Highly Regarded Investors? Mining Predictive Value from the Collective Intelligence of Reddit's WallStreetBets

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

Moritz Schneider*
moritz.schneider@guest.hpi.de
Hasso Plattner Institute, University of Potsdam
Germany

Lucie-Aimée Kaffee
lucie-aimee.kaffee@hpi.de
Hasso Plattner Institute, University of Potsdam
Germany

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

ABSTRACT

In 2021, the Reddit community of WallStreetBets (WSB) started making headlines in the mainstream media as a source of risky investment ideas called *YOLOs*, and *meme stocks* such as GameStop. The community has become infamous for its use of vulgar language, memes, and unorthodox investment strategies. While previous research has explored the community in terms of language and social dynamics, this work sheds light on the extraction of predictive value from investment advice shared on WSB to investigate Reddit's collective intelligence. We evaluate more than 1.6 million posts contributed over 4.5 years, extracting thousands of investment recommendations, which are matched and benchmarked with corresponding stock market data. To analyse and assess their value, we investigate multiple machine learning models, evaluating their effectiveness on multiple time windows under different market conditions, and exploring their latent spaces. When predicting the success of a new post, we view it within the context of the WSB community's discussions and the stock market data of the same day. Our best model's predictive performance on stock market data outperforms the S&P 500 index and other traditional investment strategies by a significant margin. We conclude that amateur retail traders posting on WSB can act as an intelligent crowd and constitute a valuable source of investment advice. Our findings yield generalizable insights on how the collective intelligence of online communities can be extracted and utilized.

CCS CONCEPTS

• **Information systems** → **Web mining**; **Social networks**; • **Computing methodologies** → *Machine learning*; • **Applied computing** → *Economics.*

KEYWORDS

Web Mining, Social Computing, Wisdom of Crowds, Collaborative Investing

ACM Reference Format:
Tolga Buz, Moritz Schneider, Lucie-Aimée Kaffee, and Gerard de Melo. 2024. Highly Regarded Investors? Mining Predictive Value from the Collective Intelligence of Reddit's WallStreetBets. In *ACM Web Science Conference (WEBSCI '24), May 21–24, 2024, Stuttgart, Germany*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3614419.3643993

Warning: This paper contains explicit language

1 INTRODUCTION

Reddit's WallStreetBets (WSB) community has become notorious for its unorthodox forms of discussing financial investments, most famously its quest in 2021 to promote GameStop (GME) [3] and other stocks that had widely been considered to have grim prospects by professional investors, but have since become known as *meme stocks* [15]. In this community, one immediately encounters vulgarity and slang ("retarded" / "regarded", "apes 🦍"), entertaining memes and videos, as well as casual discussions of highly risky investments, often with a "YOLO" (*you only live once*) attitude. Nonetheless, as the community focuses on discussing and evaluating stock market investments, the question arises whether WSB could be a collectively intelligent crowd [31, 53] and whether there is any predictive value in all of this seemingly noisy data.

In this paper, we investigate WSB's collective intelligence by evaluating whether promising investment recommendations can be identified among the diverse data encountered in the community's posts. To process and analyse this large corpus of posts, we employ explainable machine learning (ML) algorithms that indicate which part of the data contains the most promising investment signals. In particular, we pose the following research questions:

RQ1: How well does the WSB community's shared investment advice perform against traditional investment strategies?
RQ2: How accurately can we identify valuable posts with ML models to extract profitable investment advice?
RQ3: Could an investment strategy following the recommendations of a WSB community-informed ML model be more profitable than traditional strategies?
RQ4: Which part of WSB's recommendations influences the decision of the ML model most, and which patterns can we find?

*Both authors contributed equally to this research.

WEBSCI '24, May 21–24, 2024, Stuttgart, Germany

ACM ISBN 979-8-4007-0334-8/24/05
https://doi.org/10.1145/3614419.3643993

To this end, we create and examine a large dataset of raw WSB posts contributed over 4.5 years in combination with stock market data. We preprocess these posts, mine investment recommendations from the text, enrich them with relevant data from the stock market, and train a series of ML models to identify posts providing genuinely beneficial advice in terms of real-world stock market movements. For classifying WSB's investment recommendations, we define multiple prediction targets: Firstly, whether the recommendation will yield a general positive investment return. Secondly, whether the recommendation will outperform the broader market (represented by the S&P 500 index in our study). Thirdly, the percentage range in which the investment will gain or lose value. In contrast to typical ML settings, a model that is able to outperform the market, even just at a moderate overall level of accuracy, is highly valuable, as it can enable substantial gains and speaks to the predictive quality of WSB's posts. Studies show that stock-picking-focused investment strategies rarely manage to achieve this consistently, even those of large equity funds [19].

We find that raw WSB posts are fairly unreliable, as their median performance underperforms the broader market while the mean performance outperforms the market – yet, our predictive models can be remarkably successful in identifying promising posts when utilizing the context of the respective day's community discussions and stock market data, allowing us to define an investment strategy that outperforms the broader market both when it is rising as well as receding ("bull" and "bear" markets, respectively). We inspect the ML models in detail to identify what kinds of traits are most indicative of posts that exhibit predictive value and therefore contribute to WSB's collective intelligence.

Our main contributions in this work are our findings regarding the quality and predictive value of extracted investment recommendations from WSB and their performance, as well as the detailed analysis of relevant features. Although Reddit has been studied to some extent in prior work (as detailed in the following section), its potential for financial advice has remained underexplored in comparison with X / Twitter.[1] Through our work, we aim to advance the understanding of the collective intelligence exhibited by this platform, which is fundamentally different from Twitter. We open-source our code repository for reproducibility.[2]

2 BACKGROUND & RELATED RESEARCH

2.1 Reddit and WallStreetBets

Reddit is an online platform with over 70 million active participants that aspires to be "the front page of the internet". The site is organized across more than 100,000 active communities called "subreddits" (recent data accessible via www.redditinc.com). These generally focus on a specific topic, which can be as broad as world news or as specific as the latest Apple Watch model. When engaging in these communities, one can contribute content by submitting text posts, images, videos, or links. Each submission can be voted up or down once per user, resulting in the post's total score and visibility in the main feed. Submissions can be discussed via comments, which can be up- and down-voted and replied to as well. The moderators of a given subreddit can set up community guidelines for all of their members to follow, e.g., which topics are disallowed.

WallStreetBets (WSB) is by far the largest subreddit about investing (more than 14 million subscribers at the time of writing) and describes itself as "a community for making money and being amused while doing it. Or, realistically, a place to come and upvote memes when your portfolio is down" (accessible via www.reddit.com/r/wallstreetbets). The community shares diverse kinds of posts, ranging from detailed stock analyses, serious discussions, and links to news, posts of recent gains and losses on the stock market, all the way to images and videos posted as memes for mere entertainment purposes. In mid-2023, WSB ranked as the 49th largest subreddit in terms of subscriber count, but 26th for the number of posts per day, and 5th for the number of comments per day, according to the website subredditstats.com[3] – indicating a highly active and engaged community.

The WSB community has been shown to have evolved a unique form of language, combining financial factuality with entertainment [20], often characterized by vulgarity [2]: members refer to each other as "apes", "autists", and "retards" (or "regards" more recently, to avoid being banned by Reddit). Trading profits are called "tendies", investors holding unrealized losses are "bag holders", and ingenious investment ideas are expected to go "to the moon 🚀" (metaphoric for a skyrocketing stock price), just to name a few examples. At the same time, WSB shows strong group cohesion, and a loyal, good-humoured atmosphere [6]. Figure 1 shows a well-received example post that exhibits WSB's characteristic writing style.[4] While the community's moderators allow WSB's characteristically vulgar language and practice it themselves, they are strict with regard to rule violations – spam posts, hate, political opinions, suspected "pump and dump" schemes, most cryptocurrencies, and scams are policed rigorously, which can be seen by the large ratio of removed posts. This makes a significant contribution to content quality and benefits the community, which generally expects high-quality contributions.

In early 2021, WSB caused GameStop and other meme stocks to gain unprecedented attention, despite their unattractiveness from a traditional investing standpoint – which is why institutional investors were engaging in widespread short-selling. Short-selling, a speculative strategy used when stocks are deemed overvalued, involves borrowing shares to sell immediately, expecting to repurchase them at a lower price before a pre-defined time window expires. Multiple hedge funds, blamed by the WSB community for the 2008/2009 market crash, were identified as those shorting these stocks. WSB users posited that mass purchasing and holding of GameStop stocks would force them to close positions by buying back shares at higher prices, driving the price even higher – due to the community's participation, this successfully caused a "short squeeze" [34, 49] and with it extraordinary share price increases. As similar situations were attempted with other stocks, moderators eventually banned "short squeeze" posts.

While all finance-related communities share the goal of discussing profitable investment ideas, WallStreetBets' additional goal of harming hedge funds financially has led to the situation being

[1]While a search for "Twitter stock market" on Google Scholar yields more than 1.7 million results, "Reddit stock market" yields 153,000 results at the time of writing.
[2]The code is accessible via https://github.com/tbuz/Reddit_Investment_Signals

[3]Due to recent changes to Reddit API pricing, this service does not provide up-to-date data anymore.
[4]Accessible via www.reddit.com/13cn3hx

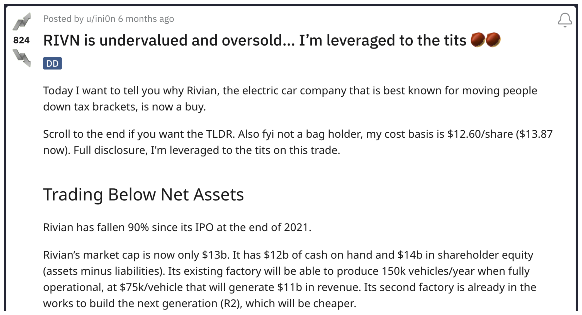

Figure 1: Example Due Diligence Post from WSB (excerpt)

interpreted as a David-vs.-Goliath battle of small retail investors against the institutions they despise. This movement united the WallStreetBets community, made headlines in the mainstream media and caused many millions of people to join the subreddit, growing the 1.8 million subscribers at the beginning of January 2021 to more than 9 million just one month later.

2.2 Collective Intelligence

The theory of collective intelligence discusses the ability of communities to achieve common goals through participation and collaboration [31, 53], which has also been reviewed in the context of online communities in social media [39, 48]. However, there is a lack of research based on large-scale social media analysis seeking to quantify the potential collective intelligence in specific domains such as the stock market. Malone et al. [35] pose four questions in their Collective Intelligence Genome to be asked when investigating whether a community could be collectively intelligent, which we answer for WSB as follows: **(1) What is being done?** – The WSB *creates* stock market analyses and *decides* on their quality using votes and comments. **(2) Who is doing it?** – A large, anonymous *crowd* of Reddit users in the WSB community (with a certain degree of moderation for improved quality). **(3) Why are they doing it?** – There is not much to gain from posting an investment strategy online other than "internet *glory*", i.e., *love* and recognition by the community. However, if the strategy is successful, the stock market rewards the author with money. **(4) How is it being done?** – WSB community members create a *collection* of investment recommendations, which are then rated via *group decisions* about whether to upvote and *individual decisions* about whether to comment on or even follow a proposed investment strategy. In addition, we aggregate daily posts to calculate a *consensus* of recommendations. These answers show that WallStreetBets fulfils the criteria for collective intelligence along the dimensions of the Collective Intelligence Genome.

2.3 Prior Work

The impact of social media has been widely studied in the context of key political or societal events, such as elections [22], the Arab Spring [27], or the #MeToo movement [36]. It has been shown that emotionally charged messages tend to spread more effectively in social media [51]. Information extraction from social media can be valuable for investigating the public sentiment on topics such as politics, companies, brands, and products, and is frequently studied especially in behavioural finance and decision sciences. Other studies have investigated whether cues on social media may yield insights regarding consumer confidence [16], a firm's stock performance [57], or even enable predicting the stock price movement of specific companies [18, 41, 52]. However, the mentioned work focuses on Twitter data.

Reddit, in contrast, does not limit the length of text submissions, which enables contributors to post short opinions or extensive reports – opening up new opportunities for social media mining, but also new challenges. Both platforms have in common an ability to post images, videos, and links to other websites, while Reddit allows for mixed-media posts and extensive response threads. In addition, Reddit's subreddit-based community structure ensures thematic focus, enabling additional investigations. For instance, prior work has used Reddit in domains such as mental health [40], parenting [50], and social norms [17], but very few studies considered the financial subreddits from an economic or quantitative perspective.

Traditional approaches for stock price prediction consider technical indicators from the stock market exclusively. Using machine learning to attempt to predict stock prices has been a very active topic of research [14, 24, 43, 45]. However, existing studies that assess social media as a potential source of information generally consider Twitter as their source of signals and primarily focus on detecting user sentiment [1, 42, 46, 47]. More recent approaches have proposed language models for Natural Language Processing (NLP) tasks related to the financial domain, e.g., two different FinBERT projects [4, 32] focused on tasks such as sentiment analysis and question answering. BloombergGPT [56] is a 50-billion-parameter large language model developed to perform multiple financial NLP tasks. Unfortunately, sentiment analysis tools developed using more general (financial) data do not work well on communities with irregular language such as WSB, which is why some projects extensively modify existing tools manually [37]. BloombergGPT is not available to the public and academics generally lack the data and compute resources to reproduce it. Moreover, these models do not support time series data, in contrast to our approach.

After the 2021 GameStop hype on WallStreetBets, researchers started to investigate this community in particular and derive insights about it: the social dynamics within WSB that led to the hype [34, 49], the battle of retail investors against Wall Street [30], the financial mechanisms behind the irregular price activity [3], the influence of retail traders on prices and volatility [25] – most of this research focuses on socio-economic and general market effects and implications. Further research assessed the implications of the GameStop events for market regulators and brokerages [29, 55] with one study analysing selected posts from an anthropological perspective [38]. Only few studies have assessed WSB from a financial investment perspective: While some focus on a specific question such as the impact of user sentiment on the price dynamics of the GameStop stock [33], and some fail to find predictive value through testing WSB-inspired trading strategies [12], we have identified successful investment strategies through quantitative evaluation of a large number of posts in previous work [10, 11][5], although

[5]We further show that WSB's discussions cover a broad range of stocks, topics, and industries [11].

another study came to the conclusion that the value of WSB diminished after its major growth [7]. While the latter references are most closely related to our work, they do not study the ML-based predictive value of posts. This work sheds novel light on the community using ML and interpretability techniques as well as reviewing different market cycles.

3 DATASET

3.1 Data Acquisition and Filtering

We obtain data from the Reddit WallStreetBets (WSB) subreddit, spanning 1,670,273 submissions posted between January 1, 2018 and July 3, 2022. This data was collected in 2022 via the Pushshift API [5], which regularly stores snapshots of posts and enables access to historic data. A disadvantage of this approach is that certain data points that change over time, e.g., a post's score, may not be up-to-date. Still, this provides a more accessible option than Reddit's official API, which is rate-limited, does not provide historic data, and has introduced a restrictive pricing model in 2023.

These posts were authored by 714,430 unique users, including the community's bots 'AutoModerator' and 'WSBVoteBot'. Of all authors, approximately 261,861 have posted more than once, 15,261 posted more than 10 times, and 157 have posted more than 100 times – this indicates that there is a relatively large core of highly active users as well as an exponential distribution, which is common in social media datasets. From the top ten authors regarding number of posts in our dataset, one is a WSB moderator, three accounts have been suspended, one has been deleted, and the other five seem to be highly active Reddit users who engage in other subreddits as well, with most of them being moderator to at least one community.

The Reddit data includes various types of attributes (more than 100 features in total), including information about the author, subreddit, platform-specific metadata, links, timestamps, and texts. The following features provide valuable information for our analysis (listed as "WSB Metadata" in Table 1): *title* and body text (*selftext*), the post's categorization within the subreddit (*link_flair_text*), *author* information, and selected metadata (listed under WSB Metadata in Table 1). Unfortunately, the features *score*, *num_comments* (number of comments), and *total_awards_received* cannot be utilized to the full extent, as the values are not up-to-date (as explained above). We exclude some of the metadata, e.g., design-related information such as font colour, links for uploaded or shared videos, and image data, which do not provide an additional information gain (e.g., the color of the category tag is always identical per category, and therefore redundant).

A majority of WSB's submissions is labelled with category tags known as "flairs" – a process that is enforced by the community's moderators. Their distribution reveals that the WSB community enjoys serious discussions and DDs (Due Diligences, i.e., analysis and presentation of a specific investment idea) for exchanging ideas to the same extent as memes, YOLOs (i.e., high-risk investments that usually involve placing all funds in a single stock or option), Gain and Loss posts, and so-called "Shitposts" for entertainment.

This categorization helps in dividing posts into two groups: those with a proactive nature suggesting an investment idea (especially DDs and Discussions), and those with a reactive nature responding to something that has already happened or a non-serious contribution (Gain, Loss, Meme, Shitpost, Satire, Donation, Question). We follow the principles of our previous work [10] by filtering the submissions along three rules, as they have been shown to significantly improve the baseline WSB performance: Firstly, we eliminate posts that have been deleted or removed, as these are usually low-quality or duplicate submissions removed by moderators due to violation of the community's rules. Secondly, we remove posts that have an empty text body – e.g., memes are usually only an image or a video with a title, but no further text description, while DDs and Discussions always include a body text. Thirdly, we exclude post categories that are of a "reactive" nature, as explained above. This results in a filtered dataset of 212,042 posts.

3.2 Enrichment with Market Data

For context and evaluation, we extract stock market data for the same time frame for all stocks of the S&P 500 index, which is considered a good indicator for the broader (U.S.) stock market, as it covers approximately 80% of the market capitalization and an estimated USD 15.6 trillion of assets as of December 2021 [28]. Although WSB discussions may encompass arbitrary globally listed stocks, in this paper we focus on S&P 500 companies – our preliminary analysis revealed that this represents a large portion of stocks discussed on WSB, while excluding the volatility and irregularity of smaller meme stocks (larger meme stocks like Tesla remain in scope).

For each stock ticker included in the S&P 500 index, we obtain various features based on stock market data for it by mining Yahoo! Finance. This data contains detailed stock price history at a daily granularity level, including opening and closing prices (*Open*, *Close*), daily *High* and *Low* prices, *Volatility*, *Volume*, as well as *Dividends* and *Stock Splits* (the latter two only if applicable for a day). For each day, we add the relative change of the stock price *since* one week, three days, and one day (*prev_t*), and *after* one day, three days, one week, one month, and three months (*change_t*) – the latter are future data with respect to the data point, which we therefore use individually as target labels during training and testing, but not as an input value in order to avoid data leakage and "cheating". We have decided on these time windows as (1) one day is the minimum granularity we can achieve with the Yahoo! Finance dataset, (2) longer windows than three months only leave a small subset of our data for the analysis, as the window is subtracted both from the beginning and the end, and (3) the other windows in between are aimed to simulate different short and mid-term investment horizons.

We further obtain and extract the recommendations of investment bank analyst reports (in particular, their positive recommendations, e.g., "Buy", "Outperform", "Positive") as supplementary information – we create a column per investment bank (e.g., *Morgan Stanley*) and set a day's value as *true* if the bank issues an investment signal on that day. In addition, we aggregate all investment bank recommendations in one column (*any_buy*). While access to full reports is expensive, the final decision, e.g., "buy", was retrievable via services such as Yahoo! Finance at the time of writing.[6] It has to be noted that investment bank recommendations are

[6]Unfortunately, recent changes in the Yahoo! Finance API have reduced the accessibility of their data.

only considered as an additional indicator once WSB has provided a signal on a day, which simulates a user reading posts on WSB and then visiting Yahoo! Finance for further research. As additional features, we compute the moving averages (MA) of the closing price for 7, 30, and 90 days for each day (*MA07*, *MA30*, *MA90*), and whether a day's stock price has been below the 30-day MA in conjunction with a buy signal of WSB (*BUY_MA30*) or any investment bank recommendation (*any_buy_MA30*). In Table 1, these features are listed in the categories "Investment Banks", "Technical Indicators", and "Stock Price Data". We are able to combine the very different datasets of WSB posts and daily financial data by aggregating the extracted WSB signals per stock on a daily granularity, which is then easily matched with the daily price data gathered per stock of the S&P 500 index from Yahoo! Finance.

The reviewed time frame from January 2018 to July 2022 contains a challenging mix of positive ("bull market") as well as negative developments ("bear market") in the stock market, affected by factors such as the COVID-19 pandemic, rising inflation, and the war in Ukraine. In these circumstances, beating the market consistently remains a challenging task for all investors.

4 METHODOLOGY

Figure 2 gives an overview of the experimental setup of this work, which we explain in further detail in this section.

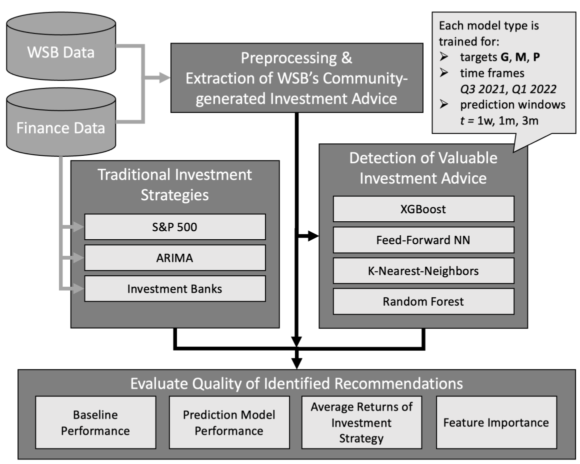

Figure 2: Overview of our experimental setup

4.1 Feature Selection and Feature Engineering

We further process the datasets to maximize the information gained from the semi-structured data (containing structured metadata and financial information as well as unstructured text data). We start by analysing the potential information gain per column and filter out columns from the Reddit dataset that do not add value for our analysis. As the Reddit dataset provides a large amount of platform-specific metadata, we exclude various features that are not valuable for our analysis or provide redundant information, e.g., a post's category ("flair") has a text, richtext, text color, background color, CSS class, and type – all of these are always identical for each respective flair text. Following this approach, we exclude more than 60 features, including various Reddit-internal meta information and features that only occur for a subset of posts.

The features categorized as "WSB Features" in Table 1 are mainly extracted from the post's unstructured text fields (*title*, *selftext*) and encoded in numerical features to be ingested into ML models. For data cleaning and feature engineering, we eliminate punctuation characters in title and body text and extract additional information – most importantly the mentioned stock tickers and their frequency, as well as wording that is related to investment recommendations: For the extraction of the S&P 500 stock mentions, we count the occurrence of tickers with (*$count*) and without (*count*) a prefixed "$" (e.g., $AAPL and AAPL for stocks of Apple, Inc.) as well as full company names. The feature *count_window* considers the average *count* over the last three days. Our initial results revealed that additional rules are necessary to eliminate false positives: single-character tickers and tickers that could be mistaken for common words or abbreviations are only counted with a prefixed "$", e.g., F (Ford Motor Company or an abbreviation for a popular swear word), DD (Due Diligence or DuPont), HAS (Hasbro). We employ a robust method to identify tickers that are mistaken as false positives: the ratio of mentions with a prefixed "$" to the mentions without a prefixed "$" is within a range of 20 – 50% for all other tickers, while it is far below 10%, in some cases even below 1% for the false positive candidates. Applying this method yields a list of "false positive" tickers that are only considered if mentioned with a prefixed "$", including 'ALL', 'ARE', 'CEO', 'DD', 'DOW', 'FAST', 'INFO', 'IP', 'IT', 'LOW', 'MA', 'NOW', 'PSA', 'SEE', 'SO', 'TECH'. For each post, we include a list of all tickers (*tickers*) and the most-mentioned ticker (*top_ticker*) as features.

To detect wording related to an investment recommendation such as buying, holding, or selling, we determine the frequency of specific words (specifically, "buy" and "call(s)", "hold", "sell" and "put(s)") and subtract the frequency of corresponding negations, e.g., "not buy", "don't buy" (*BUY*, *HOLD*, *SELL*). Additionally, we compute the same statistic with the constraint that buy-related words must occur in close (up to 20 characters[7]) proximity of a stock ticker (*BUY_ngrams*) – but not for hold and sell signals, as our evaluation has shown that these do not provide value. If a post contains at least 50% more buy-related words than sell-related ones, we consider it a buy recommendation (*BUY_post*). We add a Boolean indicator for whether the majority of all WSB posts about the specific stock posted on that day recommends a buy or a sell, also with a 50% threshold (*BUY_signal*, *SELL_signal*). For each post, we add the *date* and *weekday* (extracted from *created_utc*), the number of posts per day that have mentioned the same ticker (*posts*), and the total amount of WSB posts per day (*activity*).

Applied to the filtered dataset, this methodology yields 28,007 posts containing at least one ticker mention (of one of the S&P 500-listed tickers), of which 5,785 posts mention a ticker with a prefixed "$", and 11,522 are classified as a buy signal, while 4,902 are classified as a sell signal. As additional text-related features, we add the number of words per post (*text_lengths*) as well as sentiment values (*polarity*, *subjectivity*), which are predicted with the medium-sized, pre-trained spaCy model (*en_core_web_md*) [26]. Furthermore, we

[7]Based on our analysis of the dataset, these words usually occur closely to the ticker, e.g. "buy TSLA" or "AAPL is a buy". We have chosen 20 characters to provide a buffer and allow longer variants, too.

have experimented with word embeddings – however, this did not improve our results, as the models performed worse when the number of input features increased too much. All non-numerical attributes are encoded as categorical features by creating categories or grouping into buckets.

Table 1 provides an overview of the resulting feature set that we include in our model pipeline – for these we define six categories: *WSB Features*, *WSB Metadata*, *Investment Banks*, *Technical Indicators*, *Stock Price Data*, and *Text Information.*

Table 1: Overview of selected and engineered features for ML model training (categorical features are encoded)

Feature Category	Included Features
WSB Features	date, weekday, tickers, top_ticker, BUY_post, count, $count, BUY, HOLD, SELL, BUY_ngrams, posts, count_window, BUY_Signal, SELL_Signal, activity
WSB Metadata	author_flair_text, author_premium, created_utc, domain, is_original, is_reddit_media_domain, is_self, is_video, link_flair_text, num_comments, score, total_awards_received, upvote_ratio
Investment Banks	any_buy, Morgan Stanley, Credit Suisse, Wells Fargo, Citigroup, Barclays, Deutsche Bank, UBS, Raymond James, JP Morgan, B of A Securities, BMO Capital, Keybanc, RBC Capital, Goldman Sachs, Mizuho, Stifel, Piper Sandler, Baird, Jefferies, Oppenheimer
Technical Indicators	MA07, MA30, MA90, BUY_MA30, any_buy_MA30
Stock Price Data	Open, High, Low, Close, Volatility, Volume, Dividends, Stock Splits, close_diff, prev_1w, prev_3d, prev_1d, change_1d, change_3d, change_1w, change_1m, change_3m
Text Information	text_lengths, polarity, subjectivity

4.2 Prediction Tasks

With the aim of creating multiple prediction targets of varying difficulty levels, we define three different target variables that are each reviewed over three time windows t (one week, one month, one quarter):

G_t **– Gains:** The investment's price increases after time window t (binary classification).

M_t **– Outperforming the Market:** The investment's price increases more or decreases less than the S&P 500 index after time window t (binary classification).

P_t **– Price Movement:** The investment's price moves within range x after time window t, with x being the range represented by one of five (20%-) quantiles for the stock price change (multi-class classification).

The time windows provide an analysis of different investment horizons, which may exhibit variation in feature importance. While comparable research in the domain of finance prefers shorter time windows as short as one minute (e.g., [33]), we have chosen these longer alternatives based on multiple considerations:

(1) Our initial evaluation showed that $t <$ one week do not yield insightful results, as the differences in profitability are marginal and only become visible when reviewed over one week and beyond.

(2) We have observed that typical WSB visitors are retail investors, i.e., people working in a different job than finance and investment for longer time horizons (in contrast to day traders).

(3) It usually takes a few hours for a new Reddit post to become visible to the majority of visitors due to differences in reading activity and Reddit's standard filter ("hot"), which only shows posts if a sufficient number of users viewing a post in the "new" filter upvote it before it gets too old.

(4) The shortest t that we can analyse is one day due to the granularity of the data extracted from Yahoo! Finance.

We employ specific dataset splits into training and testing data, by using all data from January 2018 until a point of time d for training and the newer data points of the following quarter (i.e., three months) for testing. We run separate sets of experiments for two such temporal splits, with d being the end of Q2 2021, and Q4 2021, respectively, as each of these quarters has a different combination of market conditions for itself and its following quarter, making Q3 2021 ("bull market") and Q1 2022 ("bear market") our prediction targets. We use this as our only time-dependent feature in order to reduce co-dependencies. This setup allows us to assess how well the different prediction models perform under varying market conditions. We choose five bins to classify for P_t, to simplify the regression task.

4.3 Prediction Models

For our experiments, we investigate multiple ML algorithms: XGBoost, K-Nearest-Neighbors, Random Forests, and Feed-Forward Neural Networks. As it is crucial to understand feature importance for our insights into collective intelligence, we select these algorithms with some degree of interpretability.

EXtreme Gradient Boosting (XGBoost) is an ensemble method based on gradient boosted decision trees [13]. For parameter tuning, we consider a maximum depth of 15, while reducing the undersampling ratio of the training instances to 0.5 to mitigate the risk of overfitting resulting from deeper trees. **K-Nearest-Neighbors (KNN)** is a distance-based classification algorithm that considers the closest historical data points [23], in our case the $K = 3$ closest ones in terms of the Euclidean distance. **Random Forests (RF)** combine the method of Bagging [8] and the calculation of a contamination measure [9]. We use 100 trees per forest with Gini impurity as a criterion to determine the quality of a split. **Feed-Forward Neural Networks (FFNN)** are deep neural networks with three or more layers [44]. Our architecture consists of 4 fully-connected hidden layers of dimensionality 10 with ReLU activation, trained in 1,000 iterations using Adam optimization. In addition, we have evaluated Naive Bayesian and SGD-optimized linear SVM classifiers, which are omitted from further discussion due to poor performance.

4.4 Traditional Investment Strategy Baselines

In order to compare our results to traditional investment strategies (shown in Table 2), we consider the S&P 500 and two alternative baselines: an Autoregressive integrated moving average (**ARIMA**) model trained on historic data to decide daily whether to invest in any S&P 500 stock for time t, and the performance of investment banks' analyst recommendations to buy any stock of the S&P 500 index (**Banks**), all for the same time windows of our test data (Q3

2021 and Q1 2022). ARIMA is a traditional forecasting method for time series data [21] and a common baseline in finance. To produce the best forecasts for predicting future data based on historical data, a dataset with seasonality and a clear trend is required. If the forecast indicates that the index will rise, we buy the S&P 500 for the defined time frames of one week, one month, or three months. The average gain or loss for all buy decisions is then used as a reference value. We also tested Facebook's Prophet [54], which performed worse than ARIMA and was therefore omitted. For the **Banks** baseline, we have included all positive (i.e., buy) signals of the 20 largest investment banks (regarding the number of reports published, e.g., Morgan Stanley) across all stocks of the S&P 500 index and calculated the average price performance after t, in accordance with how the WSB baseline is calculated.

Table 2: WSB baseline performance for test targets gains G_t, outperforming the market M_t, and average price change of raw WSB signals (WSB) and baselines after t. Support indicates the number of investment signals. G_t and M_t show that raw WSB signal accuracy is not high, with up to 70% of signals under-performing the market (in t = 3 months, Q1 2022). WSB, however, which shows the average returns (basis for calculating price movement P_t), performs better than or equal to traditional investment strategies in multiple cases.

	Target / Baseline	t = 1 week	t = 1 month	t = 3 months	Support
Q3 2021	G_t	50%	56%	60%	3,024
	M_t	40%	41%	34%	3,024
	WSB	0.10%	1.65%	5.54%	3,024
	S&P 500	0.04%	1.50%	4.68%	61
	ARIMA	0.78%	2.26%	2.48%	30,500
	Banks	0.13%	1.76%	5.32%	768
Q1 2022	G_t	41%	36%	20%	1,705
	M_t	42%	36%	30%	1,705
	WSB	-0.77%	-3.18%	-12.72%	1,705
	S&P 500	-0.46%	-2.72%	-13.02%	63
	ARIMA	2.90%	2.50%	N/A	31,500
	Banks	0.33%	-0.26%	-4.52%	991

5 EXPERIMENTAL RESULTS

5.1 WSB Versus Traditional Investment Strategies

In a first step, we aim to establish a ground truth of WSB's collective intelligence. For this purpose, we investigate the baseline performance of all WSB investment recommendations and compare it to traditional investment strategies (**RQ1**). Table 2 provides a quantitative evaluation for our previously defined targets in this regard. In addition, we calculate raw WSB performance based on the simplest type of signal: a post has to mention at least one stock ticker (from the S&P 500 index). The general performance of the investment recommendations is acceptable in Q3 2021, as G_t, i.e., the ratio of WSB signals that have yielded profits after t, indicates that 60% show a price increase after three months and at least at 50% for the shorter-term windows. When comparing this to the S&P 500, WSB's signals do not perform as well: only 34% of investment signals outperform the S&P 500 index after three months (M_t).[8] Despite this, WSB's average performance (*WSB* in Table 2) beats the index, most likely due to a smaller set of top-performing stocks. The ARIMA model outperforms other baselines in the short term but is unable to do so at t = three months. The investment banks perform marginally above the raw WSB signals at shorter t, also beating the broader market on average. In Q1 2022, the results look considerably worse across all signal sources due to the different market conditions: only 20% of WSB's signals yield a profit after three months and 30% are able to outperform the S&P 500 (i.e., by losing less value than the index or even being profitable in some cases). Regarding mean performance, WSB performs slightly less worse than the S&P 500. The ARIMA model outperforms the alternatives for shorter t, but does not provide an investment recommendation for t = three months in Q1, and hence its average return is noted as N/A. Investment bank recommendations achieve small short term profits and lose less value on the longer time windows than WSB and the S&P 500. They thus appear to provide a certain level of quality (despite it being insufficient to achieve longer-term profitability).

In conclusion, it is clear that the general market conditions strongly affect the success of all investment strategies. In the bull market of Q3 2021, the WSB baseline performance is most successful in the longest time window t and in general quite close to the performance of investment banks, while ARIMA significantly outperforms alternatives for shorter t. In the bear market of Q1 2022, however, all strategies perform much worse, with most of them losing value. Only ARIMA manages to remain profitable in this case, but again only for shorter t. In our experimental setup, the baseline collective intelligence of WSB is already able to achieve market-level performance and even slightly outperforms it in both markets (Q3 and Q1), competing with the investment strategy of following investment bank recommendations. This is in line with and extends our previous work [10, 11]. One important factor to consider is the number of signals (support) per strategy – while investing with ARIMA in short t appears most profitable (especially in Q1 2022), this may entail much higher transaction costs, reducing the net profits. Therefore, ARIMA can be a suitable strategy for professional high-frequency traders that pay reduced or no transaction fees due to their brokerage accounts, while following a combination of WSB and investment bank recommendations may be more suitable for retail traders who invest less frequently for longer t and wish to outperform the broader market. However, investors who merely follow the broader market by investing into ETFs and do not spend time on research or trying to time the market, perform decently in the bull market, but may be better off when halting their investment activity during a bear market.

5.2 Extracting Valuable Investment Advice

Given the promising but mixed results of the WSB baseline, we next assess how accurately the best recommendations of WSB can be detected with ML classifiers (**RQ2**). The results of our ML models for all targets and time-frames are summarized in Table 3.

[8] It should be noted that in Q3 2021, M_t is more difficult than G_t, and vice versa in Q1 2022, due to the S&P 500's positive and negative development, respectively.

Table 3: Evaluation of all ML models (for gains G_t, outperforming the market M_t, and price movement P_t) for investment recommendation classification and baseline accuracies. Best models achieve >60% accuracy, indicating that our models are correct more often than they fail, with XGBoost significantly improving over the baseline in almost all cases.

Model Configuration			WSB Baseline	XGBoost		KNN		Random Forest		FFNN	
	Target	t	Accuracy	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Q3 2021	G_t	1 week	0.50	**0.52**	0.49	0.50	0.49	0.50	0.34	0.51	0.50
		1 month	0.56	0.49	0.47	0.50	0.50	**0.56**	0.40	0.53	0.54
		3 months	0.60	**0.63**	0.62	0.48	0.48	0.40	0.23	0.60	0.53
	M_t	1 week	0.40	**0.60**	0.57	**0.60**	0.59	0.40	0.23	**0.60**	0.57
		1 month	0.41	**0.62**	0.57	0.52	0.52	0.41	0.24	0.58	0.57
		3 months	0.34	**0.69**	0.61	0.66	0.61	0.66	0.52	0.64	0.62
	P_t	1 week	–	**0.26**	0.25	0.24	0.24	0.21	0.07	0.25	0.23
		1 month	–	**0.26**	0.24	0.22	0.21	0.12	0.03	0.24	0.18
		3 months	–	**0.29**	0.27	0.21	0.20	0.19	0.12	0.23	0.19
Q1 2022	G_t	1 week	0.41	0.57	0.56	0.53	0.53	**0.59**	0.43	0.48	0.47
		1 month	0.36	**0.63**	0.57	**0.63**	0.53	0.36	0.19	0.51	0.51
		3 months	0.20	0.78	0.77	0.73	0.69	**0.80**	0.72	0.63	0.66
	M_t	1 week	0.42	0.58	0.57	**0.60**	0.59	0.42	0.25	**0.60**	0.58
		1 month	0.36	**0.60**	0.53	0.59	0.50	0.41	0.23	0.63	0.59
		3 months	0.30	**0.59**	0.59	0.44	0.44	0.30	0.14	0.55	0.57
	P_t	1 week	–	0.23	0.22	**0.25**	0.21	0.13	0.05	0.22	0.16
		1 month	–	**0.33**	0.20	**0.33**	0.20	0.09	0.01	0.28	0.28
		3 months	–	**0.52**	0.48	0.44	0.42	0.17	0.10	0.49	0.46

To evaluate the different models within the context of the market, we compare their performance to the same baselines. Whenever a model achieves a higher accuracy score than the WSB baseline, this indicates that following the model's recommendations increases the chance of achieving positive investment returns (G_t) compared to following all WSB recommendations. The models trained for target M_t aim at selecting those signals that are likely to outperform the S&P 500 index.

As seen in Table 3, the XGBoost models achieve the best accuracy values, while the Neural Networks (FFNN) achieve slightly more consistent results and are often only closely behind XGBoost. The KNN models exhibit less consistent results and are hence not reliable for investment decisions. Random Forests obtain fairly inconsistent results, stemming mainly from the fact that they often degenerate to predicting just a single class – this is only beneficial to some extent in the bear market of Q1 2022, during which a majority of WSB's signals yielded negative returns. We conclude that the KNN and Random Forest algorithms are incapable of handling the task's dimensionality or complexity and thus yield underfit models with a prediction quality not far from random guessing.

Upon closer examination, the scores of XGBoost are particularly intriguing: On target G_t, the model achieves accuracies of 63% in Q3 2021 and 78% in Q1 2022 for predicting the success of an investment signal after three months. This implies that for roughly two thirds of the evaluated WSB signals, the model can correctly predict whether an investment would yield profits or not. Assuming a sufficiently balanced distribution of positive and negative yields, this would ensure a profitable trading strategy. The results on M_t are even more relevant: for 69% (Q3 2021) and 59% (Q1 2022) of WSB's investment signals, the XGBoost model is able to correctly predict whether it will have outperformed the S&P 500 after three months or not – achieving a recall of 90% in detecting the valuable recommendations, meaning that it is able to detect almost all signals that outperform the S&P 500 three months later. While not as accurate, the FFNN consistently achieves results above 50% as well – this suggests that a strategy following this model could yield profitable investments and even outperform the S&P 500. Assuming the observation from the previous section about the WSB signals being better in identifying top-performing stocks holds true, the investments should be more profitable than a strategy that follows the index.

Remarkably, the prediction results for Q1 2022 are in some cases significantly better than those for Q3 2021, which suggests that the trained models are able to differentiate between good and bad investment signals more effectively when they are in a bear market. Regarding the multi-class target P_t, a more nuanced approach is needed to assess the utility of the models. In this case, reviewing the accuracy score is not as insightful, but the average price change of the (true and false) positives can instead be evaluated, which is shown in the following section.

An investment strategy that is able to consistently beat the broader market possesses substantial potential value, as only a small ratio of even the largest equity funds is able to accomplish this [19]. Our classification models (specifically XGBoost) can be leveraged to identify WSB recommendations that exhibit predictive value with an accuracy above 60% (**RQ2**), enabling the filtering of WSB content for pieces of collective intelligence. While these accuracy scores are not as high as in other applications of ML (which often aim for 80–90% or more), in our context it can be sufficient to provide an advantage over the baseline, especially the S&P 500, to be successful and provide a valuable strategy for retail and professional investors.

Table 4: Hypothetical average returns (rounded, in %) of an investment strategy that follows all predictions of our XGBoost models (for gains G_t, outperforming the market M_t, and price movement P_t) after time t versus baselines (S&P 500, ARIMA, Investment Banks, and raw WSB baseline).

	Model target			Baselines			
t	G_t	M_t	P_t	S&P500	ARIMA	Banks	WSB
Q3 2021							
1w	0.2	0.5	-0.5	0.0	**0.8**	0.1	0.1
1m	0.8	**2.3**	1.9	1.5	**2.3**	1.8	1.6
3m	7.2	**22.69**	14.5	4.8	2.5	5.3	5.2
Q1 2022							
1w	-0.8	-0.7	0.1	-0.9	**2.9**	0.3	-0.8
1m	0.1	-1.7	-1.4	-3.3	**2.5**	-0.3	-3.0
3m	**0.1**	-11.3	-3.2	-13.1	N/A	-4.5	-12.0

5.3 Investment Returns of Community–Informed Model

Naturally, our next question is: Could an investment strategy following the recommendations of a WSB community-informed model be more profitable than traditional strategies (**RQ3**)? To answer this question, we evaluate the average returns of following the predictions of the best-performing (XGBoost) models (Table 4). The results are promising, as at least two out of three models per time window t outperform the S&P 500, suggesting that the models can indeed beat the broader market on average, showing the insightful use of WSB-derived information. The differences between the two time windows confirm that in Q3 2021, outperforming the market (M_t) is a more challenging target than gains (G_t), thus yielding a better predictive model, and vice versa in Q1 2022, due to the performance of the S&P 500 – in a bull market, the ambition is to beat the market (e.g., M_{3m}), while in a bear market, the ambition is to minimize losses (e.g., G_{3m}). In the case of our WSB-based models, both ambitions are met, with promising results during both market cycles compared to most baselines (especially on the model trained for P_{3m} in Q3). While the S&P 500, banks, and WSB baselines are quite similar, ARIMA outperforms all other strategies on short-term investment horizons but under-performs (or does not perform) in the longer term. The XGBoost models each classify approximately 1/5 to 1/2 of all WSB recommendations as promising, providing a total number of recommendations similar to the banks, and significantly fewer than ARIMA's. When applying these models in real-time without knowledge of whether the current market is a bull or a bear one, it may be best to select investment recommendations for which the XGBoost (trained for G_t and M_t) and ARIMA models are in consensus.

5.4 Most Relevant Features for Post Classification

When working with WSB in an abstracted manner, particularly in the realm of financial advice, it is important to understand how the invoked models arrive at this advice. Therefore, our final question relates to the explainability of the models: Are there particular indicators that enable the distinction between profitable advice and poor advice (**RQ4**)? In particular, we seek to understand what kind of information is most valuable for predictive purposes in the machine learning sense. Table 5 shows the distribution of top 10 most important features for each of our different XGBoost model configurations. The feature importance metric is the average weight of each feature within the set of decision trees trained within the model. For simplicity and readability, we have grouped the features into the six categories defined above.

We enrich the WSB-based investment advice features with investment bank recommendations. This enrichment only happens if there is a WSB signal on the same day, i.e., their inclusion is dependent on WSB's activity and it is the conjunction of both recommendation sources that the models are exploiting. We find leveraging this additional data to be beneficial to other models (Table 5). The XGBoost models benefit from the recommendations from investment banks across all configurations, while their importance diminishes for longer time windows t. Stock price data is the second-most important category and relevant for all targets – these features consistently become more relevant when t is larger. Signals extracted from WSB (WSB Features) are relevant for gains G_t and outperforming the market M_t, but with reduced effect in the models trained for the bear market of Q1 2022. Similarly, the technical indicators are relevant for G_t and M_t as well but seem slightly more relevant in Q1 2022. WSB metadata occur rarely among the top features and mostly for the shortest t of one week. Text information does not appear among the most important features of any model, indicating that they are either irrelevant or not accurate. The latter could be the case for the polarity and subjectivity measures due to the irregular language used in the WSB community.

The results in Table 4 show that the investment bank signals on their own are not as good as the returns of the model's predictions. As the manually extracted features are less relevant for P_t, it seems less important whether a mentioned stock ticker has been mentioned 10 or 200 times on a day – as long as a buy signal is detected, the model uses it along with additional stock market features that can be extracted from sources such as Yahoo! Finance to derive its classification results.

6 CONCLUSION

We find that WSB, despite being a community of memes and risky investment advice, can serve as a source of highly valuable signals to train classification models for investment. Our analysis results show that signals derived from WSB's semi-structured data as a baseline are already competitive against the S&P 500. Depending on the market conditions, the WSB community is able to produce investment signals with a certain degree of predictive power. By utilizing these signals together with stock market data to train community-informed ML models, we can better differentiate valuable investment signals and further improve the performance, implying substantial predictive power. These models work remarkably well in both reviewed time windows, Q3 2021 and Q1 2022, which are inherently different regarding the underlying market conditions. Our evaluation of the performance of an investment strategy following the best prediction models shows that it can yield results that outperform the broader market and relevant baselines by a significant margin, both in a bull and bear market, especially for

Table 5: Distribution of top 10 feature importances for the XGBoost models (for gains G_t, outperforming the market M_t, and price movement P_t), showing that prediction target and t have an influence, while investment bank recommendations seem to be most relevant when assessing the predictive of a WSB post.

Model Configuration	Target	t	WSB Features	WSB Metadata	Investment Banks	Technical Indicators	Stock Price Data	Text Information
Q3 2021	G_t	1w	3	1	4	1	1	0
		1m	2	1	3	2	2	0
		3m	2	0	4	1	3	0
	M_t	1w	3	1	5	0	1	0
		1m	2	0	1	2	5	0
		3m	1	0	4	1	4	0
	P_t	1w	0	0	10	0	0	0
		1m	0	0	9	0	1	0
		3m	0	0	5	0	4	0
Q1 2022	G_t	1w	0	4	3	0	3	0
		1m	1	0	6	1	2	0
		3m	1	0	4	2	3	0
	M_t	1w	0	3	4	0	3	0
		1m	2	0	2	3	3	0
		3m	2	0	0	3	5	0
	P_t	1w	0	0	8	0	2	0
		1m	0	0	9	0	1	0
		3m	0	0	6	0	4	0

the investment horizon of retail investors. We consider these results valuable, as studies have shown that consistently beating the market is difficult, even for large equity funds [19]. The analysis of our community-informed models' important features shows that our models rely on a combination of WSB-based features, stock market data, and investment bank recommendations to judge a new WSB post's predictive value. We conclude that this shows that WSB exhibits collective intelligence in its investment recommendations, which can be mined from the posts with custom-trained ML models. In future work, our methodology could be extended to different sets of stocks, e.g., the broader Russell 3000 index. Furthermore, there could be additional information that could be extracted from the unstructured texts of WSB, given more advanced approaches. Recent advances in the field of NLP could yield large language models that are able to correctly identify sentiment, polarity, and similar features within the irregular texts of WSB to provide additional, valuable features.

Broader Perspective & Ethical Considerations: Our results show that the overall performance of the model was able to beat the market. However, these results are only achieved when all of the model's decisions were followed with an investment. Following a subset of these signals, e.g., due to limited funds or time, could lead to worse performance. While the results of the proposed model beating the broader market are remarkable, investors following such a model should always try their best to understand the mechanics and decision patterns of the model in order to avoid making uninformed decisions, especially when a significant amount of money is to be invested. Automated trading is particularly risky. The hype around the GameStop stock has shown that when a sufficient amount of investors participate, the price of a stock that does not seem attractive in fundamental analysis can increase significantly. This means that if a sufficient number of investors follow the signals of a model such as the one proposed in this work, the stock price can be significantly affected, even if the signal itself was inaccurate, distorting the predictive value of such a model.

REFERENCES

[1] Prasad N Achyutha, Sushovan Chaudhury, Subhas Chandra Bose, Rajnish Kler, Jyoti Surve, and Karthikeyan Kaliyaperumal. 2022. User Classification and Stock Market-Based Recommendation Engine Based on Machine Learning and Twitter Analysis. *Mathematical Problems in Engineering* 2022 (2022), 1–9.

[2] Pratik Agrawal, Tolga Buz, and Gerard de Melo. 2022. WallStreetBets Beyond GameStop, YOLOs, and the Moon: The Unique Traits of Reddit's Finance Communities. In *AMCIS 2022 Proceedings*, Vol. 8. AIS eLibrary, 1–10.

[3] Abhinav Anand and Jalaj Pathak. 2021. WallStreetBets against Wall Street: The Role of Reddit in the GameStop short squeeze. *IIM Bangalore Research Paper* 644 (2021), 1–45.

[4] Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models.

[5] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 14. 830–839.

[6] Christian Boylston, Beatriz Palacios, Plamen Tassev, and Amy Bruckman. 2021. WallStreetBets: Positions or Ban. arXiv:2101.12110v1.

[7] Daniel Bradley, Jan Hanousek Jr, Russell Jame, and Zicheng Xiao. 2021. Place your bets? The market consequences of investment research on Reddit's Wallstreetbets.

[8] Leo Breiman. 1996. Bagging predictors. *Machine learning* 24, 2 (1996), 123–140.

[9] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[10] Tolga Buz and Gerard de Melo. 2022. Democratization of Retail Trading: Can Reddit's WallStreetBets Outperform Investment Bank Analysts?

[11] Tolga Buz and Gerard de Melo. 2023. WallStreetBets: An Analysis of Investment Advice Democratization. In *Proceedings of the 56th Hawaii International Conference on System Sciences*, Vol. 56. 2150.

[12] Ryan G Chacon, Thibaut G Morillon, and Ruixiang Wang. 2022. Will the reddit rebellion take you to the moon? Evidence from WallStreetBets. *Financial Markets and Portfolio Management* (2022), 1–25.

[13] Tianqi Chen and Carlos Guestrin. 2016. XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. https://doi.org/10.1145/2939672.2939785

[14] Guillaume Coqueret and Tony Guida. 2020. *Machine Learning for Factor Investing: R Version.* Chapman and Hall/CRC.

[15] Michele Costola, Matteo Iacopini, and Carlo RMA Santagiustina. 2021. On the "mementum" of meme stocks. *Economics Letters* 207 (2021), 110021.

[16] Piet JH Daas and Marco JH Puts. 2014. *Social media sentiment and consumer confidence.* Technical Report. ECB Statistics Paper.

[17] Sara De Candia, Gianmarco De Francisci Morales, Corrado Monti, and Francesco Bonchi. 2022. Social Norms on Reddit: A Demographic Analysis. In *Proceedings of the 14th ACM Web Science Conference 2022* (Barcelona, Spain) *(WebSci '22)*. Association for Computing Machinery, New York, NY, USA, 139–147. https://doi.org/10.1145/3501247.3531549

[18] Selin Duz Tan and Oktay Tas. 2021. Social Media Sentiment in International Stock Returns and Trading Activity. *Journal of Behavioral Finance* 22, 2 (2021), 221–234. https://doi.org/10.1080/15427560.2020.1772261

[19] Tim Edwards, Anu Ganti, Craig Lazzara, Joseph Nelesen, and Davide Di Gioia. 2022. *SPIVA U.S. Mid-Year 2022.* https://www.spglobal.com/spdji/en/spiva/article/spiva-us

[20] Yves Gendron, Alexandre Madelaine, Luc Paugam, and Hervé Stolowy. 2023. Popular expertise in financial markets: An analysis of Due Diligence posts on WallStreetBets.

[21] EP George, Gwilym M Jenkins, and Gregory C Reinsel. 1970. *Time series analysis: forecasting and control.* Wiley.

[22] Yuriy Gorodnichenko, Tho Pham, and Oleksandr Talavera. 2018. *Social media, sentiment and public opinions: Evidence from #Brexit and #USElection.* Technical Report. National Bureau of Economic Research.

[23] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. 2003. KNN model-based approach in classification. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings.* Springer, 986–996.

[24] Anish Gupta, Manish Gupta, and Prateek Chaturvedi. 2020. Investing Data with Machine Learning Using Python. In *Strategic System Assurance and Business Analytics.* Springer, 1–9.

[25] Tim Hasso, Daniel Müller, Matthias Pelster, and Sonja Warkulat. 2022. Who participated in the GameStop frenzy? Evidence from brokerage accounts. *Finance Research Letters* 45 (2022), 102140.

[26] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength natural language processing in python.

[27] Philip N Howard, Aiden Duffy, Deen Freelon, Muzammil M Hussain, Will Mari, and Marwa Maziad. 2011. Opening closed regimes: what was the role of social media during the Arab Spring? SSRN 2595096.

[28] S&P Dow Jones Indices. 2023. *S&P 500®*. https://www.spglobal.com/spdji/en/indices/equity/sp-500/overview

[29] Charles M Jones, Adam V Reed, and William Waller. 2021. When Brokerages Restrict Retail Investors, Does the Game Stop? *Columbia Business School Research Paper* (2021).

[30] Hunter Jones and Joel Hietanen. 2022. The r/wallstreetbets 'war machine': Explicating dynamics of consumer resistance and capture. *Marketing Theory* (2022), 14705931221114172.

[31] Jan Marco Leimeister. 2010. Collective intelligence. *Business & Information Systems Engineering* 2, 4 (2010), 245–248.

[32] Zhuang Liu, Degen Huang, Kaiyu Huang, Zhuang Li, and Jun Zhao. 2021. Finbert: A pre-trained financial language representation model for financial text mining. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence.* 4513–4519.

[33] Suwan Long, Brian Lucey, Ying Xie, and Larisa Yarovaya. 2023. "I just like the stock": The role of Reddit sentiment in the GameStop share rally. *Financial Review* 58, 1 (2023), 19–37.

[34] Štefan Lyócsa, Eduard Baumöhl, and Tomáš Výrost. 2022. YOLO trading: Riding with the herd during the GameStop episode. *Finance Research Letters* 46 (2022), 102359.

[35] Thomas W Malone, Robert Laubacher, and Chrysanthos Dellarocas. 2010. The collective intelligence genome. *MIT Sloan management review* (2010).

[36] Lydia Manikonda, Ghazaleh Beigi, Subbarao Kambhampati, and Huan Liu. 2018. #metoo through the lens of social media. In *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11.* Springer, 104–110.

[37] Joe Marlo, Nicole Salani, and Bilal Waheed. 2021. To the Moon: Discourse on r/wallstreetbets. https://github.com/joemarlo/wsb-discourse

[38] Norma Mendoza-Denton. 2021. "Sticking It to the Man": r/wallstreetbets, Generational Masculinity and Revenge in Narratives of our Dystopian Capitalist Age. *Anthropology Now* 13, 1 (2021), 91–99.

[39] Daniele Miorandi, Vincenzo Maltese, Michael Rovatsos, Anton Nijholt, and James Stewart. 2016. *Social collective intelligence.* Springer.

[40] Usman Naseem, Adam G Dunn, Jinman Kim, and Matloob Khushi. 2022. Early identification of depression severity levels on reddit using ordinal classification. In *Proceedings of the ACM Web Conference 2022.* 2563–2572.

[41] Thien Hai Nguyen, Kiyoaki Shirai, and Julien Velcin. 2015. Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications* 42, 24 (2015), 9603–9611.

[42] Venkata Sasank Pagolu, Kamal Nayan Reddy, Ganapati Panda, and Babita Majhi. 2016. Sentiment analysis of Twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (SCOPES).* IEEE, 1345–1350.

[43] Jigar Patel, Sahil Shah, Priyank Thakkar, and Ketan Kotecha. 2015. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert systems with applications* 42, 1 (2015), 259–268.

[44] Marius-Constantin Popescu, Valentina E Balas, Liliana Perescu-Popescu, and Nikos Mastorakis. 2009. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems* 8, 7 (2009), 579–588.

[45] Thomas Raffinot and Sylvain Benoit. 2018. Investing through economic cycles with ensemble machine learning algorithms. *Available at SSRN 2785583* (2018).

[46] Gabriele Ranco, Darko Aleksovski, Guido Caldarelli, Miha Grčar, and Igor Mozetič. 2015. The effects of Twitter sentiment on stock price returns. *PloS one* 10, 9 (2015), e0138441.

[47] Matthias Schnaubelt, Thomas G Fischer, and Christopher Krauss. 2020. Separating the signal from the noise–financial machine learning for twitter. *Journal of Economic Dynamics and Control* 114 (2020), 103895.

[48] Detlef Schoder, Peter A Gloor, and Panagiotis Takis Metaxas. 2013. Social media and collective intelligence—ongoing and future research streams. *KI-Künstliche Intelligenz* 27, 1 (2013), 9–15.

[49] Peter Kalum Schou, Eliane Bucher, Matthias Waldkirch, and Eduard Grünwald. 2022. We Did Start the Fire: r/wallstreetbets, 'Flash movements' and the Gamestop Short-Squeeze. In *Academy of Management Proceedings*, Vol. 2022. 14028.

[50] Melody Sepahpour-Fard and Michael Quayle. 2022. How Do Mothers and Fathers Talk About Parenting to Different Audiences? Stereotypes and Audience Effects: An Analysis of r/Daddit, r/Mommit, and r/Parenting Using Topic Modelling. In *Proceedings of the ACM Web Conference 2022.* 2696–2706.

[51] Stefan Stieglitz and Linh Dang-Xuan. 2013. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of management information systems* 29, 4 (2013), 217–248.

[52] Hong Kee Sul, Alan R Dennis, and Lingyao Yuan. 2017. Trading on twitter: Using social media sentiment to predict stock returns. *Decision Sciences* 48, 3 (2017), 454–488.

[53] James Surowiecki. 2005. *The wisdom of crowds.* Anchor.

[54] Sean J Taylor and Benjamin Letham. 2018. Forecasting at scale. *The American Statistician* 72, 1 (2018), 37–45.

[55] Zaghum Umar, Mariya Gubareva, Imran Yousaf, and Shoaib Ali. 2021. A tale of company fundamentals vs sentiment driven pricing: The case of GameStop. *Journal of Behavioral and Experimental Finance* 30 (2021), 100501. https://www.sciencedirect.com/science/article/pii/S2214635021000459

[56] Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. BloombergGPT: A Large Language Model for Finance.

[57] Yang Yu, Wenjing Duan, and Qing Cao. 2013. The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision support systems* 55, 4 (2013), 919–926.