**Blackcoffer**

**Data Extraction and NLP**

**Test Assignment Solution**

**1. dataExtraction.py**

* Install dependencies like **beautifulsoup4** and **textstat**.
* Keep the **Input.xlsx** file in same folder as **dataExtraction.py** file.
* Specify the path for output directory to store extracted text files.
* run the “**dataextraction.py**” file to extract all the data from web sites.

**2. textAnalysis.py**

* Install dependencies like **textstat** and **nltk**.
* Specify the path of output directory for analysis.
* run the “**textAnalysis.py**” file to generate “**sentiment analysis results.xlsx**” file.

**3. output.py**

* Keep the **Input.xlsx** file and **sentiment analysis** **results.xlsx** in same folder.
* run the “**textAnalysis.py**” file to generate one output file.