Sentiment Analysis of Stock Market in Python (Part 2)- Estimating Sentiment Scores with NLTK

Bee Guan Teo Follow

Oct 7, 2021 · 5 min read ★

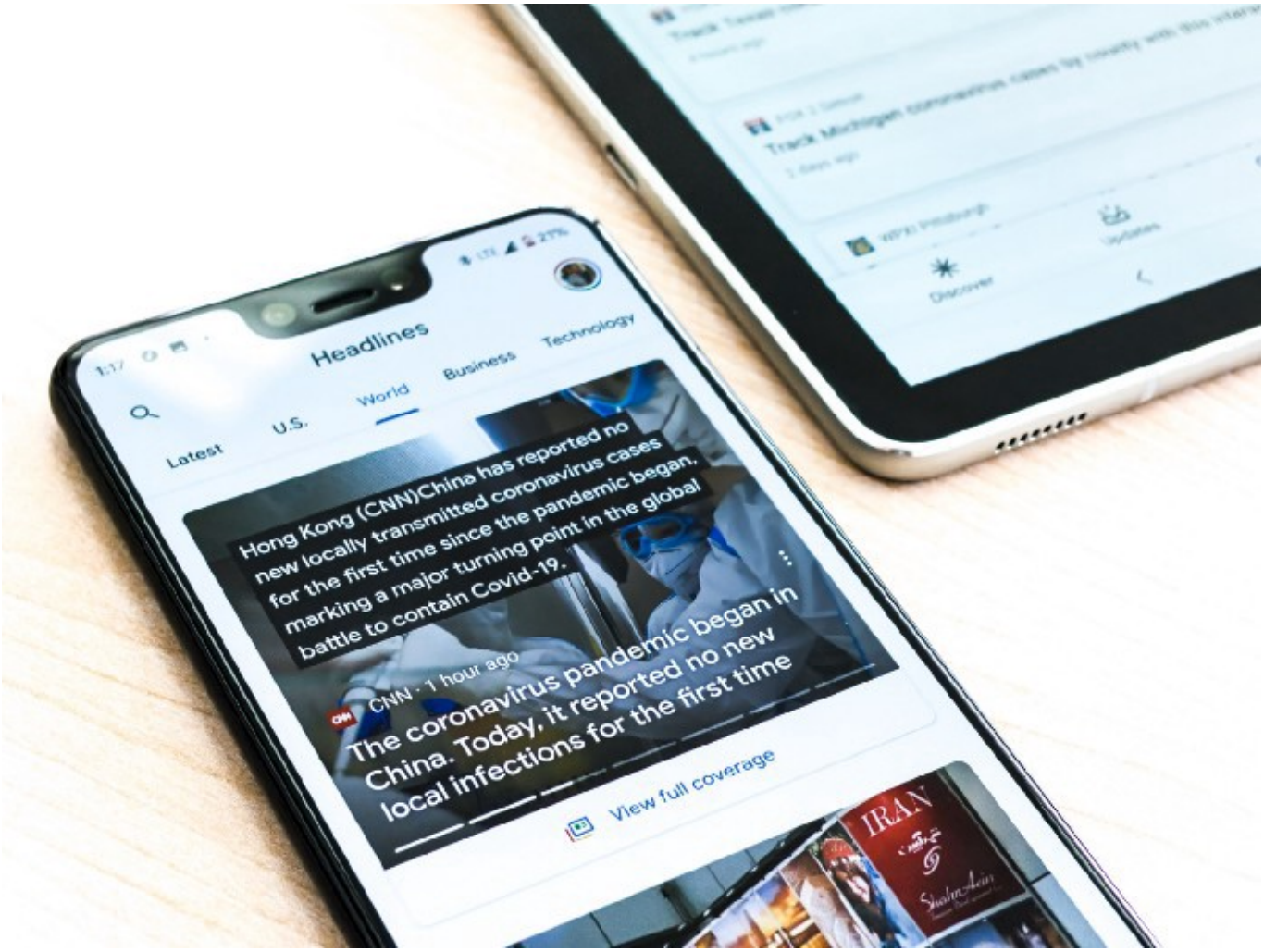


Photo by Obi Onyeador on Unsplash

This article is the second part of my earlier article about sentiment analysis of the stock market in Python. If you haven't gone through the first part of my article, you are welcome to visit the previous article to gain the overall picture of the topic:

Sentiment Analysis of Stock Market in Python (Part 1)- Web Scraping Financial News

In this Part 2 article, we are going to proceed with the pre-processed news data from Part 1 and use the **Python NLTK package** to calculate the sentiment score.

***Disclaimer**: The writing of this article is only aimed at demonstrating the steps to perform stock market sentiment analysis in Python. It doesn't serve any purpose of promoting any stock or giving any specific investment advice.*

Github

The original full source codes presented in this article are available on my Github Repo. Feel free to download it (*SentimentAnalysis_part2.py*) if you wish to use it to follow my article.

Estimating Sentiment Score with NLTK

From Part 1, we have already acquired news data from **Financial Modeling Prep (FMP)**, process it and store it in a dataframe.

date	ticker	title	text
2021-10-06 06:13	AAPL	EXG: American Market Continues To Be Overvalued, Go Abroad With This Fund	EXG: American Market Continues To Be Overvalued, Go Abroad With This Fund
2021-10-06 05:16	AAPL	Double Your Money With Zero Investing Experience	Here's a two-step plan.
2021-10-05 16:56	AAPL	Why Apple Stock Bounced Higher Today	Two Wall Street analysts offer two reasons to buy Apple stock.
2021-10-05 16:11	AAPL	What Are The Investors Doing With Apple	Someone with a lot of money to spend (and p…
2021-10-05 14:35	AAPL	Why Apple Shares Are Rising	Shares of companies in the broader technolo…
2021-10-05 12:14	AAPL	The Margin: Remembering Steve Jobs: Apple, …	Steve Jobs has been gone for a decade now, but he certainly hasn't been forgotten.
2021-10-04 16:31	AAPL	What Are The Investors Doing With Apple	Someone with a lot of money to spend (and p…
2021-10-04 16:03	AAPL	Why Apple Stock Got Bit Today	WSJ highlights the good and bad of Apple's gargantuan gaming business.
2021-10-04 15:09	AAPL	Top 5 Most Popular Stocks Hedge Funds Are Buying	The stock market saw endless turmoil last y…
2021-10-04 14:52	AAPL	Apple stock falls into 'correction' territory as it leads the Dow's losers	Shares of Apple Inc. AAPL, -2.52% dropped 2…

Image Prepared by the Author

Now we are going to use **VADER Sentiment Analyzer** from the Python NLTK package to estimate the sentiment score of that news. VADER analyzes a text and generates a sentiment score by checking the existence of positive, negative and neutral terms in the text.

1. Get Started with VADER

To ease our understanding, let us try to use VADER to estimate the sentiment score just for one news.

```
1  import requests
2  from bs4 import BeautifulSoup
3  import pandas as pd
```

```
4  from nltk.sentiment.vader import SentimentIntensityAnalyzer
5
6  ticker = "AAPL"
7  url = "https://financialmodelingprep.com/financial-summary/" + ticker
8  request = requests.get(url)
9  print(request.text)
10
11 parser = BeautifulSoup(request.text, "html.parser")
12 news_html = parser.find_all('a', {'class': 'article-item'})
13 print(news_html[0])
14
15 sentiments = []
16 for i in range(0, len(news_html)):
17     sentiments.append(
18             {
19                 'ticker': ticker,
20                 'date': news_html[i].find('h5', {'class': 'article-date'}).text,
21                 'title': news_html[i].find('h4', {'class': 'article-title'}).text,
22                 'text': news_html[i].find('p', {'class': 'article-text'}).text
23             }
24         )
25
26 df = pd.DataFrame(sentiments)
27 df = df.set_index('date')
28
29 analyser = SentimentIntensityAnalyzer()
30 print(df['text'][4])
31 print(analyser.polarity_scores(df['text'][4]))
```

SentimentAnalysis_part1.py hosted with ♡ by GitHub view raw

Line 1–27: Previous scripts to web scrape financial news data and store them in a dataframe. (*You may refer to my Part 1 article for further details*)

Line 29: Create a VADER sentiment analyzer.

Line 30: Display the fifth news text (accessed by index-4).

Line 31: Use the VADER analyzer's *polarity_scores* method to estimate the sentiment scores for the fifth news text and display the scores.

```
Shares of companies in the broader technology sector, including Apple Inc. (NASDAQ: AAPL) are trading higher,
rebounding after dipping on Monday following a rise in yields. The ris...
{'neg': 0.0, 'neu': 0.922, 'pos': 0.078, 'compound': 0.296}
```

Image Prepared by the Author

The sentiment scores are split into four parts:

1. **neg** — Negative score
2. **neu** — Neural score
3. **pos** — Positive score
4. **compound** — Normalized, weighted composite score.

The **compound score** is our target sentiment score which is ranged from -1 to 1. We can set a threshold that if the compound score is above 0.05, the sentiment is overall positive. The higher the score, the better the sentiment. In opposite, if the compound score is lower than -0.05, the sentiment is negative. If the score is between -0.05 to 0.05, the sentiment is neutral (neither positive nor negative).

In the case above, we can see the sentiment score (or compound score) is prone to positive as it is bigger than 0.05. This can be due to the existence of some positive terms in the text such as “rebounding”.

2. Applying VADER on All News Records

Now, we will use VADER to generate the sentiment scores for all our news records in the dataframe.

```
1  def calc_sentiment_score(text):
2      return analyser.polarity_scores(text)["compound"]
3
4  analyser = SentimentIntensityAnalyzer()
5  df['sentiment_score'] = df['text'].apply(calc_sentiment_score)
```

SentimentAnalysis_part2.py hosted with ♡ by GitHub view raw

Line 1–2: Define a function, *calc_sentiment_score*, that will accept news text as input and use the VADER analyser’s *polarity_scores* method to compute the sentiment score and return it as an output.

Line 4: Create a new object of Sentiment Analyzer.

Line 5: Apply the *calc_sentiment_score* function on the news text column in the dataframe. This will invoke the function to sequentially work on the news text from the dataframe and generate sentiment score (compound score) for each of the news. The resulting list of sentiment scores is assigned to a new column of the dataframe, *sentiment_score*.

date	ticker	title	text	sentiment_score
2021-10-06 17:31	AAPL	Apple Vulnerable To EU Antitrust Action - Read Why	Apple Inc's (NASDAQ: AAPL) is susceptib…	-0.5622
2021-10-06 15:13	AAPL	October an 'average' month for stocks, Albion Financial Group CIO says	Jason Ware, Albion Financial Group part…	0.4215
2021-10-06 14:43	AAPL	With return of 'Report a Problem' on App St…	Apple is making a change to its App Sto…	0
2021-10-06 14:35	AAPL	Tech Wreck: Buy These 3 Tech Blue Chips At A Discount	Interest rates continue to rise, and te…	0.4215
2021-10-06 14:13	AAPL	U.S. Justice antitrust nominee says eager to tackle more than just Big Tech	Jonathan Kanter, the third of three pro…	0.2023
2021-10-06 13:02	AAPL	Mega Caps Have Tanked - A Perfect Time To Buy Vanguard's MGK ETF	Mega Caps Have Tanked - A Perfect Time To Buy Vanguard's MGK ETF	0.5719
2021-10-06 06:13	AAPL	EXG: American Market Continues To Be Overvalued, Go Abroad With This Fund	EXG: American Market Continues To Be Overvalued, Go Abroad With This Fund	0
2021-10-06 05:16	AAPL	Double Your Money With Zero Investing Experience	Here's a two-step plan.	0
2021-10-05 16:56	AAPL	Why Apple Stock Bounced Higher Today	Two Wall Street analysts offer two reasons to buy Apple stock.	0
2021-10-05 16:11	AAPL	What Are The Investors Doing With Apple	Someone with a lot of money to spend (a…	0

Image Prepared by the Author

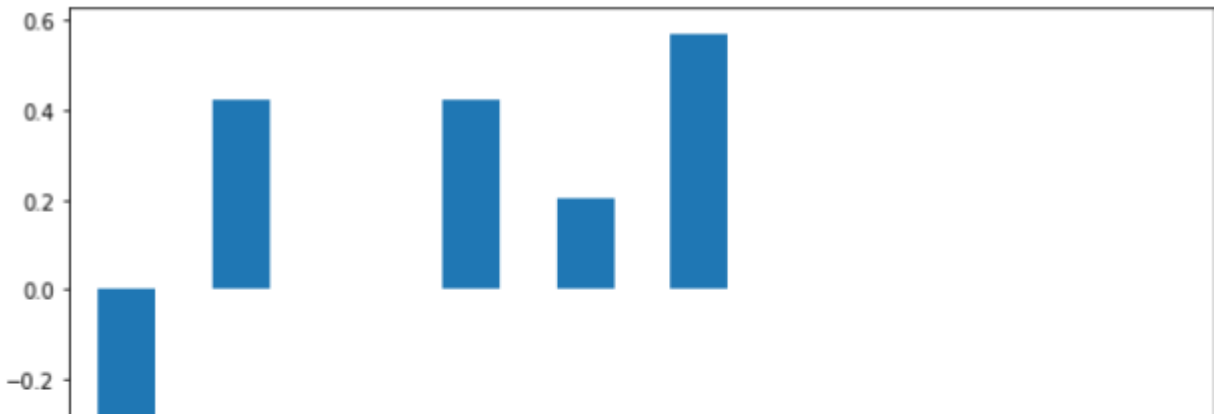
3. Visualizing Sentiment Scores on Graph

At last, we plot a bar chart of the sentiment scores to ease our interpretation and analysis of the result.

```
1   df['sentiment_score'].plot(kind='bar', figsize=(10,5))
```

SentimentAnalysis_part3.py hosted with ♡ by GitHub view raw

Line 1: Use the dataframe in-built plot method to create a bar chart for the sentiment scores.



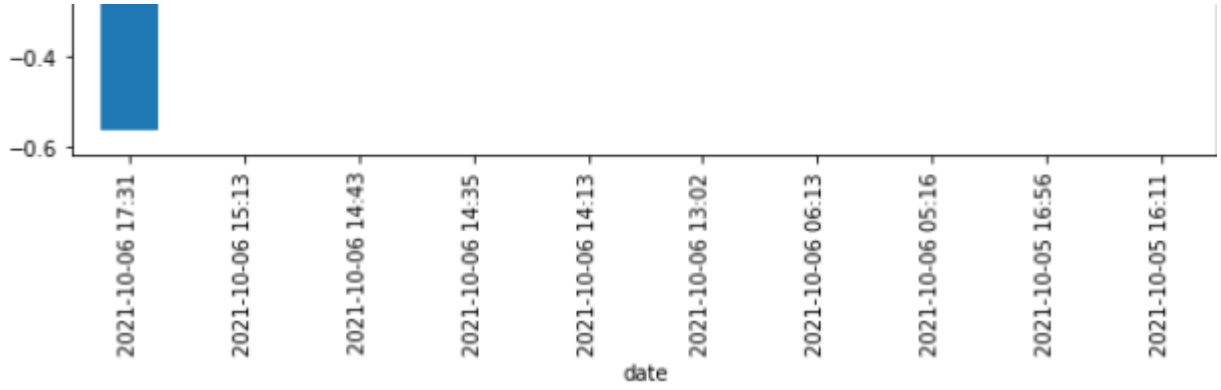


Image Prepared by the Author

From the plot above, we can observe that the recent sentiments about AAPL are quite positive. The only negative sentiment is related to an antitrust issue in the EU. The rest of the sentiments are neutral.

However, this is important to note that we can't solely depend on the sentiment scores to make our investment decision. Some common keywords in the finance context such as "bear" or "bull" might be treated as neutral terms by VADER Analyzer. Some seemingly positive keywords such as "rise" mustn't be a good sign to either buy or sell an asset. In short, we will always need to verify the score with the news context.

Conclusions

We have managed to go through these two series of articles about the sentiment analysis of the stock market in Python. The sentiment analysis presented here is based on a lexicon-based approach which is to **calculate orientation for a given text from the semantic orientation of words or phrases in the document.**

While the resulting sentiment scores may not be completely reliable due to the potentially biased evaluation of some words' usage, they still project a rough picture of the current market condition for an asset. This can be useful to help us plan for an investment at the right timing.

I hope you enjoy reading this article.

Subscribe to Medium

If you like my article and would like to read more similar articles from me or other authors, feel free to subscribe to Medium. Your subscription fee will partially go to me. This can be a great support for me to produce more articles that can benefit the community.

Reference

1. https://www.nltk.org/api/nltk.sentiment.html