

Twitter sentiment analysis of the most volatile stocks of the moment

1st Juan José Mejías

Data Science

IT Academy

2022

<https://github.com/Huanhotze>

Abstract—Extracting the data from a financial website, we located the most volatile stocks of the moment. Once we have determined which stocks we are going to work with, we will analyze what is the sentiment on twitter about them. Once the results are obtained and as a final objective, we share it in a Telegram group.

Index Terms—Scraping, Sentiment, Twitter, Telegram

I. INTRODUCTION

Within the investment world there are many ways to invest. One of them could be the analysis of the general opinion about a specific stock. Our purpose in this project is to create a market sentiment indicator.

Using the tools provided by Python to analyze text sentiment, we will analyze the market sentiment at a given time on a given set of stocks. Python's NLTK library will help us in this purpose.

Before performing the analysis we must obtain an analyzable dataset with valuable data. And with the sentiment information we could inform investors who are interested in obtaining this information.

Therefore, our project is separated into three parts or into three minor objectives to achieve a major objective:



Fig. 1. Flow information

- 1) Extract data from an investment website of the most volatile companies of the moment. In this particular case Investing.com.
- 2) Sentiment analysis in the social network of Twitter.
- 3) Reporting the result of the analysis through an instant messaging application, Telegram.

II. METHODOLOGY

As mentioned above, the project is composed of three parts, each one dependent on the previous one.

A. Extracting the most volatile stocks

Our goal is not to extract sentiment from the market or from a random stock. Our goal is to focus on the most volatile stocks at the moment. Therefore, first we will have to locate which companies are the most volatile at the moment. Within investing we find the section of the most volatile companies in the USA: Most Active - United States Stocks .

We scraped the website to get the complete table of "most active stocks" which we will pass to a dataframe.

Most Active - United States Stocks									
United States									
52 Week High 52 Week Low Most Active Top Gainers Top Losers									
Price Performance Technical Fundamental									
Name	Last	High	Low	Chg.	Chg. %	Vol.	Time		
Microsoft	234.90	238.30	230.17	-15.77	-6.29%	59.65M	13:45:29		
Tesla	226.63	230.60	218.20	+4.21	+1.90%	60.56M	13:45:30		
Apple	149.35	151.99	148.88	-2.99	-1.96%	52.61M	13:45:29		
Alphabet A	96.39	98.31	95.15	-8.09	-7.74%	58.10M	13:45:28		
Amazon.com	115.94	119.35	114.76	-4.66	-3.86%	46.77M	13:45:30		
NVIDIA	129.68	133.88	127.08	-2.93	-2.21%	37.42M	13:45:29		
Alphabet C	96.70	98.53	95.38	-8.23	-7.84%	44.72M	13:45:29		
Netflix	302.09	305.63	288.05	+11.07	+3.80%	11.21M	13:45:29		

Fig. 2. Screenshot investing table

Once we have the scraped table we will introduce it in a dataframe only with the negotiation volumen and we will add the ticker symbol. ("A ticker symbol, also called a stock symbol, is a unique code that represents a company listed on a stock exchange." [9])

	Nombre	Volumen	Ticker
15	chipotle mexican grill	629.79	CMG
1	tesla	59.95	TSLA
0	microsoft	59.42	MSFT
3	alphabet a	57.88	GOOGL
9	amd	55.55	AMD

Fig. 3. Dataframe Top 5 volatile stocks

To finish we will sort it by volatility (Volumen column) to keep the 5 most volatile ones.

B. Twitter sentiment

With the 5 most volatile stocks and with the help of the Twitter API, we extract 1000 Tweets from each of the stocks. At this point we have 5 different tables with 1000 lines each. We will work on each table separately, processing the text before applying the NLTK model. This information will go through data cleansing and preprocessing:

- Removal of unnecessary characters.
- Pass all characters to lowercase.
- Tokenize - Pass texts to a list of words.
- Eliminate very common and meaningless words (the, in, a, etc...).
- Stemming - Leave only the root of the words.
- Make all characters lowercase.

Once the text is preprocessed we will pass the NLTK model. NLTK will give us 4 different values. We will use the value 'compound'. Compound is a number that goes from -1 to +1 and gives us a negative sentiment at -1 and a positive sentiment at +1.

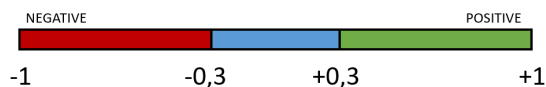


Fig. 4. Compound interval negative vs positive

Assuming a neutral range between -0.3 and +0.3. Anything to the left of that range will be negative market sentiment on the stock and anything to the right will be positive.

C. Send stock sentiment to Telegram

For the final step we will use a specific Telegram library to send the results of our analysis. After creating and configuring a Telegram bot with @BotFather we will use the Telegram Python library 'pytelegrambotapi'.

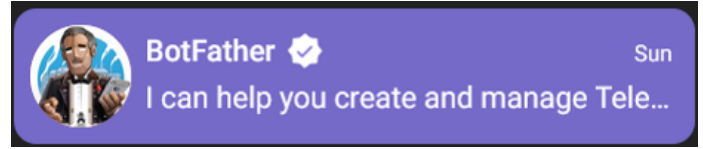


Fig. 5. Telegram Bot to create Bots

This library will allow us to send messages to Telegram to the channel we have created for this purpose "Stock Sentiment Indicator. The message received in the channel will consist of the stock ticker and the sentiment represented by an emoticon:

Positive -> 👍
 Neutral -> ↔
 Negative -> ❌

Fig. 6. Emoticons to identify sentiment of actions

III. RESULTADOS

Finally, the purpose of the project has been achieved. Our project consisted of 3 minor objectives to achieve a major goal.

- Extracting the most volatile stocks.
- Twitter sentiment analysis.
- Send stock sentiment to Telegram.

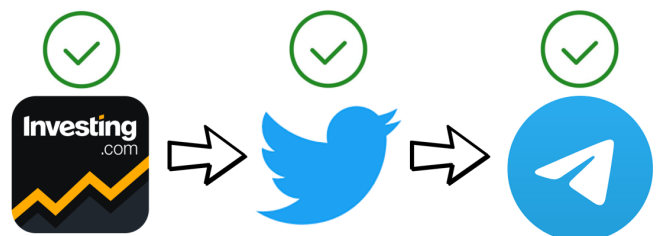


Fig. 7. Three minor goals archived!"

The major goal is archived, to receive a message in Telegram with the sentiment indicator of each of the selected stocks according to their current volatility.

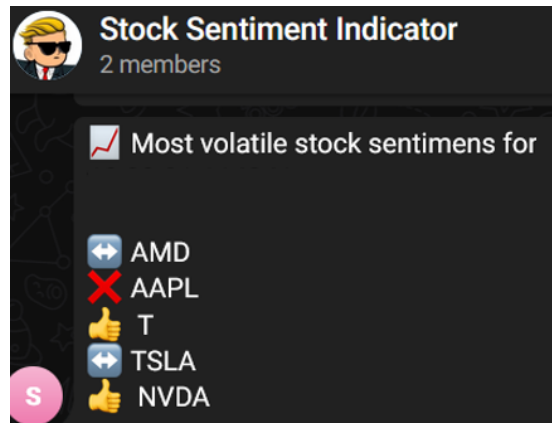


Fig. 8. Screenshot from Telegram Channel "Stock Sentiment Indicator"

IV. CONCLUSIONS

I Don't know if we can get rich with this indicator, but faith is the last thing to be lost.

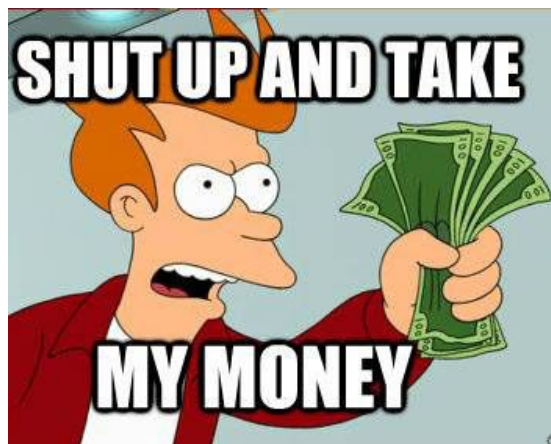


Fig. 9.

We have fulfilled the 3 necessary objectives in order to achieve the main objective, a stock sentiment indicator.

In the process we have made:

- Scraping from an investment page.
- Passed the information to a dataframe.
- Selected the stocks we were interested in according to our criteria.
- Extracted the Tweets related to these stocks.
- Adapted the text of the Tweets to be able to apply NLTK.
- Applied the NLTK model to find out the sentiment of each stock.
- Extracted the NLTK results and built a text understandable by other users.
- Sent this text by Telegram.

We have compared the results of our telegram bot with the actual change in the markets of the most volatile stocks.

Let's see if our "Stock Sentiment Indicator" matches the real market data:

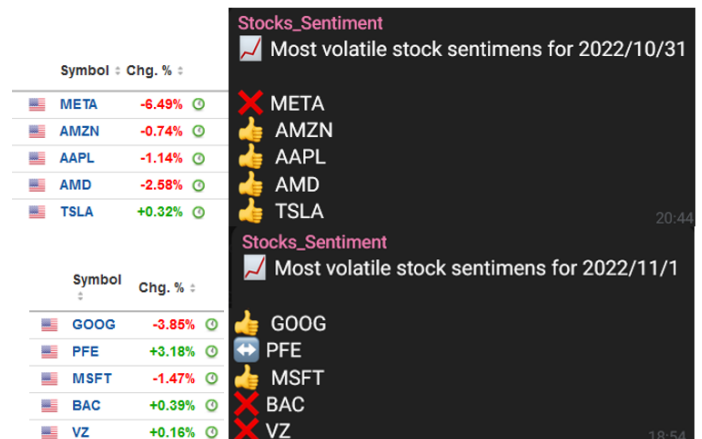


Fig. 10. Real market daily change vs sentiment indicator

When we compare our results, sentiment of the most volatile stocks on Twitter, with the actual variation of these stocks on the stock market, we see that it does not necessarily match.

We regret to have to conclude that, unfortunately, it is not advisable to use this tool to make investment decisions.

What a pity, we are not retiring from this one either!

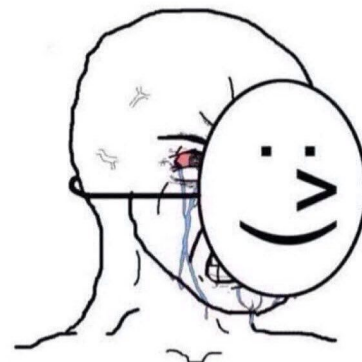


Fig. 11. Sad!

REFERENCES

- [1] NLTK Python library
- [2] Scrapy Python library
- [3] Selenium Python library
- [4] Telegram Bot API
- [5] Twitter Bot API
- [6] Youtube - Computer Science
- [7] Youtube - FRIKI del TO
- [8] Youtube - John Watson Rooney
- [9] What is a Ticker Symbol?