved. For instance, the model currently takes a few seconds to run. Although not a huge drawback, it is still possible to reduce the running time to get the same results. By passing through fewer results into the model we could greatly reduce run time.

If we had labelled data with ground truths, it would allow us to accurately examine the model's outputs and the error in our model. Hence, although labelling the data would take some time, it would improve the results the model returns.

The model also struggles with determining context in short sentences which is evident with the last sample review we passed through it. By using a different model type such as Long Short-Term Memory or Transformers, we could improve our results so context can be predicted more accurately.