**Instructions Documentation – My approach towards solution**

**Steps followed**:

1. Imported required libraries (including “requests” and “BeautifulSoup”) to get extract data from website URLs.
2. Imported input and output files to check desired info and desired columns.
3. After verifying if we can extract data from given URLs, we parsed the data from first URL and then created functions to store title and web page content in the form of a lists.
4. Created function to fetch data from all URLs in similar fashion by using for loop over created functions to extract data from all URLs and stored in a dataframe with 2 columns as “title” and “content”.
5. Imported Natural Language Toolkit library for text analysis of the content.
6. Using file handling techniques, we imported all positive and negative words from documents.
7. From different stop words documents, we imported stop words and combined them to from a single Stop words List to consider in text analysis.
8. We created a function for analyzing the sentiments of the content by defining 4 variables - Positive Score, Negative Score, Polarity Score, Subjectivity Score
9. We created a function to define the Syllable count in the words.
10. We created a function to performed readability analysis by calculating certain parameters such as Average Sentence Length, Percentage of Complex words, Fog Index, Average Number of Words Per Sentence, Complex words, Word Count, Syllable Count Per Word, Personal Pronouns in the text, Average Word Length
11. Utilized the created functions to perform textual analysis using while loop for all the contents rows
12. Defined analysis list containing all the calculated variables and parameters.
13. Created a dictionary list from analysis list in order to create final output data frame.
14. Concatenated the input dataframe (containing **URL ID** and **value**) with dataframe created using analysis list, to get the **final dataframe** with **URL ID**, **URL** & information about all the **calculated variables & parameters**.
15. Extracted the final data frame in the form of Excel as “**Output\_Data\_Structure**”.