Oil Summarization (O-SUM) Instructions Manual (V.1)

## Project Summary

Oil and Gas supply chain is a very important sector that is highly impacted by lots of news around the world. Normally, oil companies depend on very thorough analysis related to the demand, supply, seasonality and politics. Sometimes, they support some of the analysis with other news that might not be clearly related with oil and gas market, but are actually impacting the market indirectly. This news comes from variety of sources and with huge volume. They also need to be analyzed quickly in order to ensure oil and gas companies can react fast enough to global events that may impact the market volatility. Summarization of these reports is a task accomplished manually at the moment. This application comes as part of the Natural Language Processing (NLP) course of MBD master program at IE-university to help solve this problem by automating the task of text summarization.

## Instructions for Using the Application

The application comprise of three components as follows:

### NLP-OSUM-TeamC-01: Text Scrapping

The objective of this component is to scrape the oil and gas news articles from Argus Media.

To run the application:

#### 1. Prepare the web scraping

* Along with the project submission, there is a web scraping folder called “NLP-OSUM-TeamC-01”. **Make sure you download this file from GinHub (link below) and run it locally.**
* Install Docker desktop application. You can use the official website to download the it. <https://www.docker.com/products/docker-desktop/>
* Make sure that the virtual browser is configured in the local host and the port 8050 as follows: [http://localhost:8050](http://localhost:8050/)
* To make sure that you are ready to run the web scraping application you can run the following link in any web browser [http://localhost:8050](http://localhost:8050/). You should get an output mentioning the splash version.
* After these steps we are good to proceed with the running the web scraping application.

#### 2. Running web scrapping application

* Open the command prompt application.
* Navigate to the file location **“project”**.
* Type the following and click enter:
  + (For Windows) **.\Scripts\activate**
  + (For Mac or Linux) **.**[**/bin/activate**](https://colab.research.google.com/drive/1BXiyEnHroYAbVq1nx1EjJknVU6gnquAv)
* You should see the (project) virtual environment activated.
* After that you need to run the following command: **scrapy crawl argus\_news**
* When the execution gets completed, you should see an exported csv file that has all the scrapped news in the following file path **project\argus\_news\argus\_news\argusnews.csv**
* This is the direct link to the from Kaggle: [GitHub - ali-alhammad/nlp\_teamc](https://github.com/ali-alhammad/nlp_teamc)

### NLP-OSUM-TeamC-02: Classical Text Summarization Methods

* Simply, run this notebook on Kaggle.
* This is the direct link to the from Kaggle: [NLP-SUM-TeamC-02 | Kaggle](https://www.kaggle.com/code/abdelhadiahmedi/nlp-sum-teamc-02/notebook)

### NLP-OSUM-TeamC-03: Transformer Text Summarization Methods

* Simply, run this notebook on Kaggle.
* Inside the notebook, in the ***Instructions*** section:
  + Set *use\_sample\_text* to *True*to run the transformers on one sample. Otherwise Set *use\_sample\_text* to *False* to run the transformers on the dataset.
  + Set *max\_samples* to the number of sample texts that you want to summarize out of the dataset.
* This is the direct link to the from Kaggle: [NLP-OSUM-TeamC-03 | Kaggle](https://www.kaggle.com/code/abdelhadiahmedi/nlp-osum-teamc-03/notebook)