

# *How Reddit Sentiments Influenced Gamestop Stocks*

Jiaming Pang<sup>1,a,\*</sup>

<sup>1</sup>Westminster school, 995 Hopmeadow Street, Simsbury, CT 06070, US

a. apang24@westminstertools.org

\*corresponding author

**Abstract:** This paper focuses on the language usage in the GameStop (GME) share rally in January 2021. Specifically, I analyze how comments in a leading post of Reddit's r/wallstreetbets subreddit reflect the reddit user community's sentiment and emotion with various linguistic expressions. I analyzed the topics of the comments using topic modeling with Latent Dirichlet allocation algorithm, which objectively identified representative topics in the comments. Results showed that there were six clusters of topics, including two of the topic clusters that represented words that were highly correlated with stock market behaviors, such as "buy" and "sell", one cluster that suggested objective analyses of the stock market, and one cluster that presented complaining emotions. The findings indicate a direct correlation and impact between the sentiments of comments and the performance of the stock market.

**Keywords:** sentiment analysis, GameStop, social media

## 1. Introduction

In January 2021, a group of retail investors on the subreddit r/wallstreetbets noticed that a number of hedge funds were betting against GameStop, a struggling brick-and-mortar video game retailer. The retail investors decided to create a campaign to buy GameStop shares in a manner to increase the prices and forcibly remove the short-sellers (i.e. the retail investors). They aim for a phenomenon, known as "short squeezing," which usually occurs when the price of a stock unexpectedly increases, forcing the short-sellers to buy shares at a higher price in order to avoid losses. The subreddit's members were motivated in part by a desire to push back against Wall Street hedge funds that were betting against GameStop, a practice known as short selling. The hedge funds were effectively betting that the price of GameStop's stock would go down, and they stood to make a profit if that happened. The subreddit's members saw this as an opportunity to drive up the price of the stock and cause financial losses for the hedge funds. They began buying large quantities of GameStop's stock, which caused the price to surge. This event made the gamestop stock a prime example of a shortsqueeze, in which investors rebought a large amount of the stock between January 4th and 28th, driving prices up 2701.62%, up from 4.31 to 120.75 USD per unit. This push from subreddit members led to significant financial losses for the hedge funds that had bet against the company and sparked a wider debate about the influence of social media on financial markets.

The driving sentiment behind this movement was encapsulated in the phrase "I Just Like the Stock," which became a rallying cry for those who believed in the potential of GameStop. This surge

brought academic interest to the importance of sentiment analysis in understanding the dynamics of the stock market and also demonstrated the power of social media in driving financial market behavior.

Previous studies have found that media content affects the stock market in a direct manner. The stock market then affects the companies' revenue directly. Ahmad et al. built a corpus on the biggest US firms, using the data they acquired to study relationships between firm monetary returns and the tone of the media [1]. Their team used autoregressive (VAR) models, a novel approach at that time and showed how negative media expressions also negatively affected the firms at times. Sentiments in social media regarding stocks especially affected the non-professional investors to invest in a positive manner, and affected investment firms and corporate investors much less [2].

One way that media content affects stock market behaviors is through sentiments and emotions. When a stock is heavily discussed on Reddit, it can trigger a FOMO (fear of missing out) effect, prompting more investors to buy, further driving up the price. Hu et al. focus on the more particular notion of short trading instead of general stocks and trading. As mentioned earlier in paper, the threat and lure of short selling for amateur investors will actually cause the stock to increase, as seen with the gamestop stock [3]. Hu et al. find that short selling will cause the stock to increase short term but drop down after the craze has ended [3].

Long et al. mentioned in his particular study the utilization of VADER, a program that is essentially sensitive to the degree of emotions expressed in words [4]. By refining the program to the particular area by removing unnecessary words, researchers created their own lexicon with words relevant to the area. They were then able to connect the valence of a particular word to whether it was effective in affecting the stock market. The analyses suggest that the valence of words has a direct correlation with the rise and fall of the Stock Gamestop, and the study has demonstrated the power of social media and the importance of sentiment analysis in understanding the dynamics of the stock market.

Based on these previous studies, we conduct a new study on the emotion of subreddit comments during the GME short squeeze. Unlike Long et al., who examined the impact that Reddit posed on the stock rising of GME, the focus of this paper is to examine the degree of emotional changes over time using a sentient tool [4]. I collected the comments of Reddit anonymously through a corpus and analyzed them as a whole. This allowed me to collect more data, 874 comments to be exact, and be more accurate, compared with Anand et al., which specified the role of 462 major redditors serving as leaders that contributed largely to the stock [5].

To evaluate the sentiment of comments, previous studies have asked human raters to provide a score for the valence (e.g. how positive/negative it is) of a sentence. Specifically, Long et al. asked 10 people to rate the valence of comments (a rating of either negative or positive) [4]. He then related the valence with the stock's price changing. I recognized that the approach could lead to potential deficits, such as lack of agreement among different raters due to subjective rating, and potential outliers. I have fixed this issue by introducing the LDA model (Latent Dirichlet allocation), that objectively analyzes the degree of emotion to an accurate extent.

## 2. Method

### 2.1. Data

We first identified the most influential reddit thread in r/wallstreetbets during this event, titled "I SAY WHEN WE SELL" ([https://www.reddit.com/r/wallstreetbets/comments/16n3et/i\\_say\\_when\\_we\\_sell/](https://www.reddit.com/r/wallstreetbets/comments/16n3et/i_say_when_we_sell/)). We then extracted all the comments that replied to this thread using the Python Reddit API wrapper PRAW [6]. The time range of these comments was from January 6 2022 to January 27, 2022. In total, this gave us 874 comments.

## 2.2. Pre-processing

Through pre-processing, we removed words and posts that the topic model couldn't analyze effectively. These include emojis, errors in spelling, and random comments that have no correlation to the stock (e.g. hyperlinks referring to a random movie). These steps in total removed 71 posts, and left 803 comments that went through further analyses by the topic model.

## 2.3. Analysis

To get an overview of topics in these comments, we used the text analysis method, topic modeling, to analyze the comment. Topic modeling found the hidden structure of a set of documents. Specifically, we used the gensim package in Python and its wrapper of Mallet to implement a topic model with the Latent Dirichlet allocation (LDA) algorithm. The input to the model consisted of a corpus of 803 Reddit comments, with each comment treated as an individual document. We discovered semantic topics and then compared topic similarity among these topics. To decide the best number of topics, we compared the coherence values of LDA models with topic numbers ranging from 2 to 40, and chose the number that yielded the highest coherence value, which was 35. Below we describe the 35 topics found by the topic model and manually categorize them into clusters based on prior knowledge.

## 3. Results

We got 6 clusters:

- Cluster 1: buy
- Cluster 2: sell
- Cluster 3: analysis
- Cluster 4: complaint
- Cluster 5: time
- Cluster 6: other

In cluster 1, the most salient topic (topic id: 4) features the word “buy”, with a weight of 0.639. Other words that reflect a positive mood include “good”, “watch”. Some words reflect strategies, such as “diamond hand”, which indicates the strategy of not selling when the market fluctuates and holding the stock. Words in this cluster were sent out by the stockholders that wanted to profit off the situation by convincing other stockholders to not sell their shares at current price, driving up the price as an effect. The other group were the observers that saw the situation as a joke and a meme, they commented positive words to make the shortsqueeze even more severe. A large number of these observers also made comments in cluster 6 and cluster 5, with no apparent connection to the stock.

In cluster 2, the most noticeable topic (topic id: 17) is related to the word “sell”, with a weight of 0.569. Words in this cluster reflect a negative mood, such as “lose” and “scare”. Most of these words encourage people to sell the stock. As the comments in this cluster were dated mostly in January and February, where the stock was booming at a constant rate, stockholders that posted comments in cluster 2 attempted to convince others that the stock was bound to crash and to sell their shares. They would then rebuy the sold shares at a lower price and then hold shares for themselves.

In cluster 3, the most salient topic (topic id: 28) is represented by the word “other” with a weight of 0.417. Words in this topic tend to analyze the stock with opinions that are mostly objective. Words such as “option”, “short squeeze” are summaries of the stock with little emotion. A majority of the words are nouns that describe the stock's characteristics, such as “market”, “moon”, “option”, and “price”. Nouns in this cluster are objective analyses of the market with low degrees of personal opinion or emotion.

In cluster 4, the most salient topic (topic id: 7) is represented by the word “fuck” with a weight of 0.39. Words in this topic hold a complaining mood that do not address the stock directly but rather are curses and obscenities reflecting an overall negative attitude.

In cluster 5, the most weighted topic (topic id: 17) is represented by the word “day” with a weight of 0.473. Words in this topic reflect on the timing of the stock, other words such as “week” “month” also relate to the chronological nature of the stock.

In cluster 6, the most salient topic (topic id: 12) is the word “tendie” with a weight of 0.154, words in this topic have little correlation with the subject at hand, and are rather random. Other words such as “ortex” and “twitch” have weight as little as 0.001, providing little insight.

Table 1: The table of all the topics and their words and weights.

| Clusters | Topic id | Representative words  |
|----------|----------|---|
| 1: buy   | 4        | "buy": 0.639<br>"hold": 0.189                                   |
|          | 23       | "good": 0.245<br>"watch": 0.146                                 |
|          | 25       | "diamond_hand": 0.165<br>"understand": 0.094<br>"little": 0.052 |
|          | 6        | "put": 0.157<br>"rich": 0.067<br>"reddit": 0.051                |
|          | 2        | "even": 0.124<br>"say": 0.097<br>"invest": 0.079                |
|          | 1        | "many": 0.112<br>"least": 0.100<br>"gme": 0.055                 |
|          | 33       | "trade": 0.094<br>"stonk": 0.071<br>"join": 0.067               |
|          | 34       | "mean": 0.191<br>"new": 0.122<br>"come": 0.113                  |
| 2: sell  | 17       | "sell": 0.569   |
|          | 11       | "lose": 0.261<br>"line": 0.052<br>"hard": 0.050                 |

Table 1: (continued).

|              |    |   |
|--------------|----|---|
|              | 15 | "short": 0.218<br>"still": 0.194<br>"actually": 0.083 |
|              | 22 | "let": 0.213<br>"never": 0.106                        |
|              | 24 | "see": 0.197<br>"lot": 0.089<br>"year": 0.072         |
| 3: analysis  | 28 | "other": 0.417<br>"send": 0.223<br>"much": 0.049      |
|              | 32 | "stock": 0.391<br>"game": 0.154                       |
|              | 0  | "share": 0.320<br>"sell": 0.055<br>"cover": 0.045     |
|              | 10 | "also": 0.232<br>"contract": 0.188<br>"reduce": 0.174 |
|              | 29 | "right": 0.107<br>"hand": 0.082<br>"happen": 0.080    |
|              | 18 | "need": 0.095<br>"guy": 0.072<br>"retard": 0.071      |
|              | 30 | "go": 0.190<br>"money": 0.142<br>"know": 0.116        |
|              | 16 | "call": 0.196<br>"option": 0.194<br>"price": 0.113    |
| 4: complaint | 7  | "fuck": 0.390<br>"man": 0.116<br>"shit": 0.082        |
|              | 9  | "want": 0.327<br>"bleed": 0.033                       |

Table 1: (continued).

|          |    |   |
|----------|----|---|
|          | 3  | "position": 0.219<br>"close": 0.126<br>"laugh": 0.025 |
|          | 13 | "get": 0.390<br>"people": 0.224<br>"wsb": 0.052       |
|          | 5  | "movie": 0.237<br>"big": 0.120<br>"squeeze": 0.111    |
|          | 19 | "make": 0.173<br>"well": 0.129<br>"way": 0.128        |
|          | 14 | "speech": 0.029<br>"freedom": 0.020                   |
| 5: time  | 27 | "day": 0.473<br>"dip": 0.102<br>"ve": 0.083           |
|          | 8  | "time": 0.425<br>"robinhood": 0.079                   |
|          | 20 | "week": 0.120<br>"try": 0.085<br>"today": 0.057       |
|          | 31 | "market": 0.078<br>"look": 0.066<br>"think": 0.062    |
| 6: other | 12 | "tendie": 0.154<br>"boyfriend": 0.052                 |
|          | 21 | "twitch": 0.001                                       |
|          | 26 | "twitch": 0.001                                       |

#### 4. Discussion

In this paper, I analyzed the intricate relationship between the Gamestop stock and the sentiments of a particular reddit forum. Through the usage of a sentiment analysis model, the Latent Dirichlet allocation (LDA) algorithm, we were able to confirm hypothesis that the sentiments of the comments had a direct correlation and impact on the performance of the stock. Separating the comments into six clusters of different types of opinions and emotions, we found high correlation with the “weight”, or coherence values of the emotion of the language in 4 out of the six clusters. These coherence values

also meant the degree that the sentiment of the language could potentially affect the stock, as in cluster one, with a correlation of encouraging investors to buy the stock, or in cluster 2, where investors are cautioned against buying and advised to sell. Regardless of the opinions, it is undeniable that the sentiment of these comments has a recognizable impact on the decisions of these investors.

What our study possibly could have improved on, was selecting similar cases to the GME stock and comparing these cases alongside each other. Due to the nature of the GME explosion being so special, it is likely that this phenomenon will or has been a one timed outlier. If the GME stock's effect was truly an outlier, then there serves little purpose to study the case without a direct comparison to other similar cases that could allow us to further reflect on the connection between emotions, sentiments in comments and the performance of a stock.

Furthermore, due to this one timed event occurring in the past, it is difficult for us to replicate the experiment with the scale and accuracy 1:1, making it difficult to affirm our conclusions compared with other studies.

## 5. Conclusion

On a more grand scale, the study conducted shows the volatility in the stock market, and how something so miniscule as comments in a reddit forum can affect the stock's performance drastically. Although a short-squeeze explosion of this nature is highly unlikely to occur again in any short time, the situation brought up questions such as transparency within the stock market, market manipulation, and the risks and rewards of volatile investment strategies. Reddit's user base is primarily composed of retail investors, who might have limited financial resources compared to institutional investors. However, when they band together and coordinate their efforts, they can create substantial buying or selling pressure on certain stocks, causing prices to fluctuate.

To conclude, this GME case study serves as a captivating illustration of the potential for online communities to challenge conventional market behaviors and influence stock prices in unprecedented ways. This study has proved how Reddit sentiments can add another layer of volatility and speculation to the market, making it essential for investors to conduct thorough research and analysis before making their decisions.

## References

- [1] Ahmad, K., Han, J., Hutson, E., Kearney, C., & Liu, S. (2016). *Media-expressed negative tone and firm-level stock returns*. *Journal of Corporate Finance*, 37, 152-172.
- [2] Dong, H., & Gil-Bazo, J. (2020). *Sentiment stocks*. *International Review of Financial Analysis*, 72, 101573.
- [3] Hu, D., Jones, C. M., Zhang, V., & Zhang, X. (2021). *The rise of reddit: How social media affects retail investors and short-sellers' roles in price discovery*. Available from SSRN: <https://ssrn.com/abstract=3807655>.
- [4] Long, S., Lucey, B., Xie, Y., & Yarovaya, L. (2022). *"I Just Like the Stock": The Role of Reddit Sentiment in the GameStop Share Rally*. *Financial Review*.
- [5] Anand, A., & Pathak, J. (2022). *The role of Reddit in the GameStop short squeeze*. *Economics Letters*, 211, 110249.
- [6] Boe B. (2023). *PRAW: The Python Reddit API Wrapper*. <https://github.com/praw-dev/praw/> [Online; accessed 2023-06-29].