AMCIS 2022 Proceedings                                                    SIG Social Computing

Aug 10th, 12:00 AM

# WallStreetBets Beyond GameStop, YOLOs, and the Moon: The Unique Traits of Reddit's Finance Communities

Pratik Agrawal
*University of Potsdam*, pratikkumar.vijaykumar.agrawal@uni-potsdam.de

Tolga Buz
*Hasso Plattner Institute*, tolga.buz@hpi.de

Gerard de Melo
*Hasso Plattner Institute*, gdm@demelo.org

Follow this and additional works at: https://aisel.aisnet.org/amcis2022

# WallStreetBets Beyond GameStop, YOLOs, and the Moon: The Unique Traits of Reddit's Finance Communities

*Completed Research*

**Pratik Agrawal**
University of Potsdam, Germany
agrawal1@uni-potsdam.de

**Tolga Buz**
Hasso Plattner Institute /
University of Potsdam, Germany
tolga.buz@hpi.de

**Gerard de Melo**
Hasso Plattner Institute /
University of Potsdam, Germany
gdm@demelo.org

## Abstract

While the effect of established social media on stock markets has been thoroughly investigated, the recent surge in retail investing and the emergence of different finance-related Reddit communities with unique new traits have led to new research questions. In this work, we aim to understand the linguistic and thematic characteristics and differences of the largest financial Reddit communities, r/WallStreetBets, r/stocks, and r/investing. Using different techniques for the analysis of linguistic features and topic modeling, we identify keywords and phrases that are most prominent in each community and determine each community's thematic focus and risk affinity. An analysis of users that post on all of these communities confirm these findings, as they appear to adapt to the respective target audience when posting. The stock returns for each community prove consistent with their respective risk profile. Overall, we conclude that understanding these communities can help investors in making more informed investment decisions.

### Keywords

WallStreetBets, Reddit, finance, stock market, retail investing.

## Introduction

The value of extracting insights from social media for financial decision making is well-established—applications such as high-frequency and algorithmic trading are becoming more complex and powerful and seeing increasing adoption among professional stock market investors, despite being reviewed critically (Bell et al. 2012; Godfrey 2013; Beverungen 2019). Using big data and machine learning, these algorithms may incorporate market data, news reports, and raw data from established social media platforms such as Twitter and Facebook, on which millions of users share their opinions on diverse topics, including events, institutions, and people. However, recent months have brought significant growth in retail investing (i.e., by individual, non-professional investors) and with this, a new movement of finance-related communities on the Reddit platform, potentially stemming from a combination of COVID-19 lockdowns and young adults' increasing interest in investment opportunities, accompanied by the rise of low-cost trading apps for mobile devices. For instance, the user numbers of the stock trading app Robinhood increased from 13 million in 2020 to 22.5 million in 2021 (Curry 2022). Such individuals also often rely on social media platforms to consume finance-related information and decide on investments. Considering this recent surge, the community of WallStreetBets (WSB) has risen to prominence–As shown in Figure 1, its user base has grown rapidly, from 1.6 million in November 2020 to 9 million in February 2021, and to over 11 million in November 2021 (subredditstats.com 2022). Many users have been drawn in by the hype surrounding the GameStop stock (GME) since early 2021 and the prospect of taking part in a new movement (Stewart 2021).

This movement has attracted an influx of users to other finance-related Reddit communities such as r/stocks and r/investing, as well.

These communities provide a new form of social exchange for anonymous individuals, whose contributions are intentionally subjective and do not build on any form of authority associated with institutions, media outlets, or public figures, but rather just on their track record within the community. They nonetheless engage in detailed discussions of fundamental and technical analysis of securities. These properties differentiate Reddit communities from more established, well-analyzed social media such as Twitter or Facebook, which focus on contributions of specific channels or on trending topics. Additionally, Reddit communities moderate their discussions by the mechanisms of up-/downvoting and commenting, as well as the rules set in their community guidelines. In order to contribute to the research in this field, we aim to understand the unique characteristics and behavior of the largest finance communities on Reddit, r/WallStreetBets (WSB), r/stocks, and r/investing, and how they differ both from one another and from traditional communities. Specifically, we employ quantitative methodologies to identify patterns in the content and language of each community's posts and discern their differences. We investigate their characteristics and core topics and assess to what extent they may be considered a valuable source of information for professional investors, businesses, and retail investors.

We probe the following questions and answer them in this work: (1) What are the characteristics and differences of the selected communities regarding their content, topics, and language? (2) How do the returns of the top stocks mentioned in each of these communities compare? (3) What are the most popular themes in each of these communities and do they have a different focus? (4) Does the data of users who actively posted in all three communities provide valuable insights to confirm or contradict our findings? (5) Has the GME hype affected or even changed these communities?

We (1) present the methodology we have applied to extract and analyze stock market related insights from Reddit communities, (2) provide a detailed comparison of stock returns for the top stocks mentioned in the selected communities, (3) assess the thematic focus of each community using topic modeling and linguistic feature analysis, (4) present the change in user behavior when they post on WSB vs. other finance subreddits, and (5) train and analyze a machine learning model that classifies submissions into the respective subreddits for further review of our quantitative results.
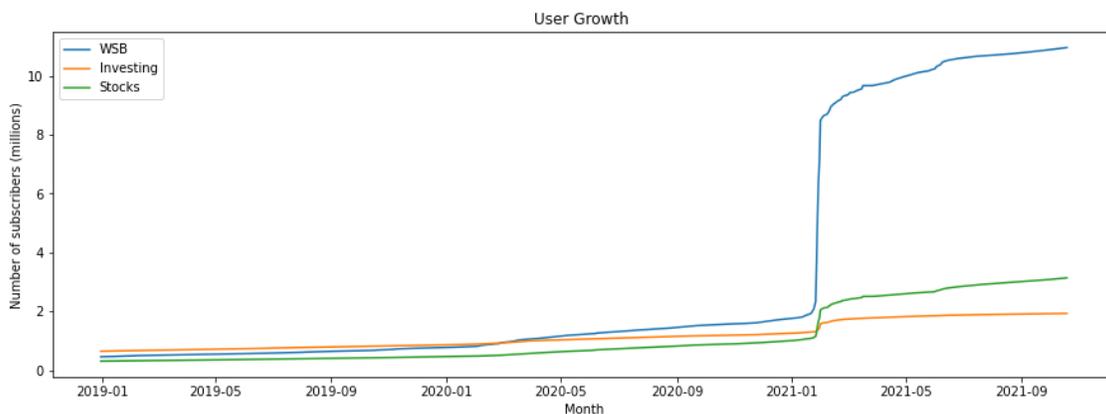


**Figure 1: User Growth of WSB, r/stocks, r/investing**

## *Background*

Reddit is an anonymous, community driven social media website with more than 50 million daily active users organized across more than 100,000 active communities (last updated January 2021; redditinc.com 2022). In Reddit, posts are organized by subject into user-created communities called "subreddits", which each cover a specific topic such as politics, science, video games, music, sports, and many more (Anderson 2015). There are multiple subreddits dedicated to investment topics such as r/stocks, r/investing, r/WallStreetBets, r/pennystocks, and more recently communities discussing cryptocurrencies (Hartwig 2021). The subreddit r/WallStreetBets has come to prominence in the hype around the GameStop stock (GME), in which retail traders appear to have orchestrated investments with the goal of causing what is

known as a short squeeze. The latter forced hedge funds that had been shorting GME to close their positions at a significant loss (Herrman 2021). In contrast, the r/stocks subreddit defines itself as a place for almost any post related to stocks, with a tagline stating that it is a community of "Investing and trading for all". At the time of writing, r/stocks has 3.7 million users/subscribers, having grown from close to 1 million back in December 2020 (r/stocks - wiki 2022). The r/investing community currently has 2 million subscribers, having grown from around 1.25 million in December 2020. This community has strict rules governing what sorts of posts are deemed acceptable. It defines itself as a community primarily focused on capital markets. The wiki section of the community clearly mentions how one is to "conduct herself like an adult while posting and not indulge in personal attacks" (r/investing - wiki 2022). Our study focuses on these three subreddits due to them being the largest finance-related communities in terms of the number of users.

### *Related Work*

The influence of social media on stock markets in general have been studied for many years already, with research suggesting that social media sentiment influences stock market volatility (Piñeiro-Chousa 2017; Jiao 2020), and that social media sentiment can be utilized for stock market prediction (Bollen 2011; Nguyen 2015; Awan 2021). However, these studies primarily consider Twitter data as well as news and blogs, rather than topic-oriented communities such as Reddit. Meanwhile, research on retail investors has shown that they tend to focus on similar types of stocks (Kumar 2006) and possess predictive power to make investments on information before it is incorporated into stock prices (Boehmer 2021), which leads to the question of whether this type of discourse can be found in Reddit communities.

While there have been multiple recent studies on WSB, they are mostly qualitative in nature: Boylston et al. (2021) conducted semi-structured interviews with 13 WSB users to understand how WSB developed one of the largest and most loyal user bases on Reddit. Chohan (2021) analyzed reasons for the rise in popularity of WSB and argues about the importance of nomenclature in subreddit groups to reinforce a common culture. Andersson (2021) qualitatively analyzed top comments of moderator-hosted threads through critical discourse analysis and found an ethos of pursuing economic profit, punishing the hedge funds, and standing up for the weak. Buz et al. (2021) quantitatively assessed how the GME hype may have affected the quality of stock recommendations in WSB. However, to date, no prior work has quantitatively assessed in detail how the language and topics of WSB differ from those of other finance communities and how it evolved between the pre vs. post GME hype time periods. There have been quantitative studies focusing exclusively on WSB discussions about GME and potentially related price changes in the stock market (Anand et al. 2021; Reijmer 2021; Olofsson 2021; Wang et al. 2021). Bradley et al. (2021) examined consequences of due diligence (DD) submissions on WSB on the stock market.

To the best of our knowledge, no prior work has yielded comparative analyses with other finance subreddits nor temporal analyses of the evolution of WSB's language and topics, which we contribute in this work. Additionally, we investigate the effects of the GME hype on the language of WSB and other finance subreddits, and review how top stock tickers in WSB have performed compared to top tickers mentioned in r/stocks and r/investing over the last three years.

## Methodology

The data of WSB, r/stocks, and r/investing was collected by means of the Pushshift API (Pushshift API 2022). This interface provides enhanced functionality, search capabilities, and allows for gathering the necessary historical data from pre-selected time frames. We obtained three years of submission data from the three considered subreddits for the time period spanning from January 1, 2019 to November 26, 2021. This data comprises a total of 861,752 posts from WSB, 154,506 from r/stocks, and 127,769 from r/investing. Within this data, WSB has 396,127 unique users, while r/stocks and r/investing have 84,609 and 73,951 unique users, respectively (where users here are defined as authors of at least one submission). By combining these datasets, we identify 5,637 common users that have posted in all three subreddits at least once, authoring 64,631 posts in WSB, 25,451 in r/stocks, and 21,520 in r/investing in total.

For topic modeling, WSB submissions were filtered based on a few selected categories. Within a subreddit, submissions may be tagged with keywords known as "submission flairs" to assign them categories or labels. Only submissions with the flair tags 'Discussion', 'DD', 'YOLO', 'News', 'Options', 'Stocks', 'Technical Analysis', 'Fundamentals', 'Technicals', 'Daily Discussion', 'Chart', or 'Futures' and having a number of

comments greater than one were considered for further analysis, as these tags designate posts that are meant to offer insights on specific markets or stocks, while other categories such as 'Meme', 'Gain', 'Loss' tend to serve an entertainment purpose or are published post-hoc after investments have been made and time has passed. This filtering yields 229,180 submissions for WSB. In the case of r/stocks and r/investing, which do not accept reactive posts about gains, losses, or memes, after filtering for posts with more than one comment, 84,394 and 111,475 submissions, respectively, remain for the topic modeling dataset.

To perform quantitative and temporal analyses, various linguistic features such as n-grams, verbs, adjectives, nouns etc. are created using the NLP library SpaCy (SpaCy 2022) for the title and body (also called "selftext") of the collected submissions. A comparative analysis of these linguistic features for all three subreddits can then be performed. Using a list of all companies listed on the U.S. stock markets from NASDAQ (nasdaq.com 2021), stock tickers are extracted from the title and body of the submissions. Lastly, to compare our quantitative results and validate our hypotheses, we apply machine learning algorithms to classify the submissions into subreddits and extract the most important features for each subreddit.

## Analysis Results

### *Choice of Words and Emojis*

We performed an analysis of submissions posted on the selected subreddits in order to understand their respective language usage patterns. Our primary motivation for this analysis is to identify unique linguistic characteristics and themes of each community and compare these characteristics and themes. Notable differences become apparent when language characteristics such as usage of words, adjectives, emojis, verbs etc. are analyzed. To find each subreddit's characteristic linguistic features (words, emojis etc.), we extracted features whose relative occurrence frequencies in one subreddit are greater than the sum of the respective frequencies in the other subreddits. For example, if the word 'moon' is used more often in WSB than in r/stocks and r/investing combined, the word 'moon' is deemed a part of WSB's characteristic linguistic features.

Tables 3 and 4 provide insights into each subreddit's characteristic words and verbs. For WSB, we find emojis and profane terms as well as WSB-specific slang (e.g., "moon", "yolo", "ape", etc.) dominating. The WSB community also focuses on topics such as options trading, meme stocks, and short squeezes. This suggests that they seek quick ways to make money, while adopting a very casual, vulgar attitude in their interactions. At the same time, however, WSB seems to have a strong sense of community, mentioning "unite" and "belong" often. In r/stocks, in contrast, we observe words relating to individual stock discussions, fundamental analysis, and terminology specific to the finance domain (e.g., quarterly report, etc.)—although the community also appears to react to what have become known as "meme stocks", which often originate from WSB discussions. The r/stocks community appears to engage in broader discussions related to specific stocks and attempts to identify undervalued securities. r/investing rather invokes words relating to index funds, assets, housing, retirement, and general advice, which point towards a broader, more formal discussion of general, more conservative investment options. We also encounter words such as "risk" being among r/investing's top words, indicating a more risk-aware mindset. The r/investing community thus focuses on a more varied set of assets.

| Subreddit | Characteristic Words |
|---|---|
| **WSB** | minimum, squeeze, ban, moon, 🦍, ape, fellow, gamestop, pltr, diamond, weekly, 🚀, fuck, amc, gme, tendie, boy, bb, 🙌, nok, 💎, fucking, stonk, rocket, retard, yolo, shit |
| **r/stocks** | discuss, quarter, cap, stock, idea, news, apple, trend, past, global, meme, fundamental, report, discussion |
| **r/investing** | fund, asset, advice, house, invest, ask, information, financial, pay, point, risk, need, money, answer, debt, question, dividend |

**Table 3: Characteristic Words of WSB, r/stocks, and r/investing**

| Subreddit | Characteristic Verbs |
|---|---|
| **WSB** | 'going', 'shit', 'deserve', 'laugh', 'scare', 'wear', 'donate', 'unite', 'rip', 'bless', 'squeeze', 'cry', 'fuck', 'pray', 'fucking', 'bleed', 'degenerate', 'drink', 'manipulate', 'print', 'belong' |
| **r/stocks** | 'revise', 'upgrade', 'earning', 'plug', 'stock', 'chat', 'discuss', 'thought', 'rate', 'eps', 'taper', 'undervalue', 'pricing', 'reiterate' |
| **r/investing** | 'invest', 'yield', 'employ', 'consult', 'acorn', 'risk', 'rent', 'earn', 'warrant', 'mind', 'retire', 'time', 'allocate', 'withdraw', 'investing', 'saving', 'emerge', 'calculate', 'contribute' |

**Table 4: Characteristic Verbs of WSB, r/stocks, and r/investing**

Emojis such as 🚀, 💎, 🙌, 🦍 are very prominent on WSB, while usage of the same emojis remains insignificant in the other two communities (Table 5). Notable examples include the rocket emoji, used in reference to possible gains, which leads the statistic, having become highly popular in 2020. It is followed by the diamond and hands emojis (often used in combination to indicate "diamond hands", i.e., someone who holds a security for an extended period of time despite market turbulences, leading to high profits) and the gorilla emoji (often invoked on WSB to casually refer to community members as "apes").

| Emoji | WSB | r/stocks | r/investing | | Emoji | WSB | r/stocks | r/investing |
|---|---|---|---|---|---|---|---|---|
| 🚀 | 48.64 | 2.06 | 0.96 | | 🦍 | 2.88 | 0.03 | 0.01 |
| 💎 | 8.40 | 0.17 | 0.11 | | 🔥 | 0.77 | 0.17 | 0.16 |
| 🙌 | 3.04 | 0.05 | 0.02 | | 👀 | 0.23 | 0.06 | 0.02 |

**Table 5: Number of times emojis appeared per 100 posts on average**

## *Popular and Unique Topics*

BERTopic (Grootendorst 2020) with its default hyperparameters was utilized for deep learning-driven topic modeling. BERTopic provides interpretable topics with salient words as topic descriptions. To identify unique topics, the intersection of these salient words for each topic with the words of every topic of other subreddits is performed. If more than five words are the same, the topic is deemed a common topic. The remaining topics are considered unique.

From the top 30 topics of each subreddit, we have identified the following topics as unique: "diamond hands", "squeeze", "mooning" for WSB; "S&P 500", "portfolio", "index funds" for r/stocks; "dividends", "401k", "retirement" for r/investing. An analysis of these topics supports the conclusions from the previous section: The topic modeling results from WSB's submissions reveal unique phrases (e.g., going to the moon), which are all part of WSB's idiosyncratic slang and hint at a high-risk, high-reward approach to investing. The topics given above are referred to based on the top word from the set of words identified by the topic modeling algorithm for each topic. For example, the dividends topic encompasses the words "dividends", "dividend", "special", "payout", "paying". For r/stocks, the unique topics refer to broader investment decisions, e.g., index funds. In r/investing, we obtain unique topics that indicate discussions of long-term investments with reliable payouts. These results confirm that the selected Reddit communities each have a different thematic focus, which we have observed in their language.

## *Most Discussed Stocks*

Table 6 lists the most frequently discussed stock tickers across the three communities ordered by their cumulative count. As expected, GME and AMC dominate WSB, due to their popularity as *meme stocks*. We find that a reliable method to detect meme stocks is to simply look for stocks that are significantly more often discussed on WSB than in the other communities. The top-ranked tickers in r/stocks and r/investing appear to belong to more conventional, large, and established companies such as AMD, MSFT (Microsoft), AAPL (Apple). However, all three communities have in common that—when particular stocks are discussed—these tend to be growth stocks, often from a technology context.

| | Ticker | WSB | r/stocks | r/investing | | | Ticker | WSB | r/stocks | r/investing |
|---|---|---|---|---|---|---|---|---|---|---|
| *1* | GME | **67.2** | 13.77 | 6.7 | | *13* | SPCE | **2.9** | 2.0 | 0.8 |
| *2* | AMC | **33.8** | 8.4 | 2.7 | | *11* | AMZN | 1.4 | **2.9** | 2.3 |
| *3* | TSLA | **8.1** | 7.4 | 5.4 | | *12* | BABA | 1.3 | **3.2** | 1.6 |
| *4* | BB | **13** | 2.4 | 1.0 | | *14* | COIN | **2.9** | 0.1 | 2.7 |
| *5* | PLTR | **7.3** | 3.9 | 1.2 | | *15* | UBER | 1.0 | 2.0 | **2.2** |
| *6* | NIO | 3.2 | **6.9** | 2.2 | | *16* | SNDL | **3.1** | 1.6 | 0.4 |
| *7* | NOK | **8.8** | 2.0 | 0.7 | | *17* | SPOT | 1.2 | 1.6 | **2.2** |
| *8* | AMD | 2.8 | **5.1** | 3.0 | | *18* | FB | 1.1 | **2.0** | 1.8 |
| *9* | AAPL | 2.1 | **5.0** | 3.3 | | *19* | FORD | 0.9 | **2.1** | 1.7 |
| *10* | MSFT | 1.8 | **4.2** | 2.6 | | *20* | PLUG | 1.0 | **2.4** | 1.1 |

**Table 6: Top Tickers of WSB, r/stocks, r/investing (Ticker occurrence per 1,000 posts)**

| Most mentions on | Ticker | Last 3 Months | Last 6 Months | Last 1 Year | Last 3 Years |
|---|---|---|---|---|---|
| **WSB** | **GME** | -38.52 | -21.46 | +144.19 | +987.05 |
| | **AMC** | -53.53 | -39.84 | +239.22 | +34.81 |
| | **BB** | -32.93 | -29.94 | -41.97 | -16.90 |
| | **NOK** | -0.53 | -6.06 | +37.93 | -10.60 |
| | **PLTR** | -40.60 | -39.73 | -59.35 | +21.09 |
| **r/stocks or r/investing** | **AMD** | -14.96 | +10.48 | +28.74 | +431.12 |
| | **AAPL** | +14.75 | +17.63 | +27.06 | +303.99 |
| | **MSFT** | -10.20 | +4.61 | +22.86 | +179.41 |
| | **AMZN** | -13.03 | -6.92 | -6.02 | +76.25 |
| | **BABA** | -27.17 | -36.22 | -54.36 | -26.42 |

**Table 7: Relative Stock Price Change (%) Over Different Time Periods
(at time of writing; data from Google and Yahoo! Finance)**

In Table 7, we present the relative price changes over four different time windows for the stock tickers mentioned significantly more on WSB as well as tickers mentioned significantly more on r/stocks and r/investing. All three communities have in common a focus on stocks from the technology sector with high growth potential but somewhat higher risk. Due to the general market circumstances in recent months, many of these companies' stocks have lost in value (except for AAPL, i.e., Apple)—this is due to a correction of technology stocks occurring in January and February 2022, followed by geopolitical instability in Ukraine. With respect to the returns over the last six months, the trend is clearer: the stocks most discussed by WSB have lost in value, while the choices of r/stocks and r/investing include some profitable positions among other low performers from the technology sector. For the last three years, we can observe that AMD, AAPL, and MSFT have given an average of 300% returns, while WSB's success is mostly due to the meme stocks GME and AMC. The larger, more established technology companies such as Apple and Microsoft have been discussed more prominently in the rather conservative communities of r/stocks and r/investing and proven to be safer bets within recently falling markets.

Therefore, we conclude: (1) WSB tends to favor riskier growth stocks that might perform well in bullish markets but poorly in bearish markets. (2) The more conservative communities also favor comparably risky technology companies, but have a stronger focus on larger, more established ones, thus faring better during the bearish market. (3) In general, neither of these communities adopts a particularly broad view of the markets and thus an investor may wish to avoid considering any community as their single source of information.

### *Patterns of all three Subreddits' Common Users*

For future analysis, one may consider common users that have posted in all three communities to investigate whether they confirm the linguistic and thematic characteristics observed above. Table 8 shows that the same users indeed change their language to adapt to the particular subreddit they post in: While choosing vulgar language and WSB terminology when posting in WSB, the same users switch to more formal language to suit the discussions when they post in r/stocks and engage in more nuanced discussions and use comparably conservative language when posting in investing. In WSB, these users use emojis excessively, while almost completely omitting them from their vocabulary in the other communities. These common users also mention different stocks in their posts depending on where they post: Table 8 shows that these users focus on TSLA, NIO, and AMD in r/stocks, while they discuss GME, AMC, and PLTR more in WSB, following the general differences between the communities. These users confirm the findings from above, as they employ a different type of language and address different topics when posting in each of the communities.

| Number of times emojis appeared per 100 posts on average | | | | | Top Tickers of WSB, r/stocks, r/investing (Avg. number of tickers per 1000 posts) | | | |
|---|---|---|---|---|---|---|---|---|
| **Emoji** | **WSB** | **r/stocks** | **r/investing** | | **Ticker** | **WSB** | **r/stocks** | **r/investing** |
| 🚀 | 33.55 | 1.42 | 1.05 | | **GME** | **19.4** | 11.5 | 7.8 |
| 💎 | 5.29 | 0.27 | 0.11 | | **AAPL** | 2.0 | **5.3** | 3.9 |
| 🙌 | 1.77 | 0.09 | 0.06 | | **AMC** | **7.5** | 5.5 | 3.0 |
| 🦍 | 1.54 | 0.03 | 0.02 | | **MSFT** | 1.7 | **5.0** | 4.0 |
| 🔥 | 0.40 | 0.20 | 0.15 | | **PLTR** | **5.1** | 4.8 | 2.7 |
| 👀 | 0.16 | 0.09 | 0.05 | | **AMD** | 2.3 | **5.8** | 3.9 |

**Table 8: Common Users - Emojis & Tickers**

### *Most Discriminative Features based on an ML Submission Classifier*

We have trained three machine learning classifiers (Logistic Regression, Linear SVM, Multinomial Naive Bayes) to predict from which subreddit a post originates based on the submission's title and body text, represented using TF-IDF vectors, which are composed of two terms: the first term is the normalized Term Frequency (TF, i.e., the number of times a word appears in a document), divided by the total number of words in that document; the second term is the Inverse Document Frequency (IDF), computed as the logarithm of the number of the documents in the corpus divided by the number of documents where the specific term appears. The Logistic Regression classifier, with one-vs.-rest classification, performs best in terms of accuracy, obtaining an accuracy of 60.1%. A random representative sample of a total of 63,885 (WSB: 21544, r/stocks: 23176, r/investing: 19465) submissions was split into a training set of 42,802 (67%) samples and a test set of 21,083 (33%) samples. Table 9 lists the most important features for each subreddit from the logistic regression model. These features are extracted based on the importance coefficient values of the model for each class label (WSB, r/stocks, and r/investing).

| Subreddit | Discriminative Features |
|-----------|-------------------------|
| **WSB** | yolo, retards, retard, tendies, autist, autists, porn, 💎🙌, retarded, fucking, apes, melvin, ape, 🚀🚀, boys, 🌈🐻, 🚀🚀🚀, 🚀🚀🚀🚀, fuck, autism, gang, gme, autistic, clov, papa, degenerates, confirmed, dfv, thanks, memes, yolod, 💎🤲, fucked, hands, blocking, allowing, jpow, stonk, lmao, shit, fuckers, bois, 69 |
| **r/stocks** | dgly, stock, shll, dkng, term, airlines, watchlist, etfs, enph, workhorse, utility, pink, gnus, plug, sqqq, mgm, codx, metaverse, alpp, dividend, recap, taper, mrna, etsy, bp, stoploss, salesforce, mojo, logistics, mro, fiasco, shopify, sq, question, synopsis, scyx, ties, thoughts, dodge, sava, poised, tsx, depot |
| **r/investing** | spitballin, bonds, estate, invest, ira, bitcoin, business, 401k, investment, etfs, loans, asset, retirement, assets, acorns, crowdfunding, bond, earn, questions, vanguard, startups, private, housing, gdp, equity, property, investments, buffett, intelligent, basic, compound, resident, investor, fund, economy, moronic, dca, yield, crypto, reinvest |

**Table 9: Discriminative Features for Subreddit Prediction using Logistic Regression**

The results from the classifier in Table 9 corroborate our hypotheses and results from the previous sections. We observe WSB's vulgar language, jargon, and an increased usage of emojis among the salient features for WSB. For r/stocks, we notice terms relating to industries such as airlines and logistics, specific stocks, and ETFs. Among r/investing's important features, we observe terms such as "401k", "bonds", "fund", "property", and "retirement", which points to the classifier correlating r/investing submissions with conservative investment assets (but also cryptocurrencies, which seem to be a higher-risk outlier). The top features based on the feature importance coefficients from our classifier models confirm our findings that WSB's characteristics are quite distinctive, while r/stocks and r/investing have a more conservative setting.


### *Effect of GME Hype on r/stocks and r/investing*

Our analysis shows that the language on r/stocks and r/investing changed to some extent after the GME hype. In r/stocks, n-grams such as 'buy gme', 'amc stock', '💎💎', and in r/investing the n-grams 'meme stocks', 'buy gme' were some of the top post-hype bi-grams. Additionally, some users from other communities took interest in WSB after the GME hype and became active in the community: Out of 24,040 r/stocks users who were not active in WSB in 2020, 2,584 (10.74%) became active in 2021. Similarly, 8.20% (1,808) of 22,028 r/investing (previously non-WSB) users first started posting on WSB in 2021. This tells us that events in WSB have profoundly influenced other communities and some users from those may have started observing WSB more closely. Nonetheless, we hypothesize that the very large number of new WSB members in 2021 may have also been due to users newly joining Reddit after hearing about WSB in mainstream media – making the GME hype a major catalyst for the recent emergence of WSB and similar subreddits. This could explain why r/stocks and r/investing have grown during the GME hype as well.

## Discussion

Our analysis provides evidence of fundamental differences between WSB and other finance communities such as r/stocks and r/investing. An analysis of the most frequently mentioned words and topics suggests that (1) members of WSB engage in non-traditional high-risk investment discussions utilizing vulgar language but showing signs of a strong sense of togetherness, while (2) r/stocks' members tend to follow more conventional strategies and are interested in evaluating stock performances (utilizing technical and fundamental analyses) and quarterly earnings reports, and (3) r/investing focuses on low-risk long-term investment assets from diverse areas. This observation is further strengthened by the stock returns analysis. While all are interested in growth stocks with a technology context, the more conservative r/stocks and r/investing communities rather favor established companies such as AAPL, AMD, and MSFT, which have lost less value in the bear market of early 2022. We also find that a simple method of identifying meme stocks is to seek stocks that are mentioned in WSB substantially more frequently, in relative terms,

compared to r/stocks and r/investing. Users who have posted in all three communities adopted the linguistic style of the respective community, e.g., using emojis such as 💎, 🙌, and 🦍 almost exclusively when posting in WSB. From a user behavior perspective, we conclude that these users have understood the differences between the communities and have learned to adapt to each community's norms. Our detailed analysis and the examination of our trained machine learning classifier confirms WSB's characteristics that differentiate it from the other subreddits. Our conclusions agree with previous qualitative findings (Boylston et al. 2021), that humor plays a vital role in promoting in-group cohesion within WSB. WSB's GameStop hype has been a major catalyst of growth for finance-related subreddits, especially WSB, but it does not seem to have changed the inherent nature of communities like r/stocks and r/investing. Although we find that the meme stocks like GME and AMC are also frequently discussed on r/stocks and r/investing, we do not see a pattern of their users shifting to higher-risk investments.

Retail investors as well as financial analysts can utilize these subreddits to understand which stocks are being discussed currently, track the sentiment towards them, and identify meme stocks. Particularly, these communities can aid in understanding retail traders' reactions to recent events and news. However, while extracting information from these sources, one should be aware of their different risk profiles. Reviewing the consensus between all three communities will help to assess the risk associated with a discussed stock – if mostly discussed on WSB, it will likely be a high-risk meme stock, while one mostly discussed on r/investing might be safer but low-yield. Stocks discussed across all communities could be the most promising, being safe enough for the more conservative traders while offering potential gains that intrigue higher-risk traders.

## Conclusion

We have shown that WSB is unique in terms of its language and terminology, which are an essential part of the community's identity and a clear difference to r/stocks and r/investing. Still, none of the communities should be used as a single source of information due to their thematic focus. As increasing numbers of retail traders consider new sources of information for their decision making, the role these communities play in stock market dynamics is bigger than ever. From a business or an investor's perspective, considering the unique traits of different finance-related communities can be valuable to identify important topics and signals among retail traders and how they are discussing them. Furthermore, our analysis shows that the different subreddits provide a nuanced picture of what moves retail traders with different risk affinity. Our insights about the characteristics of these communities can help researchers and investors assess their potential and value. For future work, we see multiple possibilities, e.g., analyzing usage metadata and comment activity to understand user behavior on a large scale, applying further, more advanced machine learning techniques for distinguishing important features, qualitative evaluation to determine how good the training and expertise within the community is, or an analysis of the dynamics of how meme stocks emerge.

## REFERENCES

Anand, A. and Pathak, J. 2021. "The Role of Reddit in the GameStop Short Squeeze" *IIM Bangalore Research Paper No. 644* (Available at SSRN: https://ssrn.com/abstract=3873099)

Anderson K. 2015. "Ask me anything: what is Reddit?" *Library Hi Tech News* 32 (5), 8–11

Andersson E. 2021. "The Mobilisation of r/wallstreetbets : A critical discourse analysis of GameStop short squeeze comments." (Available at http://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-54971)

Awan, M. J., Rahim, M. S. M., Nobanee, H., Munawar, A., Yasin, A., & Zain, A. M. 2021. "Social media and stock market prediction: A big data approach." *Computers, Materials & Continua, 67(2), 2569-2583*.

Bell, D. and Gana, L. 2012. "Algorithmic Trading Systems: A Multifaceted View of Adoption" *2012 45th Hawaii International Conference on System Sciences*

Beverungen, A. 2019. "Algorithmic Trading, Artificial Intelligence and the Politics of Cognition." In: Andreas Sudmann (Hg.): The democratization of artificial intelligence. Net politics in the era of learning algorithms. Bielefeld: transcript 2019, S. 77–93.

Boehmer, E., Jones, C. M., Zhang, X., and Zhang, X. 2021. "Tracking retail investor activity." *The Journal of Finance 76.5: 2249-2305* (Available at https://onlinelibrary.wiley.com/doi/10.1111/jofi.13033).

Bollen, J., and Huina M. 2011. "Twitter mood as a stock market predictor." *Computer 44.10: 91-94.*

Boylston, C., Palacios, B., Tassev, P. and Bruckman, A. 2021. "WallStreetBets: Positions or Ban." (Available at ArXiv abs/2101.12110)

Bradley D., Hanousek Jr. J., Jame R., and Xiaoe Z. 2021. "Place your bets? The market consequences of investment research on Reddit's Wallstreetbets"

Buz, T. and de Melo, G. 2021. "Should You Take Investment Advice From WallStreetBets? A Data-Driven Approach." (Available at arXiv preprint arXiv:2105.02728)

Chohan, U. 2021. "Too Big to Fail, Too Small to Win: The Counter-Hegemony of WallStreetBets" (Available at SSRN: https://ssrn.com/abstract=3849770)

Curry, D. 2022. "Stock Trading & Investing App Revenue and Usage Statistics (2022)" (Available at https://www.businessofapps.com/data/stock-trading-app-market)

finance.yahoo.com. 2022. (Available at https://finance.yahoo.com/)

google.com/finance. 2022. (Available at https://www.google.com/finance/)

Godfrey, P. 2013. "Social media and high-speed algorithmic FX trading are a dangerous mix" *Euromoney.com - Foreign Exchange* (Available at https://www.euromoney.com/article/b12kjvh0jp12xc/social-media-and-high-speed-algorithmic-fx-trading-a-dangerous-mix)

Grootendorst, M. 2020. "BERTopic: leveraging BERT and c-TF-IDF to create easily interpretable topics." (Available at https://doi.org/10.5281/zenodo.4381785).

Hartwig, J. 2021. "Reddit's Top Investing and Trading Communities" (Available at https://www.investopedia.com/reddit-top-investing-and-trading-communities-5189322)

Herrman, J. 2021. "Everything's a Joke Until It's Not." (Available at www.nytimes.com/2021/01/29/style/gamestop-wallstreetbets-reddit.html)

Jiao, P., Veiga, A., and Walther, A. 2020. "Social media, news media and the stock market." *Journal of Economic Behavior & Organization 176: 63-90.*

Kumar, A., and Lee, C. MC. 2006. "Retail investor sentiment and return comovements." *The Journal of Finance 61.5: 2451-2486* (Available at https://doi.org/10.1111/j.1540-6261.2006.01063.x).

nasdaq.com. 2021. "Stock Screener" (Available at https://www.nasdaq.com/market-activity/stocks/screener)

Nguyen, T. H., and Shirai, K. 2015. "Topic modeling based sentiment analysis on social media for stock market prediction." *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers).*

Olofsson S. 2021. "$GME To The Moon: Mapping Memetic Discourse as Discursive Strategy in Reddit Trading Community r/WallStreetBets during the GameStop Short Squeeze Saga" (Available at http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-196319)

Piñeiro-Chousa, J., Vizcaíno-González, M., and Pérez-Pico, A. M. 2017. "Influence of social media over the stock market." *Psychology & Marketing 34.1: 101-108*

Pushshift API. 2022. "Pushshift Reddit API Documentation" (Available at https://github.com/pushshift/api)

redditinc.com. 2022. "Reddit by the Numbers" (Available at www.redditinc.com)

Reijmer, N. 2021. "Contagion of Narratives: An Example of the WallStreetBets Saga", Radboud University. (Available at https://theses.ubn.ru.nl/handle/123456789/11363)

r/investing - wiki. 2022. "Lose money with friends!" (Available at https://www.reddit.com/r/investing/wiki/index)

r/stocks -wiki. 2022." Stocks - Investing and trading for all." (Available at https://www.reddit.com/r/stocks/wiki/index)

scikit-learn. 2022. "Machine Learning in Python" (Available at https://scikit-learn.org/stable/)

SpaCy. 2022. "Linguistic Features" (Available at https://spacy.io/usage/linguistic-features)

Stewart E. 2021 "The GameStop stock frenzy, explained" (Available at https://www.vox.com/the-goods/22249458/gamestop-stock-wallstreetbets-reddit-citron)

subredditstats. 2022. "r/wallstreetbets stats" (Available at https://subredditstats.com/r/wallstreetbets)

Wang, C. and Luo B. 2021. "Predicting $GME Stock Price Movement Using Sentiment from Reddit r/wallstreetbets" (Available at https://aclanthology.org/2021.finnlp-1.4.pdf)