State Wise Stock Market Sentiment Analysis Using Web Scraping

Aim of the Project :

The project's main aim is state-wise stock market sentiment analysis for targeting the audience with correct social media keywords planning and advertisement. This project uses web scraping tools to collect text-based data from platforms like Facebook, Twitter, Pinterest, and Instagram. The data is then processed using a natural language processing algorithm for specific keywords, as well as WordCloud for graphical representations of word frequency, which emphasize words that appear more frequently. The end output is a list of keywords that can be used for state-wise sentiment analysis of the stock market for social media marketing and advertisements.

Procedure:

1) Data Collection

- **Data Extraction using web scraping**:

  Requirements: chrome driver (link: https://chromedriver.chromium.org), install and import libraries like selenium, requests, by tesseract (OCR tool used for reading texts from images), pillow.

- **Running web scraping code:**

  - Add your login id, password and keywords and run.
  - Eg for keywords: learn stock market Bangalore, learn stock market Delhi etc.
  - The web scraping code returns a text-based output as keywords.txt (eg: learn stock market Delhi.txt)
  - Run the same codes for multiple keywords
  - Extracted data is stored in text files.

2) Data Cleaning

- **Merging the text files:**

  - Used merging tools to combine all the documents.

    - Merging tool: https://products.aspose.app/words/merger/txt
    - Merged text files are stored in merged text files.

- Importing the merged .txt file
- Run the text file using keyBERT to get keywords and word cloud to get a visual representation of the frequency of the keyword.
- Save the word cloud model to get keywords png image, word cloud images of both the states are saved in word cloud output images.

Working:

1. Run, the code logs in to the profile and collects data based on given search keywords, it collects data from posts.
2. The web scraping tools run on chrome driver and it is controlled by automation text software selenium.
3. The output text file is combined using the merging tool.
4. The merged text file is processed using the NLP algorithm KeyBERT.
5. KeyBERT provides keywords for sentiment analysis.
6. Wordcloud gives an image output consisting of multiple keywords based on word frequency.

Expected Results:

The result is a list of keywords that can be used for state-wise sentiment analysis of the stock market. By analyzing the keywords output, a user can better plan their social media advertising by understanding what types of keywords they need to use, as well as the way people from different states view the stock market, and by using WordCloud, the keywords are visualized based on word frequency extracted from the input text file. The word cloud (image) is used to understand the overall sentiment of an individual state towards the stock market in a single image whereas KeyBERT helps in determining the specific keywords that need to be considered before selecting any keywords and designing advertisement creatives.

WordCloud: A word cloud is a collection, or cluster, of words depicted in different sizes. The bigger and bolder the word appears, the more often it's mentioned within a given text and the more important it is.

Framework required:

- Python 3.6 and above
- Matplotlib for visualization
- keyBERT for keywords extraction
- WordCloud for graphical representation of keywords with high frequency