Implementing BERT Model

1. Open ‘Preprocess.ipynb’ , run all cells. It uses the ‘wikihowAll.csv’ dataset and generates the
   stratified dataset. Then it splits the stratified dataset into passages less than 512 words,
   passages between 512 and 1024 words and passages between 1024 and 2048 words. It saves the data as ‘WikiHow_sample_leq512.csv’,
   ‘WikiHow_sample_in1024.csv' and
   ‘WikiHow_sample_in2048.csv'.
2. Open ‘BERT_testing.ipynb’, and run all cells. It uses ‘WikiHow_sample_leq512.csv’ and generates summaries for the passages using the BERT_large model, then calculates a ROUGE score for each summary (up to the 36,200th summary due to GPU limitations), as well as the average ROUGE scores for all summaries generated. These ROUGE scores are saved in ‘averages_WikiHow_sample_leq512_results_BERT.csv’. To change which dataset is being processed, change the filename from ‘WikiHow_sample_leq512.csv’ to ‘WikiHow_sample_in1024.csv’ or ‘WikiHow_sample_in2048.csv’, and save the ROUGE metric averages to be saved in ‘averages_WikiHow_sample_in2048_results_BERT.csv’ or ‘averages_WikiHow_sample_in2048_results_BERT.csv’. The correlation values are also printed and the correlation plot too. The 3 datasets for ranging text length can be combined and then the combined data’s correlation plotted also.
3. To calculate the BM25 score, open ‘Experiment3.ipynb’, and change all filename references from 'WikiHow_sample_all_withsummary.csv’ to ‘WikiHow_sample_leq512_results_BERT.csv’. Run all cells. This file calculates BM25 scores for each summary and counts the number of summaries gained higher BM25 scores generated by each model.