CSPB 3112 Professional Development in CS
Matthew Scott
August 12, 2024
Github: https://github.com/gr8tscott/gr8tscott.github.io/blob/main/README.md

Final Project Report: Stock Sentiment Analysis Using OpenAI

Introduction

In the modern financial landscape, timely and accurate sentiment analysis of news articles can significantly impact investment strategies. This project leverages the OpenAI platform to analyze news articles about publicly traded companies, aiming to provide actionable sentiment—buy, sell, or hold. The project not only seeks to enhance my proficiency with OpenAI's tools but also to bridge a personal interest in stock market analysis with advanced artificial intelligence techniques.

Background

Vision and Motivation

The core vision of this project is to build an application that utilizes the OpenAI framework to analyze news articles and provide sentiment predictions. This initiative was inspired by my earlier experience as a stockbroker, where I manually assessed numerous articles daily to gauge public sentiment for stock trading. The idea of automating this process through a code-driven approach has been a longstanding goal, merging personal interests with professional development in software engineering. This project represents a culmination of years of self-teaching, learning about algorithmic trading, and transitioning into software engineering.

Goals

The specific goals of the project include:

- Gaining expertise in OpenAI's development tools.
- Enhancing skills in web/data scraping.
- Improving proficiency with APIs.
- Deploying the application for accessibility via web or app to showcase in interviews.

Methodology, Materials, Goals, & Methods

Methodology and Materials

1. **Tools and Technologies:**
    - **OpenAI**: For sentiment analysis of news articles.

- **Web Scraping**: To extract relevant content from news websites.
        - **APIs**: To fetch stock price data and interact with external services.
        - **Flask**: For developing the web application.
        - **PostgreSQL**: To manage and store stock data.
  2. **Development Process:**
        - **Web Scraping**: Implemented to fetch articles from CNBC.
        - **Sentiment Analysis**: Integrated OpenAI to analyze article content and determine sentiment.
        - **Database Management**: Utilized PostgreSQL to store and manage stock data.
        - **Deployment**: Deployed the application on Render with a PostgreSQL database.

Methods

  1. **Data Collection**:
        - Fetch stock-related news articles using web scraping techniques.
        - Use the Polygon API to retrieve current stock prices.
  2. **Data Processing**:
        - Analyze the scraped article content using OpenAI to derive sentiment.
        - Store and manage the processed data in a PostgreSQL database.
  3. **User Interface**:
        - Developed a web interface using Flask where users can input URLs, view sentiment analysis results, and track stock predictions.

Results/Discussion

Results

The project successfully implemented the core functionalities:

- **Sentiment Analysis**: The application accurately analyzed the sentiment of news articles using OpenAI.
- **Data Management**: Integrated PostgreSQL for effective data management.
- **Web Scraping**: Extracted relevant news content from CNBC articles.

However, the project faced limitations:

- **Data Scraping Challenges**: Initially limited to scraping CNBC articles, which restricted the ability to analyze multiple companies simultaneously. These issues still persist and will be addressed in future iterations.
- **Deployment Issues**: Encountered connectivity and performance issues with the PostgreSQL database on Render.
- **API Use**: Both OpenAI and Polygon API had limitations in how many API calls that could be made per minute/day which slowed development progress can added to project cost for subscriptions.

- ○ OpenAI has a certain amount of credits/tokens you need to buy in order to use their tools. The more characters in your prompt and the more processing power needed to complete it.
- ○ Financial API Polygon restricted users to 5 API calls per minute and only 25 per day for their free version which would be consumed very quickly during testing phases.

Solutions to some of these limitations:

- **Prompt Crafting**: The more characters (letters, numbers, spaces, etc) that you give the AI model to analyze, the more tokens that it takes to analyze your input and craft and output, increasing cost and time. The solution to this was a carefully crafted prompt that specified in as few words as possible the goal and minimized the output. For example:
  - ○ **Prompt**: "You analyze news articles and give a sentiment: 'Buy', 'Sell', or 'Hold' based on the content. Output only the company ticker symbol and colon followed by your sentiment grade. Do not summarize."
  - ○ **Input**: "Delta Airlines was severely impacted by the technology outages caused by…"
  - ○ **Output**: "DAL: Sell"
- **Monetary Investment**: I had hoped to spend as little as possible on this project but found that every stage from deployment of the webpage and hosting the database to setting up the APIs were too limited for Free versions so I upgraded to the base level paid versions. Total investment: ~$65.

Discussion

The project demonstrated successful integration of OpenAI for sentiment analysis and effective use of web scraping and database management. Despite these successes, challenges with data scraping and database connectivity highlighted areas for improvement. I found that the action of scraping data is fairly simple; the practice of refining that data so that it is more attractive and consumable is much more challenging. Moving forward, the plan includes expanding the data sources beyond CNBC and optimizing the application's performance to handle larger datasets and multiple companies concurrently.

The project also achieved its goal of providing a functional and visually appealing application, showcasing the ability to analyze stock sentiment and manage stock data efficiently. Future work will focus on addressing current limitations and further refining the application's capabilities as well as tinkering with occasional bugs and errors.

Deployed Site (sometimes finicky, working on debugging, see images below for use case)

First Look: Database of current stocks displayed that were added weeks ago. Tracking of whether the initial prediction to buy, sell or hold was correct.

← → C   gr8tscott-github-io.onrender.com

Stock Sentiment App v1.0

Paste a URL from CNBC here

SEND URL TO AI

Your input:

Article Title:

Article Content:

AI Response:

Stock Predictions

Ticker	Current Price	Traded Price	AI Response	Date	Correct Prediction
AAPL	$217.70	$218.54	Buy	2024-07-24	False
GOOGL	$162.26	$172.63	Hold	2024-07-24	False
MSFT	$406.30	$428.90	Sell	2024-07-24	True
LUMN	$4.62	$1.52	Buy	2024-07-24	True

REFRESH TABLE

After Entering a News URL: Article is scraped, cleaned, analyzed by AI, API call for last stock price, added to database, and finally displayed.

← → C   gr8tscott-github-io.onrender.com

Stock Sentiment App v1.0

https://www.cnbc.com/2024/08/09/elizabeth-warren-letter-to-tesla-chairwoman-robyn-denholm-about-musk.html

SEND URL TO AI

Your input: https://www.cnbc.com/2024/08/09/elizabeth-warren-letter-to-tesla-chairwoman-robyn-denholm-about-musk.html

Article Title: Tesla's board faces questions from Elizabeth Warren about Musk's corporate 'entanglements'

Article Content: TSLAX owner Elon Musk (L), and Sen. Elizabeth Warren (D-MA).ReutersSen. Elizabeth Warren, D-Mass., is asking the Tesla board if it's looked into CEO Elon Musk's use of company resources to benefit his other ventures, including SpaceX and xAI.Tesla's Board of Directors appears to be failing to meet its fiduciary duties to Tesla's shareholders by neglecting to address company CEO Elon Musk's apparent conflicts of interest, Warren wrote in a 10-page letter to Tesla Chairwoman Robyn Denholm on Thurs

AI Response: TSLA: Sell

Current stock price for TSLA: $197.6

Stock Predictions

Ticker	Current Price	Traded Price	AI Response	Date	Correct Prediction
AAPL	$217.70	$218.54	Buy	2024-07-24	False
GOOGL	$162.26	$172.63	Hold	2024-07-24	False
MSFT	$406.30	$428.90	Sell	2024-07-24	True
LUMN	$4.62	$1.52	Buy	2024-07-24	True
TSLA	$197.60	$197.60	Sell	2024-08-12	False

REFRESH TABLE