Sentiment Analysis Project Documentation

Introduction:

This document serves as a comprehensive guide for understanding and running the sentiment analysis script developed as part of a project. The primary goal of the sentiment analysis is to analyze the sentiment of text data, providing insights into its positive or negative nature. The script calculates various metrics, including positive score, negative score, polarity score, subjectivity score, and readability metrics, to provide a holistic view of the sentiment of the text data.

1. Approach to the Solution:

Data Scraping:

In the initial phase of the project, we gathered text data relevant to our analysis. This involved employing web scraping techniques to extract data from various online sources. The collected raw text data formed the foundation for our sentiment analysis.

Data Cleaning:

Following data acquisition, we proceeded to clean the raw text data to prepare it for sentiment analysis. The cleaning process involved several steps, with the primary focus on removing irrelevant information and noise from the text. This included removing stop words, which are common words that do not carry significant meaning in the analysis context. Additionally, we formatted dictionaries containing positive and negative words, which are essential for sentiment analysis.

Sentiment Analysis:

The core aspect of the project involved performing sentiment analysis on the cleaned text data. This process entailed calculating various sentiment metrics to gauge the overall sentiment of the text. Key metrics included:

- Positive Score: Total count of positive words in the text.

- Negative Score: Total count of negative words in the text.

- Polarity Score: A score indicating the overall polarity (positive or negative) of the text.

- Subjectivity Score: A score indicating the subjectivity of the text.

- Readability Metrics: Metrics such as average sentence length, percentage of complex words, Fog Index, average number of words per sentence, complex word count, syllable per word, personal pronouns count, and average word length were calculated to provide additional insights into the readability and complexity of the text.

2. How to Run the Script:

To execute the sentiment analysis script and generate the output, follow these detailed steps:

Step 1: Install Dependencies:

Ensure that the following dependencies are installed on your system:

- Python (version X.X.X or later)

- NLTK library (install using `pip install nltk`)

- CMU Pronouncing Dictionary (download using `nltk.download('cmudict')`)

Step 2: Download the Script:

Download or clone the sentiment analysis script (“Data\_Extraction and Sentiment\_Analysis.py`) from the project repository.

Step 3: Navigate to Directory:

Open a terminal or command prompt and navigate to the directory containing the script.

Step 4: Run the Script:

Execute the following command to run the script-“ python sentiment\_analysis.py”

Step 5: Review Output:

Upon completion, the script will generate an output CSV file (“output\_file”.csv`) containing the calculated metrics for each text file analyzed.

3. Dependencies:

Ensure the following dependencies are installed:

- Python (version X.X.X or later)

- NLTK library

- CMU Pronouncing Dictionary