Blackcoffer Data Extraction and NLP

Test Assignment

Mayank Gajwe | mayank@haxnovr.io | 836-935-3251

Overall Approach:

1. **Using TextBlob:** I opted to use the TextBlob library for sentiment analysis. TextBlob offers pre-trained models that can analyze the sentiment of text more accurately and handle various contexts compared to a dictionary-based approach.
2. **Customizable Word Lists (Optional):** The solution also included the option to incorporate custom positive, negative, and stop word lists. This allows for domain-specific sentiment analysis beyond the general sentiment provided by TextBlob.

Breakdown of Functions:

- analyze_text **function:** This function takes the text as input and performs the following steps:
    - **Preprocessing:** The text is converted to lowercase and stop words are removed.
    - **Sentiment Analysis using TextBlob:** A TextBlob object is created, and its sentiment analysis features are used to obtain sentiment polarity and subjectivity scores.
    - **Word Count** the function calculates the occurrences of positive and negative words from the custom lists within the cleaned text.
    - **Returns a dictionary:** The function returns a dictionary containing the calculated sentiment scores (polarity and subjectivity) and custom word counts.

Addressing Errors:

The explanation also covered potential errors that might occur during the process, such as:

- Empty text after preprocessing (handled by checking for empty text before division in subjectivity score calculation)
- Missing or inaccessible stop word files (addressed by checking for directory existence and potential file access issues)

By following this approach, I aimed to provide a more robust and customizable solution for sentiment analysis using TextBlob while offering the flexibility to include domain-specific sentiment analysis through custom word lists.

Thank you,

Mayank Gajwe

mayank@haxnovr.io