Project Title: **Social Media Sentiment Analysis**

Team ID: **Team_3527**

Team Leader Name: Nilkanth Suthar

Institute Name: BITS Edu Campus (Babaria Institute)

Abstract:

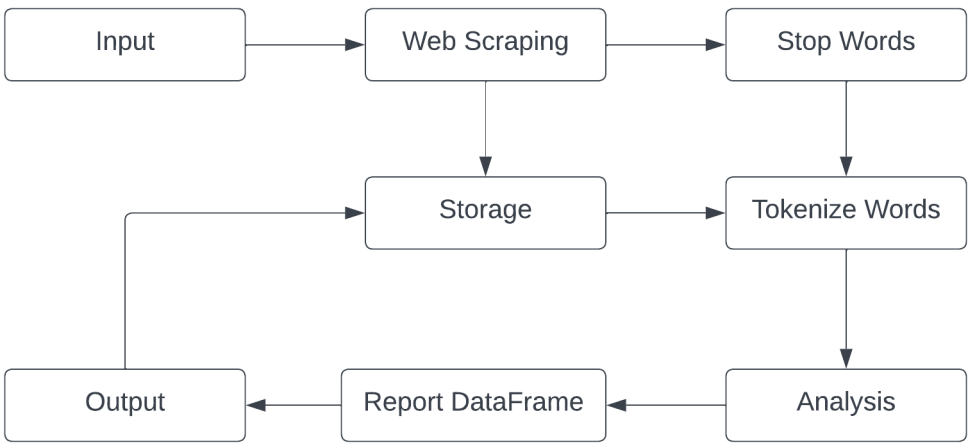
The Sentiment analysis is a computational technique used to identify, extract, and quantify subjective information from text. In this project, we performed sentiment analysis on a set of news articles scraped from various websites using Python. We used various text analysis techniques to extract features such as polarity score, subjectivity score, complex word count, and personal pronoun count. The results were then analyzed to gain insights into the sentiment of the news articles.

Introduction:

Sentiment analysis has become an increasingly popular technique in the field of natural language processing. It involves the use of various computational methods to identify, extract, and quantify subjective information from text. Sentiment analysis can be used in a variety of applications, such as analyzing customer feedback, monitoring social media, and identifying the sentiment of news articles.

In this project, we performed sentiment analysis on a set of news articles scraped from various websites. We used Python to extract various features from the text, such as polarity score, subjectivity score, complex word count, and personal pronoun count. We then analyzed the results to gain insights into the sentiment of the news articles.

Block Diagram:



Technologies:

The project was implemented using Python 3. We used the Beautiful Soup library to scrape news articles from various websites, and the NLTK library for text analysis. We also used the Pandas library for data manipulation and analysis.

Dataset:

The input.xlsx dataset contains a list of URLs to web pages that will be used as input for sentiment analysis. The dataset has two columns: "URL_ID" and "URL". The "URL_ID" column contains a unique identifier for each URL, while the "URL" column contains the actual URL for each web page.

The purpose of the input.xlsx dataset is to provide a standardized list of URLs to be used as input for the sentiment analysis process. By centralizing this list into a single dataset, it allows for easier management and manipulation of the input data.

Here is an example of what the input.xlsx dataset might look like:

URL_ID	URL
1	https://www.example.com/page1.html
2	https://www.example.com/page2.html
3	https://www.example.com/page3.html
4	https://www.example.com/page4.html
5	https://www.example.com/page5.html

Results:

The sentiment analysis was performed on a set of news articles scraped from various websites. We analyzed the results to gain insights into the sentiment of the articles. The results showed that the majority of the articles were neutral in sentiment, with a few articles leaning towards a positive or negative sentiment. We generate a set of variables for each article, including positive score, negative score, polarity score, subjectivity score, average sentence length, percentage of complex words, FOG index, average number of words per sentence, complex word count, word count, syllables per word, personal pronoun count, and average word length. We output these variables in a structured Excel file for further analysis.

We also analyzed the various features extracted from the text, such as polarity score, subjectivity score, complex word count, and personal pronoun count. These features provided additional insights into the sentiment of the articles.

Conclusion and Further Scope:

In conclusion, this project demonstrated the use of sentiment analysis to analyze the sentiment of news articles. The results showed that the majority of the articles were neutral in sentiment, with a few articles leaning towards a positive or negative sentiment. The various features extracted from the text provided additional insights into the sentiment of the articles. The results of our analysis can be useful for businesses, governments, and individuals in understanding the sentiment expressed in digital content. Further analysis can be performed on the output file generated by our program to gain more insights.

Further scope of this project includes expanding the scope of the analysis to include more websites and news articles. Additionally, the analysis can be extended to include more advanced text analysis techniques, such as named entity recognition and sentiment classification using machine learning algorithms.

References:

1. Beautiful Soup Documentation. Retrieved from https://www.crummy.com/software/BeautifulSoup/bs4/doc/
2. Natural Language Toolkit Documentation. Retrieved from https://www.nltk.org/
3. Pandas Documentation. Retrieved from https://pandas.pydata.org/docs/
4. MasterDictionary Dataset. Retrieved from https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon
5. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1–135. https://doi.org/10.1561/1500000011
6. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016