Stock and Cryptocurrency News Summarization and Sentiment Analysis

Ganesh Nalla(E21CSEU0303)-EB17

Shivangi Goyal(E21CSEU0587)-EB12

INTRODUCTION

The primary goal of this project is to create a tool that automatically summarizes financial news articles for specific stock and Crypto tickers using deep learning techniques. By automating this process, customers should be able to save time and effort when getting critical information on the equities in which they are interested.

This Python program creates a stock news summarizer using deep learning. The tool accepts a stock or cryptocurrency url symbol as input, then utilizes its algorithms to crawl relevant news pieces and provide concise summaries that include all of the crucial facts.

This application is intended for busy traders and investors who must keep up with a constant stream of financial news. Despite their lack of time to read every news piece, they must stay current on company changes and market trends.

PROBLEMS ADDRESSED

This Stock News Summarizer addresses many issues that investors and traders confront as they navigate the ever-increasing flood of financial news.

1. **Information Overload:**
   a) Financial news sites produce a large number of items every day.
   b) Keeping track of all the required information for various stocks or cryptocurrencies might be difficult.
2. **Time Constraints:**
   a) Investors and traders frequently make time-sensitive judgments.
   b) Scrolling through lengthy articles may not be possible in such cases.
3. **Extracting Key Points:**
   a) Even if investors manage to read the news, collecting key conclusions from each item might be difficult.
   b) They may find it difficult to distinguish between important and unnecessary information.

SOLUTION

This Python code uses Streamlit to develop a simple web application for summarizing and assessing the sentiment of stock and cryptocurrency news stories. Here's an overview of the code and its features:

1)Imports and Initialization:

- The code includes important libraries such as streamlit for developing the web app, transformers for interacting with the pre-trained Pegasus model, BeautifulSoup for web scraping, requests for retrieving websites, and pipeline for sentiment analysis.
- It defines the model name (human-centered-summarization/financial-summarization-pegasus), which refers to a pre-trained Pegasus model that is especially built to summarize financial news.
- The tokenizer (tokenizer) and model (model) are then loaded using the model name that was supplied.
- Finally, a sentiment analysis pipeline (sentiment) is built with the transformers library.

2)Function for Scraping and Processing Articles (scrape_and_process)

- This function accepts the URL of a news article as input.
- It uses the requests library to obtain the webpage's content.
- BeautifulSoup is then used to parse the webpage's HTML content.
  The function uses soup.find_all('p') to find all paragraph elements (
  on the webpage.
- It collects the text content of each paragraph and merges it into a single string (article).

3)Function for Summarizing Articles (summarize)

- This function, which is marked with @st.cache_data(), is used to cache summaries to increase efficiency.
- It accepts the processed article text (article) as input.
- The function initially uses the tokenizer to encode the article text into a format acceptable for the Pegasus model.
- The max_length option restricts the length of the input text to 512 tokens, preventing the model from being overloaded.
- The truncation=True parameter guarantees that any text that exceeds the limit is immediately truncated.
- The encoded article is then given into the Pegasus model (model.generate), which produces a summary.
- The function sets up the model to create summaries with a maximum length of 300 tokens (max_length=300) and uses beam search with 5 beams (num_beams=5) to increase summarization quality.
- The early_stopping=True option terminates the summarizing process once a sufficient summary has been created, saving computing time.

- Finally, the resulting summary is decoded into human-readable language by the tokenizer (tokenizer.decode) and returned.

4) Streamlit App Design:

- The code uses Streamlit to produce a visually stunning and user-friendly online application.
- It uses st.markdown to build custom CSS styles for formatting the app's layout, such as titles, input sections, buttons, and summary/sentiment containers.

5) User Interaction and Information Display:

- Thewebapp uses st.markdown to show the headline "Stock and Cryptocurrency News".
- It has an input region (st.text_input) where users may enter the URL of the news they wis h to summarize.
- Using st.button, we construct a button titled "Get Summary and Sentiment Analysis".
- When the user selects the button, the message "Please wait while we fetch and analyze th e news..." is presented on st.info to indicate that processing is occurring.
- To mimic processing time, the code implements a 3-second delay using time.sleep(3).

6) **Summary and Sentiment Analysis (upon button click):**

- The scrape_and_process method is used to obtain the article text from the given URL.
- The summarize function generates a brief summary of the article.
- Sentiment analysis is conducted on the resulting summary using the sentiment pipeline.

7) **Displaying Output:**

- The resulting summary and sentiment analysis findings are shown in distinct parts using s t.markdown.
- The summary is presented with a bolded title "Summary:" followed by the actual summar y content, all formatted within a visually different container created using custom CSS.
- Similarly, the sentiment analysis result is presented with the title "Sentiment Analysis:" fo llowed by the sentiment classification (likely positive, likely negative, or neutral) and any further information offered by the sentiment analysis model.

Overall, this code shows a well-structured and user-friendly Streamlit application that uses deep learning to summarize and assess sentiment in stock and cryptocurrency news stories.

LIMITATIONS

1) **Dependency on Web Scraping:**
   a) The program uses web scraping to retrieve article material.
   Websites' structures change regularly, potentially destroying scraping functionality.
   b) Websites may also use anti-scraping methods to keep bots from flooding their servers.
   This may necessitate changing the scraping algorithm to cope with the revised website structure.
2) **Limited Summarization Accuracy:**
   a) Pegasus, like other huge language models, is not flawless. The produced summaries may not always capture all of the details of the original article.
   b) The accuracy of the summaries can be influenced by variables such as the intricacy of the article's language or the inclusion of factual inaccuracies in the original article.
3) **Sentiment Analysis Shortcomings:**
   a) Sarcasm, irony, and complicated phrase constructions can all be difficult for sentimen t analysis programs to detect.
   b) The emotion ascribed to the summary may not always fully represent the overall tone of the article.
   Furthermore, sentiment analysis frequently uses a binary (positive/negative) or three-way (positive/negative/neutral) categorization, which may not capture the entire range of emotions expressed in the text.
4) **Lack of Contextual Understanding:**
   a) The model lacks a thorough grasp of the financial markets or the individual companie s/cryptocurrencies discussed in the articles.
   b) This might result in summaries that lack context or overlook important financial facts.

CONCLUSION

This stock news summarizer helps busy investors and traders by automatically:

- Finding relevant articles based on user-supplied stock or cryptocurrency tickers.
- Creating succinct summaries that capture the main themes of each story, saving consumers time from reading long news articles.
- Sentiment analysis is used to determine the general tone (positive, negative, or neutral) surrounding a specific stock or cryptocurrency, which helps evaluate market perception.
- Overall, it simplifies the process of keeping up with the newest financial news, helping consumers to make educated investing decisions.