**Methodologies**

**1. Approach to the Solution**

1. **Data Loading:**
   * Load the input data from an Excel file using pandas. The data consists of URLs that need to be processed.
2. **Article Extraction:**
   * For each URL in the DataFrame, send an HTTP request to fetch the webpage content.
   * Parse the HTML content using BeautifulSoup to extract the article's title and body.
   * Save the extracted articles into separate text files for further analysis.
3. **Text Analysis:**
   * **Tokenization & Cleaning:**
     + Use spaCy to tokenize the text and clean it by removing stopwords and punctuation.
   * **Readability Analysis:**
     + Calculate the Gunning Fog Index to assess the readability of the text.
   * **Sentiment Analysis:**
     + Load custom stopwords and sentiment words from specified folders.
     + Count positive and negative words to determine the polarity and subjectivity scores.
   * **Additional Metrics:**
     + Calculate various metrics like average word length, syllables per word, and the count of personal pronouns.
4. **Data Storage:**
   * Store the calculated metrics in an Excel file with columns for each variable.

**2. Running the Python Script**

1. **Setup:**
   * Ensure that all required dependencies are installed. These include:
     + pandas
     + requests
     + beautifulsoup4
     + spacy
     + nltk
     + syllapy
     + openpyxl
   * Download and set up the necessary language models and stopword lists:
     + For nltk, download the stopwords using nltk.download('stopwords').
2. **Execution:**
   * Save the provided Python script as a .py file (e.g., text\_analysis.py).
   * Run the script using the Python interpreter by executing the following command in your terminal or command prompt:

python NLP\_CODE.py

* + Ensure that the input Excel file, stopwords folders, and output directories are correctly set up in the script.

**3. Dependencies Required**

* **Python Libraries:**
  + pandas - For handling data in Excel files.
  + requests - For fetching web content.
  + beautifulsoup4 - For parsing HTML.
  + spacy - For tokenization and text processing.
  + nltk - For natural language processing tasks (e.g., stopwords).
  + syllapy - For counting syllables in words.
  + openpyxl - For Excel file operations (if required).
* **Additional Resources:**
  + Stopwords files for custom stopwords processing (stored in StopWords folder).
  + Positive and negative words files (stored in MasterDictionary folder).