# **Objective**

The objective of this assignment is to extract textual data articles from the given URL and perform text analysis to compute variables that are explained below.

# **Data Extraction**

Input.xlsx

For each of the articles, given in the input.xlsx file, extract the article text and save the extracted article in a text file with URL\_ID as its file name.

While extracting text, please make sure your program extracts only the article title and the article text. It should not extract the website header, footer, or anything other than the article text.

# **Data Analysis**

For each of the extracted texts from the article, perform textual analysis and compute variables,

# **Variables**

Definition of each of the variables given in the “Text Analysis.docx” file.

Look for these variables in the analysis document:

1. POSITIVE SCORE
2. NEGATIVE SCORE
3. POLARITY SCORE
4. SUBJECTIVITY SCORE
5. AVG SENTENCE LENGTH
6. PERCENTAGE OF COMPLEX WORDS
7. FOG INDEX
8. AVG NUMBER OF WORDS PER SENTENCE
9. COMPLEX WORD COUNT
10. WORD COUNT
11. SYLLABLE PER WORD
12. PERSONAL PRONOUNS
13. AVG WORD LENGTH