# ****Blackcoffer Article Analysis****

This project analyzes articles by extracting:

* **Sentiment scores** (positive, negative, polarity, subjectivity)
* **Readability metrics** (Fog Index, complex words, word counts, syllables, personal pronouns)

The results are saved in a excel output file.

## ****Project Structure****

Project/

│

├── Data\_Extraction.py

├── Sentiment\_analysis.py

├── Analysis\_Functions.py

├── Writing\_Scores.py

├── config.py # Contains PATHS & output fieldnames

├── StopWords # Folder with given stopword files

├── MasterDictionary # Folder with positive & negative words

├── Input.xlsx # File with URL\_ID and URL columns

├── Articles # Folder where scraped articles (.txt) are saved

└── Output.xlsx # Final output file (.xlsx)

## ⚙️ ****Setup****

### 1️⃣ Install dependencies

Run these in terminal or Colab:

*pip install requests beautifulsoup4 nltk spacy pandas openpyxl*

*python -m spacy download en\_core\_web\_sm*

Download NLTK stopwords:

*import nltk*

*nltk.download('stopwords')*

## ****How to Run****

### ****Step 1 – Modifying Config****

### **Specify Path according to your system in PATHS**

### ****Step 2 -- HowExtract Articles and Generate Sentiment & Readability Scores****

Compute metrics and export final results:

*python Main.py*

This:

* Uses Input.xlsx and saves files in Articles.
* Loops through all .txt files in Articles
* Analyzes each for sentiment + readability
* Writes a .xlxs file to the output path specified in config.py.

## ****How It Works****

| Step | Details |
| --- | --- |
| **Sentiment Analysis** | Uses custom StopWords + MasterDictionary (positive & negative lists). |
| **Readability Analysis** | Uses **NLTK stopwords** to clean text before counting words and computing metrics. |
| **Scraping** | Uses requests + BeautifulSoup to download articles from URLs. |