

Victor Morgan

10/25

CSC-570

Dr. Osunmakinde

[YouTube Video](#)

Conversion to my AI field of choice

For my Natural Language Processing program I naturally fitted it to my theme of finance. I wanted it to be adaptable to certain problems only faced in the area of finance. This program's goal is to be able to analyze a website provided by the user, or text, and give them the sentiment of the text surrounding the company. The user can then use the sentiment of the article to decide if they want to invest in that company's stock, not invest, or do further research.

The PEAS are as follows:

Performance : The agent's ability to extract meaningful words from the article and correct identification of the article's sentiment.

Environment : The HTML page provided or the direct text provided

Actuators : Predicting the text's sentiment based on the remaining meaningful words

Sensors : The raw html or text input. The list of stopwords, custom and default. NLTK's Vader's lexicon and pre-defined sentiment.

Since I'm dealing with stocks again, I found a lot of the websites have antiscraper frameworks implemented, which go directly against my program's ability to pull their words. I designed my program to be resilient against denials and allow for elevated user inputs and decisions

Feature Engineering

For this assignment I continued to focus on the user-friendly aspect of how my AI should function, continuing to make my program more flexible, and resilient to denials of access

I first wanted the user to be able to provide input on words they also consider meaningless. I noticed things like punctuation and words like "That", "Was", and "a" getting past, so the user is now able to curate a list of words that will also be removed from both bodies of text provided by the user. This provides more accurate sentiment analysis based on how the user views the nature of the text they provided.

I also implemented a backup for the event that my agent is denied access to a website. Upon being denied access, my agent handles the error instead of crashing, notifies the user of the error, and then allows them to enter some of the text they've copied from the website and proceed with the analysis as normal.

In conjunction with allowing the user to enter their own text manually to be analyzed, I implemented headers which will allow my agent to get into certain websites, but I have noticed this only works once or twice per website.

I also implemented a nltk module called VADER which I used for the analysis of the sentiment of certain words. I then used that to decide if the overall text is negative or positive. Since I'm also in the domain of finance, I manually added a handful of words that might not be considered in the same way by VADER. This makes the lexicon more aware of finance based words for more accurate sentiment analysis.

Conclusion

In conclusion I notice that finances face stricter restrictions when it comes to allowing the analysis of external AI devices. This requires a more strategic approach when it comes to designing AI for these analyses. Through my program, I've also learned to express my wordcloud as qualitative data, scoring each word that is meaningful to decide if the text is positive, negative, or neutral. I've implemented resilience to allow for users to still analyze websites that have restrictions that render my agent unable to access it.

I also wanted to remain a certain level of resilience for my agent, so that in the case that it encounters a website it's unable to access, it can navigate the situation by allowing the user to input their own text to be analyzed. In the future I'd like to navigate this situation by allowing the user to also check a local file that the user may have been able to download containing the same information to be analyzed.