

Final Year Undergraduate Project 2021/22
Queen Mary University of London

WallStreetBets Stock Sentiment Analysis

Interim Report



Bishes Limbu
School Of Electronic Engineering and Computer Science

ABSTRACT

A report for a Natural Language Processing project that aims to analyze a sentiment of a stock from a subreddit (sub-forum) called, WallStreetBets (WSB), on the popular forum site Reddit.

Contents

- ABSTRACT 2
- CHAPTER 1: INTRODUCTION 4
 - 1.1 INTRODUCTION 4
 - 1.2 PROBLEM STATEMENT 4
 - 1.3 AIM 4
 - 1.4 OBJECTIVES 5
 - 1.5 USE CASES..... 5
 - 1.6 RESEARCH QUESTIONS 5
 - 1.7 REPORT STRUCTURE 5
- CHAPTER 2: EDUCATION 6
 - 2.1 LITERATURE REVIEW 6
 - 2.1.1 THE EFFECT OF SOCIAL MEDIA ON THE STOCK MARKET 6
 - 2.1.2 TRADITIONAL LONG-TERM INVESTING VS TRADING ADVICE FROM WSB..... 6
 - 2.1.3 EFFECTIVENESS OF SENTIMENT ANALYSIS ON PREDICTING STOCK MOVEMENTS 6
 - 2.2 EXISTING APPLICATION REVIEWS..... 7
 - 2.2.1 MEMESTOCKS.ORG..... 7
 - 2.2.2 YOLOSTOCKS.LIVE..... 7
 - 2.2.2 TALKWATER 8
 - 2.3 SUMMARY 8
- CHAPTER 3: ANALYSIS 9
- Bibliography 10

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

Due to the accessibility of the internet and the rise of social media, retail investors discuss in large numbers about the stock market on social media platforms. Discussions can range from sharing their extravagant successes and plummeting losses or sharing their theories and predictions about a certain stock. Reddit, one of the most popular forum sites has been a cornerstone of these discussions, with many users flocking to the site daily, to read or share content. Some theories about a selected stock can become very popular and when the market responds with rising prices that appear to confirm the predictions, it can create a large cult-like following and a wave of retail investors buying the selected stock, raising the price of the stock dramatically. These stocks are labeled “meme stocks” (Hayes, 2021) and during the latter half of 2020 the stock market was struck by many of these meme stocks, where the followers have managed to coin a hefty profit on their positions for the stock.

1.2 PROBLEM STATEMENT

Other than reading through hundreds of posts and comments on Reddit, there are no apps or an easier method to quickly evaluate the sentiment around a surrounding stock. There are websites available that include the number of mentions a certain stock has garnered but they do not include sentiment analysis. Sentiment analysis tools and apps are largely available for other functions and social media sites. However, due to the problem being new and in a niche area, these tools are not designed to tackle this problem with meme stocks.

1.3 AIM

My web app will incorporate a web scraper PRAW (The Python Reddit API Wrapper) that scrapes posts and comments from a popular subreddit (sub-forum) on Reddit called WallStreetBets and will analyze the sentiment surrounding a selected stock by the user, by weighting the sentiment on terminology and slang used to form a score for the 3 sentiment outcomes: positive, neutral, and negative. The app will also include another analysis option, where the user can see the most-mentioned stocks as the data scraped can be used to easily calculate this as well. The app will provide a shortcut for the user, as they will not have to read through posts and comments on the forum, to track and analyze a certain stock’s popularity and potential. It will display the results of the searched stock in a graph where the user can easily understand the information and sentiment.

1.4 OBJECTIVES

- To research about the meme stocks.
- To investigate the influence of social media on stock market trading.
- To investigate and explore current options into tracking meme stocks.
- To investigate and explore current tools and apps for sentiment analysis.
- To develop a web app using Django back-end and React front-end.
- To implement a web-scrapper using PRAW (The Python Reddit API Wrapper) to get data from user posts and comments from a selected stock.
- To use the data from the scraper to analyze the sentiment around a certain stock.
- To display the results from the analysis as a graph, using python graphing tools.
- To perform user testing.
- To evaluate the effectiveness of the application.

1.5 USE CASES

- User can select a stock to search.
- User can view the sentiment around the searched stock.
- User can adjust the timeframe of the sentiment analysis.
- User can view the top mentioned stocks.
- User can view the results in a form of a graph on a website.

1.6 RESEARCH QUESTIONS

- How does social media affect stock market trading??
- How successful is investment advice from WallStreetBets?
- How effective is sentiment analysis on predicting stock price movements?

1.7 REPORT STRUCTURE

Chapter 2 will present the findings from the research conducted. It will have a literature review which will include background research and will review the existing applications available on the market.

Chapter 3 will be the analysis, where it will include the technical aspects of the project development. It will include requirements, risks, use-cases, and diagrams about the development.

CHAPTER 2: EDUCATION

2.1 LITERATURE REVIEW

2.1.1 THE EFFECT OF SOCIAL MEDIA ON THE STOCK MARKET

Andre Kostolany, one of the most successful investors of the 20th century once stated: “Facts only account for 10% of the reactions on the stock market; everything else is psychology.” Kostolany believed that the markets are highly impacted by emotional reactions. (Nasdaq, 2019)

With the emergence of social media in the last decade, the ability to share opinions via tweets, forum posts and blogs are quick and easy. People can share the viewpoints on certain stocks and opinions on the market to masses of people. Depending on the sentiment of the opinion, it can cause a widespread influence on other users on social media, generating a broadcast of information about a certain stock that can triggers emotional market reactions in a blink of an eye.

Examples of this can be seen as on Jan 27, 2021, the value of GameStop (GME) skyrocketed to \$347. The 95% change in value was attributed to a reddit forum WallStreetBets (WSB) and a user named ‘RoaringKitten’, Keith Gill. Keith Gill had recognized that hedge funds who were betting on GameStop’s demise had shorted the stock in amounts that were absurd and unnatural. He then proposed a ‘play’ that retail investors would buy the stock and hold it (not sell the stock). This in turn resulted in hedge fund’s shorts to expire, forcing them to buy back shares at losses, however when there was a following of users from WSB holding the stock, it drove up the price dramatically in a short squeeze. (Service, 2021)

2.1.2 TRADITIONAL LONG-TERM INVESTING VS TRADING ADVICE FROM WSB

There has been a study conducted that analyses the success rate and compares the investing methods and advice from WSB. Detailed analysis of stocks has been conducted, where the time frames of 1 day, 3 days, 1 week, 1 month and 3 months have been measured. On average WSB portfolio stocks have grown 3-4 times more than traditional long-term trading through the S%P 500. During the dates of January 2019 and December 2020 if an investor had followed the WSB portfolio they would have seen a growth of 16% after 3 months and even up to 32.03% if additional signals in form of posts were detected by the investor. The data shows that since 2019, with good awareness to buy and sell signals from posts on WSB, the average WSB investor outperformed the S&P500 investor. (Tolga Buz, 2021)

2.1.3 EFFECTIVENESS OF SENTIMENT ANALYSIS ON PREDICTING STOCK MOVEMENTS

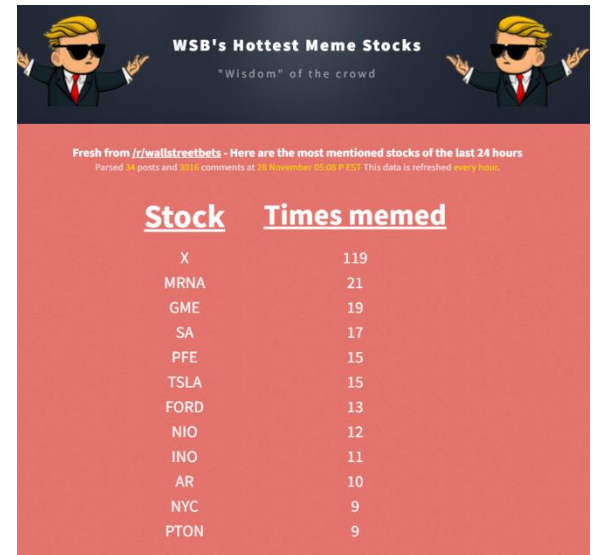
A study used tweets from twitter to judge the correlation between the sentiment of tweets about a certain stock and the price movement. A classifier was trained with words that are subjected to positive or negative sentiments about stock movement. Trained on a large data set the tweets were classified into 3 categories: positive, neutral, and negative. The accuracy of the results ranged from 69.01% to 71.82%, varying from model type. The results showing a strong correlation exists between the rise/fall of stock prices of a company to public opinions and emotions expressed on twitter through tweets. (Venkata Sasank Pagolu, 2017)

2.2 EXISTING APPLICATION REVIEWS

2.2.1 MEMESTOCKS.ORG

A simplistic webpage, that shows the amount of mentions a certain stock has garnered on WSB in the past 24 hours, with the use of PRAW (Python Reddit API Wrapper). When you click on the stock name it redirects you too the stock's yahoo page where you can view its attributes, while if you click the number of mentions it leads you to the list of posts on reddit. You can also change the order of the list by clicking the "Times memed" to start the list in ascending order, while doing the same for the stock but alphabetically.

It does not provide any graphing of the results, which can be utilized by users to understand the data in another form. Furthermore, it does not provide any sentiment analysis.



2.2.2 YOLOSTOCKS.LIVE

Produced by a COVID Modeler and Data Scientist Youyang Gu, YoloStocks track stock ticker mentions being discussed in every comment and thread from multiple subreddits and aggregates the data live. It also includes a filter mechanism called Real Mentions, which remove duplicate or similar posts to filter out bots and spam that could skew the data set.

Users can view the dataset in a top 15 table, furthermore they can choose different subreddits, other than WSB to gain data from. Additionally, there are options to pick the timeframe of data gathering. The options include, last 100 days, last 14 days and the last 24 hours. It also provides plots of the data in 2 graphs, showing the detailed mentions and timeline of posts. However, it does not include sentiment analysis.

Top 15 Tickers: r/WallStreetBets (Past 24h)

Total Real Mentions*: 2,447 (-15%

% of mentions from top 15 tickers: 49%

Rank	Ticker	Real Mentions*	Prev Rank
1	SPY	182 (7%)	1
2	GME	175 (7%)	4
3	TSLA	110 (4%)	6
4	LCID	109 (4%)	68
5	WISH	108 (4%)	3
6	Z	89 (4%)	17
7	PLTR	88 (4%)	5
8	VIX	81 (3%)	14
9	AMD	48 (2%)	16
10	MRNA	47 (2%)	2
11	QQQ	42 (2%)	22
12	NVDA	33 (1%)	7
12	V	33 (1%)	54
14	INTC	31 (1%)	44
14	BNTX	31 (1%)	10

*After removing duplicates/bots/spam

Choose a subreddit to plot:

Select...

- ☐ Last 100 days ☐ Last 14 days
☒ Last 24 hours

Data updated: November 28, 2021 05:46pm

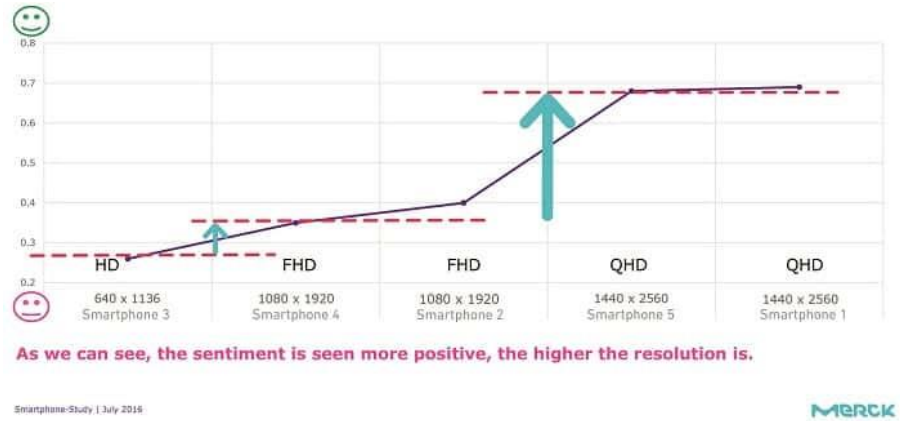
ET

(less than a minute ago)



2.2.2 TALKWATER

A service for companies where a company can search themselves on social media to learn insights and sentiments around their company. It provides a quick brand overview which allows the company to see what is being said about their brand online. Other than the sentiment it provides further statistics such as demographics, geographies, engagement, and volume. This helps companies maintain their brand reputation, as they can launch changes when sentiment about their company changes, which in turn improves customer experiences, as the you can view how customers feel about the company's product or service. Talkwater provides great graphing resources, data analytics and sentiment analysis, however it is only exclusive to companies. They do not provide the tracking of sentiment analysis for stocks.



2.3 SUMMARY

In summary, the research concludes that sentiment analysis of WallStreetBets is worthwhile, as investors with the WSB portfolio outperform, on average, a traditional investor. Furthermore, it shows there is a correlation between the sentiment of stocks on social media and the price of the stock. Additionally, there are inadequate sentiment analysis tools or apps to track these meme stocks on social media. The webapp for this project will provide that analysis tool.

CHAPTER 3: ANALYSIS

3.1 Risk Assessment

RISK	LIKELIHOOD	EFFECT	SEVERITY	PREVENTION ACTIONS
Unable to learn and use PRAW	2	Without PRAW it is near impossible to do the project as a web scrapper is essential to create the data set for this project.	5	Use resources online and documentation provided by PRAW to learn and use it effectively to scrape Reddit for data.
Unable to find a service to host the website/webapp	2	Without a service to host the webapp, the website cannot be accessed online by other and they will need to download the website and files locally to view and run the web app.	3	Find a suitable platform to host the web application.
Time Management	3	Project is incomplete and many of the features missing or not functioning properly.	5	Work on project weekly and do not neglect it.
Loss of Files	1	Project files are lost due to corruption or unforeseen circumstances.	5	Use GitHub as version control and to keep a back-up of all resources used in the project.
WallStreetBets Removal	1	The planned subreddit can be removed by reddit if deemed necessary. Will cause issues with the project as WSB is the subreddit that has been researched about.	2	No prevention actions can be done, however incase if it does get removed you can switch subreddits.
Unable to find a suitable Python Data Visualization module	1	Without the module, the graphing would be impossible, resulting in the app missing a key feature.	5	Find a suitable graphing module.

Bibliography

Hayes, A., 2021. *Meme Stock*. [Online]

Available at: <https://www.investopedia.com/meme-stock-5206762>

[Accessed 24 11 2021].

Nasdaq, 2019. *How Does Social Media Influence Financial Markets?*. [Online]

Available at: <https://www.nasdaq.com/articles/how-does-social-media-influence-financial-markets-2019-10-14>

[Accessed 24 11 2021].

Service, B. N., 2021. *Social stock: how social media and influencers affect financial markets*. [Online]

Available at: <https://bunewsservice.com/social-stock-how-social-media-and-influencers-affect-financial-markets/>

[Accessed 25 11 2021].

Tolga Buz, G. D. M., 2021. *Should You Take Investment Advice From WallStreetBets? A Data-Drive Approach..* [Online]

Available at: <https://arxiv.org/ftp/arxiv/papers/2105/2105.02728.pdf>

[Accessed 26 11 2021].

Venkata Sasank Pagolu, K. N. R. G. P. B. M., 2017. *Sentiment analysis of Twitter data for predicting stock market movements*. [Online]

Available at:

https://ieeexplore.ieee.org/abstract/document/7955659?casa_token=3WT2WEKzZTAAAAAA:qPEepUxYbkme9zo8EOWQVlykqtKOLB4gmzZI4Qd3ha2Pf09rvLp6aQS2wZAOvpn17wsNz0Uw

[Accessed 26 11 2021].