Product Requirements Document (PRD)

Product Name: Stock Sentiment Analyzer
Owner: Chinmay Dhamapurkar

Objective/Background:

To provide a sentiment analysis of news headlines related to specific stocks from the S&P 500. This assists investors in understanding potential market movements influenced by media perceptions.

Features and Functionality:

Ticker Listing:

- Fetches and displays a list of S&P 500 tickers from Wikipedia.
- Provides an interactive option to select a specific ticker or analyze all.

News Fetching:

- For a selected ticker, fetches the latest news headlines using the yfinance library.
- Handles potential data fetching errors, ensuring the user is informed if news cannot be fetched.

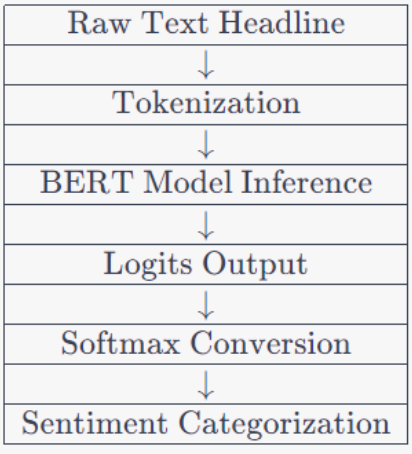
Sentiment Analysis:

- Uses the BERT model from the transformers library to perform sentiment analysis on fetched news headlines.
- Classifies sentiment into five categories: very negative, negative, neutral, positive, and very positive.

Output Display:

- Displays the analysis results in a structured table format using prettytable. The table includes:
  - Ticker Symbol
  - News Title
  - Sentiment
  - Stock Name
  - Link to the original news article

Model Architecture Visualization:



Technical Deep Dive:

Transformers and BERT Model:

- Transformers are a type of deep learning model, particularly designed for handling sequences.
- BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based machine learning technique for NLP pre-training.
- It understands the context from both left and right (bidirectional) of a word in a sentence, rather than just one side.
- Our implementation uses a pre-trained BERT model for sentiment analysis.

Tokenization:

- The BERT model requires input in a specific format, so raw text headlines are tokenized using the provided tokenizer.
- Tokenization breaks down text into smaller pieces, called tokens.
- Special tokens, like [CLS] and [SEP], are added to fit BERT's requirements.

Model Inference:

- Tokenized inputs are fed into the BERT model.
- BERT processes the sequence and outputs logits for each sentiment category.
- Softmax is applied to the logits to derive probability distributions across the sentiment categories.

Softmax Explained:

- Softmax is a mathematical function that converts a vector of numbers into a probability distribution.
- For sentiment classification, it gives the probability for each sentiment category, ensuring they sum up to 1.

Sentiment Categorization:

- The sentiment category with the highest probability post-Softmax application is chosen as the final sentiment for the headline.

User Flow:

- User starts the program.
- The program fetches and displays the first 50 tickers from the S&P 500 list.
- User is prompted to select a specific ticker or choose 'ALL' to analyze all tickers.
- The program fetches news headlines for the selected ticker(s).
- Each news headline is analyzed for sentiment.
- The program displays a table with the ticker, news title, sentiment, stock name, and the link to the original news.

Dependencies:

- Libraries:
  - yfinance - For fetching financial news data.
  - requests - For making HTTP requests.

    - BeautifulSoup - For web scraping Wikipedia to fetch the list of S&P 500 tickers.
    - transformers - For using the BERT NLP model.
    - prettytable - For displaying results in a table format.
    - torch - For model input processing and predictions.
- External Data:
    - Wikipedia's list of S&P 500 companies
    - News headlines for tickers (from yfinance)

Potential Future Enhancements:

Historical Sentiment Analysis: Analyze sentiment over time to identify trends.

Graphical Display: Plot sentiment scores over time to visualize sentiment trends.

User Profiles: Allow users to save specific tickers and preferences for future sessions.

Expand Data Sources: Integrate more financial news sources for comprehensive analysis.

Notification System: Notify users when there's a significant sentiment change for a tracked stock.

Constraints and Assumptions:

- The product currently supports only S&P 500 tickers.
- Sentiment analysis accuracy depends on the BERT model and might not always reflect the actual sentiment.
- There is a potential delay between real-time events and their reflection in the news data from yfinance.