**Instructions to Run the Article Feature Extraction Project**

**This document provides step-by-step guidance on setting up and running the article feature extraction project. Follow the instructions below carefully.**

**Prerequisites**

1. **Python Installation:**
   * **Ensure Python 3.8 or later is installed on your system. You can download Python from** [**python.org**](https://www.python.org/)**.**
2. **Required Libraries:**
   * **The necessary Python libraries are listed in the requirements.txt file. Install these libraries using the instructions provided below.**
3. **Input Files:**
   * **Ensure the following files are present in the project directory:**
     + **Input.xlsx: Contains the URLs for article extraction.**
     + **MasterDictionary/positive-words.txt: File containing positive words.**
     + **MasterDictionary/negative-words.txt: File containing negative words.**

**Steps to Set Up the Environment**

1. **Clone or Download the Project:**
   * **Place all project files, including the script, input files, and dictionaries, in a single directory.**
2. **Install Dependencies:**
   * **Open a terminal or command prompt.**
   * **Navigate to the project directory using the cd command:**

**cd <path-to-project-directory>**

* + **Install the required libraries using the following command:**

**pip install -r requirements.txt**

1. **Verify Setup:**
   * **Ensure that the article\_data folder exists (if not, it will be created during execution).**
   * **Confirm that the Input.xlsx file contains valid URLs in the URL column.**

**Running the Project**

1. **Run the Python Script:**
   * **Execute the script using the following command:**

**python project\_blackcoffer.py**

1. **Monitor the Logs:**
   * **A processing.log file will be generated in the project directory. Use this file to monitor the script's execution and troubleshoot any errors.**
2. **Generated Outputs:**
   * **Extracted Articles:**
     + **Articles will be saved as .txt files in the article\_data folder.**
   * **Processed Data:**
     + **A file named processed\_articles.csv will contain the paths of the extracted articles.**
   * **Extracted Features:**
     + **Features will be saved in a file named article\_features.csv.**

**Key Notes**

* **Modifying Input Data:**
  + **To analyze different articles, replace the Input.xlsx file with another file containing the required URLs.**
* **Dependencies:**
  + **If additional libraries are required, add them to the requirements.txt file and reinstall dependencies using the pip install -r requirements.txt command.**

**Troubleshooting**

1. **Dependency Errors:**
   * **If a library is missing or outdated, ensure the correct version is specified in requirements.txt and reinstall.**
2. **Invalid URLs:**
   * **Verify the URL column in Input.xlsx. Remove or correct any invalid or inaccessible URLs.**
3. **Permission Issues:**
   * **Run the terminal or command prompt as an administrator to avoid permission errors when creating or accessing files.**
4. **File Not Found Errors:**
   * **Ensure all input files (e.g., dictionaries, Input.xlsx) are placed in the correct directory.**

**# - The final out file name – (article\_with\_analysis,csv)**