Data Extraction and NLP Test Assignment

Approach to the problem statement:

1. First, I made a single list of all the stop words from all the text documents.
2. Then I converted every stop word into small case to maintain uniformity throughout the project.
3. Further made a list of all the positive and negative words and converted them into small case. Then updated this list of positive and negative words by removing the words which were also present in the stop words list.
4. Later, I scrapped data from each web page and stored it in a variable.
5. Then using NLP the extracted data was tokenized into small parts and text analysis was carried out.

How to run the file?

1. The file is written in python script.
2. Firstly, python should be installed on your machine.
3. Then using Anaconda or Jupyter Notebook or Visual Studio you can run the file.