Mood Swings the Market: Role of Reddit Sentiments in the Gamestop Saga

Jasleen Madan

July 2025

1 Abstract

Existing literature has emphasized the role of social media platform Reddit in the GameStop (GME) saga, highlighting the influence of investor sentiment on retail investing, herd behavior, and the broader impact of online social networks on financial collective action. However, current sentiment analysis methods on Reddit in the context of the GameStop, largely reliant on lexicon-based approaches and static sentiment dictionaries, lack financial domain-specific training leading to misclassification; highlighting the need for more context-aware models. To address this gap, I analyze threads from the r/WallStreetBets subreddit from December 15, 2020 to February 28, 2021 applying a context-aware, prompt-based sentiment analysis framework using the current state-of-the-art large language model Gemma 3 (12B parameters). This study investigates the predictive potential of Reddit activity, post popularity, and sentiment data in forecasting GameStop (GME) stock prices at 5-minute intervals using a non-autoregressive Long Short-Term Memory (LSTM) based multistage framework. The results indicate that while price data alone provides a baseline level of predictive accuracy, augmenting it with Reddit activity metrics (such as number of comments and votes indicated by scores) at 5-min level leads to increase in R^2 and a decrease in Mean Absolute Error (MAE), indicating improved predictive accuracy. In particular, models combining price data with Net sentiment (Positive- Negative) aggregated at same temporal resolution yield the highest performance gains, suggesting that social sentiment dynamics influence price movements, particularly in bullish and euphoric market conditions. Examining the relative predictive relationship adds to the expanding research on "meme stocks" and the influence of investment forum discussions on intraday stock price fluctuations.

Keywords: Meme stocks, Natural language processing, Social media-driven trading, Financial forecasting, Reddit activity, Sentiment analysis

2 Introduction

2.1 Background

GameStop (GME), an American retailer specializing in electronics and video games, has been around since the 1980s. However, its performance declined from the mid-2010s due to the rise of online shopping and the shift toward digital game downloads. The COVID-19 pandemic in 2020 further worsened its struggles, impacting brick-and-mortar retailers worldwide. As a result, short sellers, particularly hedge funds, increasingly bet against GameStop, with some anticipating its bankruptcy. This caused the price of GME's stock to fall from around $45 at the end of 2015 to just $3–$5 by mid-2020. Thus, GameStop (GME) became the most heavily shorted stock on Wall Street by major hedge funds, with nearly 140% of its public float sold short.

By April 2020, GameStop had already attracted attention on Reddit, where users noted its exceptionally high short interest of 84% Kochkodin (2021). At the same time, activists like Keith Gill strongly advocated for its undervaluation and growth potential Popper and Browning (2021), citing factors such as stronger-than-perceived financials, an upcoming console cycle, and a strategic shift toward e-commerce.

WSB users, likely motivated by both profit opportunities and resentment toward institutional investors Chohan (2022), coordinated efforts to trigger a short squeeze, compelling short sellers to cover their positions at substantial losses. On January 20, 2021, GameStop closed at $39.12 on the New York Stock Exchange (NYSE), but within a week, its stock surged to $347.51, marking an extraordinary 788% increase. The rally continued, reaching a peak of $483.00 on January 28, 2021, with over one million shares failing to deliver—solidifying the short squeeze's success. According to Mancini et al. (2022), this unprecedented market event resulted from large-scale coordination on social media, driving collective action that significantly impacted the stock market.

2.2 Previous Works

The GameStop rally, fueled by the "to the moon" movement on r/WallStreetBets, has garnered significant academic attention, offering insights into the market impact of online investor coordination and discussions on investment forums Corbet et al. (2021), Lyócsa et al. (2022), Anand and Pathak (2022), Long et al. (2023),Mancini et al. (2022). In parallel, Robinhood, a platform that democratizes access to financial markets, has been shown to attract attention-driven traders Banerji et al. (2021), a demographic also targeted on Reddit,Anand and Pathak (2022), Long et al. (2023), highlighting the importance of studying retail investors' behavior during such events.

The Gamestop saga also highlights the role played by media in shaping financial markets. A vast literature has explored how media coverage, including news outlets, social media, and financial platforms, shapes public perception, drives market trends, and triggers price fluctuations. One

strand of research has examined the role of traditional financial news sources in influencing behavior. Atkins et al. (2018) have shown that information extracted from textual news sources can be used in predicting directional changes in market volatility.Corbet et al. (2020), Anamika and Subramaniam (2022) have studied the impact of macroeconomic news and new headlines in cryptocurrency markets. The impact of media is especially pronounced during stock market bubbles Dyck and Zingales (2003),Campbell et al. (2012), Bhattacharya et al. (2009) and recession and stressed economic times. Wu et al. (2004),Garcia (2013)

Lately, the focus has shifted to analysing social media, particularly Twitter and Reddit, in shaping retail investor behaviour, since these platforms are capable of disseminating information to wider audience at a faster pace and use informal language which individual investors are more sensitive and responsive to Rennekamp and Witz (2021),Gu and Kurov (2020). Tollefson (2023) state that younger investors aren't afraid to turn to untraditional sources of information for investment advice. Individual investors are increasingly turning to social media platforms for investment advice from their peers due to constraints of limited attention and information overload from various sources Chohan (2022), Warkulat and Pelster (2024),Kaur et al. (2018). Moreover, Kadous et al. (2022) find that individual investors often rely on low-quality investment advice from social media, especially when it is negative, due to a lack of awareness about its predictive value and a tendency to accept information automatically. Investors using low-fee brokerages are particularly vulnerable to social media influence, especially from subreddits, which often leads them to make irrational financial decisions that harm their financial well-being and contribute to investment manias Gale (2022).Warkulat and Pelster (2024) show that meme investors in particular, on average, engage in speculative behavior when trading on attention and lose significant amounts of money doing so.

2.2.1 Why do sentiments matter?

Given the growing influence of social media on retail investor behavior, it is essential to focus on how social media sentiment shapes investment decisions and market dynamics. The role of sentiment in financial markets has had theoretical foundations since the 1990s, with a substantial body of literature examining how individual investors are often irrational and influenced by psychological biases Shiller (1990), Odean (1998), Puzzle et al. (1995), Barber and Odean (2008).

Recent research has found similar results. Madaan and Singh (2019) conducted an exploratory study on the impact of behavioral biases, such as overconfidence, anchoring, disposition effect, and herding, on investment decisions of individual investors on the National Stock Exchange, showing that overconfidence and herding bias significantly influence investment decisions. Semenova and Winkler (2021) find empirical evidence of psychological contagion among WSB users, where an initial set of investors attracts a larger and larger group of excited followers—net of any fundamental price movements. The study by Youssef and Waked (2022) examined herding behavior in the cryptocurrency market during the COVID-19 crisis, finding that herding tendencies emerge during high

volatility and crises, fueled by media coverage that stirred panic and frenzy among investors. Cookson et al. (2023) found that investors on StockTwits, particularly self-described bulls and bears, tend to seek out information that aligns with their existing beliefs, forming echo chambers. Similar findings have been observed across global markets Rehan and Umer (2017), Athur (2014), Kumar and Goyal (2016).

Sentiments are particularly studied in the stock market because they directly influence investor behavior, often leading to decisions driven by emotions rather than rational analysis. Shleifer and Summers (1990) argues that changes in investor sentiment, influenced by irrational beliefs, affect security returns, as arbitrage is limited due to the risks faced by rational investors, offering a superior alternative to the efficient markets approach. Thus, sentiments help explain why markets sometimes deviate from fundamental values, as investors' collective feelings of optimism, fear, or uncertainty can drive buying or selling activity Lerner and Keltner (2001), Schmeling (2009), Dong and Gil-Bazo (2020), Fernandez-Perez et al. (2024).

As a result, the role of sentiments and social media has been the subject of extensive research. In their seminal paper, Bollen et al. (2011) analyze the ability of social media, particularly Twitter, to reflect society's collective mood and its influence on the stock market, using Profile of Mood States (POMS) to assess public mood across multiple dimensions and demonstrate its predictive power the Dow Jones Industrial Average (DJIA). Gu and Kurov (2020) explore the predictive value of firm-specific sentiment from Twitter messages on stock returns, finding that Twitter sentiment forecasts returns without reversals. Their results suggest that Twitter provides new information, particularly about analyst recommendations, price targets, and earnings, which accounts for about one-third of its predictive power. Liang et al. (2020) compared the forecasting ability of three sentiment indices based on positive and negative posts from social media, newspaper articles, and online news, finding that while traditional newspapers do not influence Chinese stock markets, both social media and online news have significant predictive power.

2.2.2 Related works on Gamestop and investor sentiments

Coming to case in point, there are several studies focusing on the sentiment surrounding GameStop (GME), investigating how WallStreetBets (WSB) activity, conversation sentiment, and user interactions can predict retail trading behavior and GME returns. These works apply linear regression models and machine learning techniques to analyze the relationship between social media sentiment and stock price fluctuations. Celestin and Vanitha (2021) found that while traditional news sources could not predict Gamestop returns, sentiments expressed on Reddit forum were strongly predictive. Long et al. (2023) investigated the impact of sentiments from Reddit on GameStop's intraday returns at 1,5,10 and 30 min level, using a customized sentiment analysis dictionary based on the Valence Aware Dictionary and Sentiment Reasoner (VADER) [1] to analyze comments from r/WallStreetBets.

[1] `https://github.com/cjhutto/vaderSentiment`

They found that while net investment sentiments affect GME returns during bullish market conditions, there is limited influence during bearish market movements. Anand and Pathak (2022) found a significant predictive association between the tone of discussions on the WSB subreddit and GME returns, volatility, bid-ask spreads, and volumes (Dec 2020–Mar 2021), using sentiment lexicons and manual adjustments for WSB slang and emojis. They observed the impact was largely driven by a small group of 462 influential subredditors. Hu et al. (2021) used an adapted the Loughran and McDonald (LM) dictionary, incorporating WallStreetBets-specific slang and emojis, to measure the tone of Reddit submissions and found positive tones predicted higher future returns, especially when agents had higher influence. Kim et al. (2023) classified Twitter and Reddit posts into positive and negative sentiments, measuring the proportion of each sentiment and found that both the valence and volume of these posts significantly impacted GME's intraday transaction volumes and contributed to irrational trading behaviors. Notably, the study highlights how individual investors—particularly on platforms like r/WallStreetBets—collaboratively shared sentiment and trading strategies, fostering a collective sense of being well-informed, or social informedness.This perceived informedness, despite limited reliance on traditional financial analysis, mobilized coordinated trading actions that substantially amplified both the volatility and trading volume of GME stock. Wang and Luo (2021) combined VADER for sentiment polarity and BERT model for sentiment embeddings (vector based on the meaning of the posts' words) to predict price movements and found combination of both to yield highly accurate predictions. Zheng et al. (2021) used the NTUSD-Fin market sentiment dictionary, specifically designed for financial social media, to provide more accurate sentiment analysis compared to off-the shelf VADER or general sentiment models and found GME's stock price could be partly explained by popularity of the dominant topic, topic cohesiveness, and sentiment divergence between interacted users on r/WSB. Fernandez-Perez et al. (2024) analyzed emotions at a more granular level than sentiments, utilizing EmoLex [2] and EmTract [3]—emotion lexicons based on Plutchik's wheel of eight emotions—and found that joy positively predicts returns just before the GME bubble peaks, fear at the bubble's peak, and anger after the bubble bursts.

2.2.3 Gaps in current works

While lexicon-based approaches like VADER dictionaries are popular in finance literature due to their simplicity and lack of training data requirements, they are prone to significant noise because they are not trained on financial datasets. Tools such as VADER, ANEW, and EmoLex, although customizable and backed by years of research, often fail to account for context and language nuances, leading to potential misclassification.Wang and Luo (2021) demonstrated that off-the-shelf VADER sentiment analysis produces disappointing results on text from r/wallstreetbets due to frequent use of sarcasm, irony, slang, investment jargon, emojis, and non-standard word usages specific to the subreddit or investment domain. Fernandez-Perez et al. (2024) also acknowledge that Emolex contain unequal distribution of words across eight emotions and constraints in capturing context, ambiguity, and evolving language. Long et al. (2023) attempts to address this limitation by customizing

[2] Source:http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

[3] Source:https://github.com/dvamossy/EmTract/blob/main/README.md

VADER with human-annotated sentiment scores, but the underlying constraints still persist. Kant et al. (2024) employed both fine-tuned (Bidirectional Encoder Representations from Transformers)[4] BERT-Sentiment and VADER to offset any potential shortcomings of either method. Ferraro and Sperlì (2024) use roBERTa [5] to extract sentiments from different Online Social Networks (Reddit and Twitter) and predict daily stock prices using multiple AI models. Kim et al. (2023) also used fine-tuned BERT-based models quantifying the proportion of positive and negative sentiments.

Despite these efforts, existing models still lack domain adaptation, robustness to informal language or nuanced understanding of the context, instruction-following capability, and the ability to generalize across dynamic financial discourse. Most of the current work either relies on static rule-based systems or general-purpose transformers not fine-tuned on financial or social media-specific corpora with limited few-shot adaptability in sentiment analysis tasks. This leads to reduced effectiveness in parsing sentiment from communities like r/wallstreetbets, where users often communicate with high levels of sarcasm, memes, and domain-specific terminology Zhang et al. (2023c), Zhang et al. (2023a), Yang et al. (2020).

3 Data and Methodology

3.1 Data

This paper utilizes two primary datasets: Reddit data from the r/wallstreetbets subreddit (`https://www.reddit.com/r/wallstreetbets/`) and minute-level stock price information from FirstRate Data (`https://firstratedata.com/`). The r/wallstreetbets subreddit was selected for the GameStop analysis because it became a central hub for retail investors during the GameStop short squeeze. The subreddit is known for its high volume of discussions related to stock trading, often characterized by memes, sarcasm, and a strong sense of community-driven market action. Historical intraday stock price information is collected from FirstRate data which sources stock data directly from major exchanges and fully adjust for both splits and dividends. All stock info is rigorously tested for accuracy and completeness.

3.1.1 Reddit data

The Reddit data was originally obtained by scraping using the Pushshift API, which provided access to archived Reddit data prior to its decommissioning in April 2023. The datasets are now hosted on Academic Torrents [6], curated by the user Watchful1 [7]. Each subbreddit data is contained in two files - posts and comments. Posts on Reddit are main content submissions, while comments are responses to posts or other comments, both with associated metadata.

[4]Source: `https://huggingface.co/google-bert/bert-base-uncased`
[5]Source: `https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment.`
[6]`https://academictorrents.com/details/1614740ac8c94505e4ecb9d88be8bed7b6afddd4`
[7]`https://www.reddit.com/user/Watchful1/`

For the purposes of this paper, Reddit post data spanning from December 15, 2020, to February 15, 2021, is considered, comprising 878,898 unique post entries. Each entry includes the post title, body, timestamp, score, upvote ratio, and the number of comments received. The upvote ratio represents the proportion of upvotes to total votes, while the score is the difference between upvotes and downvotes on a post or comment. Popular threads tend to attract more upvotes and comments from forum participants. Therefore, the number of comments on a post serves as a proxy for Reddit activity, capturing both the volume of engagement and user interest in a particular discussion, while the post score reflects its overall popularity and virality.

The decision to exclude comments from the analysis is due to a low signal-to-noise ratio, as more than 50% of the body text was removed or deleted by Reddit moderators during the squeeze period. This significantly complicates sentiment analysis, as sentiment in comments often depend on the original body text.However, all post titles were retained. To create a comprehensive dataset for textual analysis, the post titles were appended to the body text to form the fulltext for each post. The data is then cleaned by removing deleted entries, irrelevant characters, duplicate comments, bot-generated content, and URLs. Additional variables are created to mark trading hours and trading days. Posts were further classified based on the timeline of Gamestop squeeze.

- **Pre-Squeeze (before January 13, 2021)**
  Reflects baseline discussions before mainstream media attention and substantial price volatility. On January 13th Keith Gill publicly touted Gamestop stock on Reddit causing 50% spike.[8]

- **During (January 13–27, 2021)**
  Captures the rapid escalation in GME stock prices, extreme trading volume, significant media attention, and peak retail investor activity, culminating on January 27 which stock price reaching the peak as major short sellers close positions at significant losses.

- **Post-Squeeze(January 28–February 15, 2021)**
  Defined by brokerage (Robinhood and other platforms) trading restrictions, price corrections, increased regulatory scrutiny (SEC), and shifting community sentiment. Malz (2021)

Further information on in-depth Gamestop timeline can be found in the reference provided.[9] Table 1 reports the daily activity and scores after filtering. The analytical sample shows posts made during the trading day and market hours after accounting for holidays. As shown in Figure 1 and Figure 2, which present daily post scores and submission counts on r/WallStreetBets across different phases of the GameStop squeeze, the trading restrictions imposed on January 28 corresponded with a noticeable surge in discussions and overall platform activity with some posts receiving score as high as 417276. The dip in the submissions post-squeeze was due to r/WallStreetBets moderators temporarily setting the subreddit to private to manage the influx of new users.

Due to the nature of the forum, it is difficult to distinguish retail investors from other r/WallStreetBets users. Therefore, following the approach outlined in Long et al. (2023) and Wang and

[8]source: YahooFinance2021

[9]`https://www.thestreet.com/investing/stocks/a-timeline-of-the-gamestop-short-squeeze`

Luo (2021), all r/wallstreetbets participants are assumed to be speculative retail traders.

Dataset A: Reddit Posts					
Full Sample					
	Total posts	**Daily average**	**Total comments**	**Daily Comment average**	**Max Scores**
Pre squeeze	28965	998.79	1430090	49313.44	70557
During squeeze	160071	10671.40	2922977	194865.13	211767
Post squeeze	689862	36308.52	7408135	389901.84	417276
Analytical Sample - Active trading days 9:30 AM EST to 4:00 PM EST					
Pre squeeze	10012	345.24	140198	4834.41	70557
During squeeze	68556	4570.40	780309	52020.60	197252
Post squeeze	268426	14127.68	1904771	100251.10	231278

Table 1: *This table reports the summary statistics for the Reddit posts (Dataset A). In Full Sample, all Reddit posts within the sample period are shown containing 878,898 entries whereas in the analytical sample only those posted on weekdays during the stock market trading hours (9:30–16:00 Eastern Standard Time) are considered which contains a total of 346,994 posts. The sample period from December 15, 2020 to February 15, 2021.*

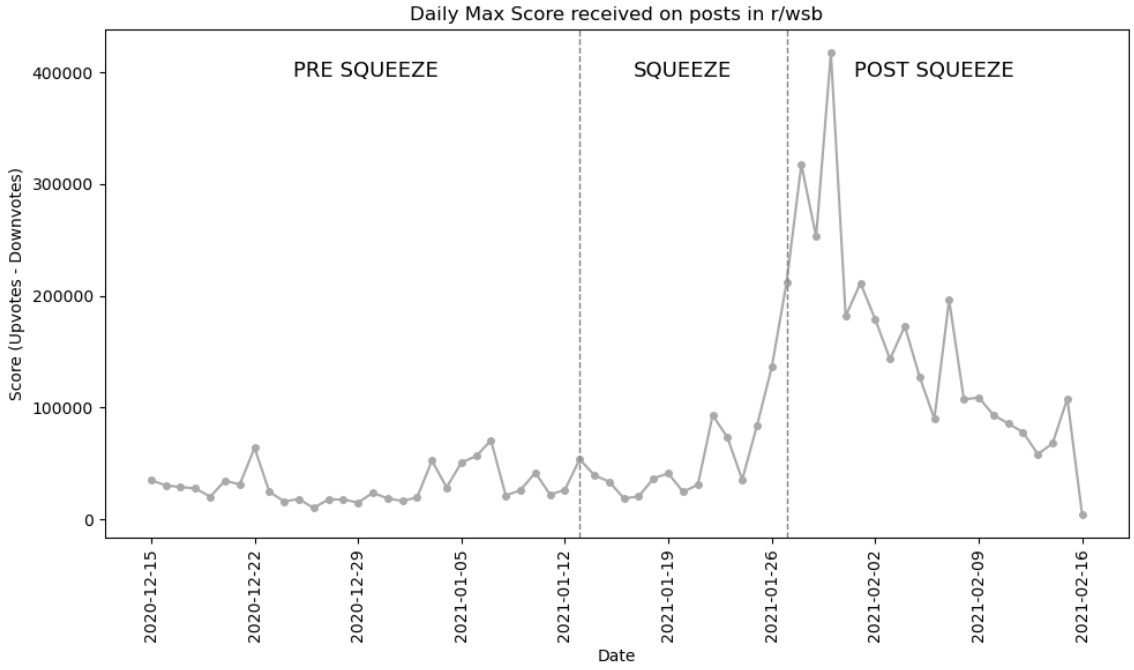

Figure 1: *Daily maximum post scores on r/WallStreetBets*

3.1.2 Stock data

Stock data for GameStop (ticker: GME), collected from FirstRate Data, includes Price (Open, High, Low, Close) and Volume information at various time intervals—1-minute, 5-minute, 1-hour, and daily— covering the period from December 15, 2020, to February 15, 2021 during regular trading hours (9:30 AM ET to 4:00 PM ET) on market days. For all the primary analysis, GME closing

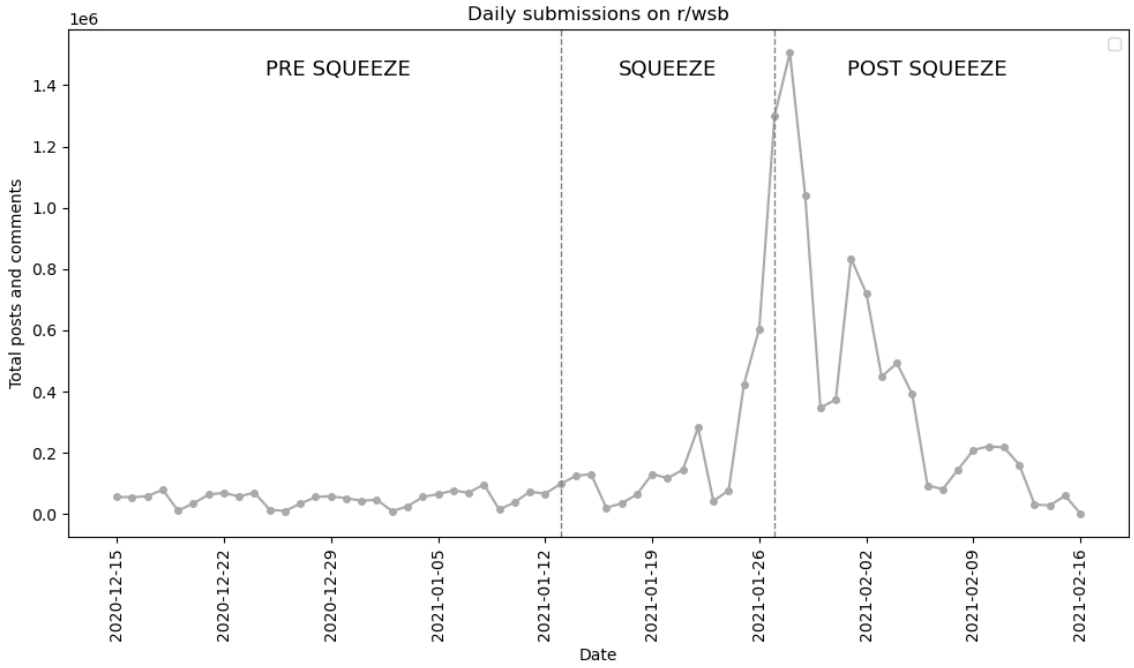

Figure 2: *Daily submissions on r/WallStreetBets*

stock price is considered.

Figure 3 plots the daily closing price of GME alongside the S&P 500 Index (SPX), a benchmark index representing the performance of 500 large-cap U.S. companies and widely regarded as a key indicator of overall U.S. stock market performance. The plot also includes GME vs IWM (iShares Russell 2000 ETF) of which GME is a constituent. It is evident that while both the broader market (as reflected by the SPX) and the Russell 2000 ETF (IWM) followed a similar trajectory— experiencing a dip in the early weeks of 2021 likely due to pandemic-related uncertainties— GME deviated sharply from this trend, emerging as a notable anomaly.

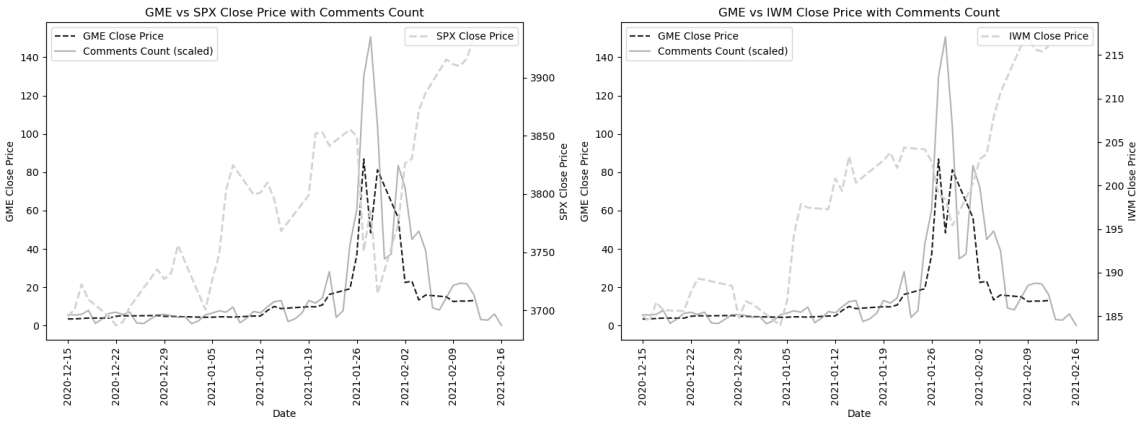

Figure 3: *Daily closing price of GME vs. SPX (left) and GME vs. IWM (right), alongside comment volume on r/WallStreetBets. GME's price movement aligns more closely with Reddit activity than with broader market indices.*

3.2 Methodology

The analysis focuses on posts made during trading hours, from 9:35 AM to 3:55 PM EST, excluding the first and last 5 minutes of the day to avoid biases related to market open and close, as is standard in market microstructure research. After-market hours were not considered for analysis since only particular subset of retail investors participate in extended hours trading. Most spikes in Reddit submissions and scores occur in market hours during the squeeze. See Figure A.1 in the Appendix for further details. All the primary analysis on Reddit activity, scores and sentiments are conducted at a 5-min frequency level. Given the volatility of GME during the squeeze period, daily or hourly frequencies were deemed unsuitable for capturing the rapid market movements. However, 1-minute intervals were also avoided, as they do not provide sufficient time for sentiment to accumulate and for investors to react. Therefore, a 5-minute interval was chosen, striking a balance between capturing sentiment and allowing enough time for emotional responses and market adjustments. Prior studies support the claim that sentiments on Reddit and Twitter impacts returns with a delay of at least 5 minutes. (Fernandez-Perez et al. (2024), Long et al. (2023), Renault (2017)).

3.2.1 Overview

To model the temporal dynamics of GME stock prices, a Long Short-Term Memory (LSTM) network is employed, well-suited for capturing sequential dependencies in financial time series data. This is because classical models suffer from two major disadvantages: (1) they rely on strong assumptions such as linearity, stationarity, and the presence of autocorrelation. Since the considered timeframe centers around the short-squeeze period and includes its prelude and aftermath — an altogether volatile setting - linear models have suboptimal performance. (2) they are limited to regression tasks and struggle with capturing multivariate relationships and long-term temporal dependencies, which are crucial when modeling complex patterns in sentiment-driven stock price movements.

In contrast, LSTM models can learn and model multivariate, complex, nonlinear temporal relationships directly from the data, making them more robust in capturing the intricate dynamics between Reddit sentiment and stock price movements, especially when sentiment data is noisy or sporadic with minimal feature engineering. Althelaya et al. (2018), Selvin et al. (2017), Siami-Namini et al. (2018), Ferraro and Sperlì (2024)

The input to the LSTM consists of a sliding window of 12 consecutive observations of GME stock prices sampled at five-minute intervals, thereby encapsulating a one-hour context of historical pricing information. The model follows a non-autoregressive design, using only past observations within each window without feeding back its own predictions. To mitigate the risk of reward hacking—where the model might trivially replicate the most recent price due to minimal changes over short intervals—a deliberate one-hour gap is introduced between the end of the input sequence and the prediction target. This design compels the model to learn both short-term fluctuations and longer-term structural patterns in stock price behavior, rather than relying solely on immediate past

values. The network is trained using the Mean Squared Error (MSE) loss function, which penalizes large deviations between the predicted and actual prices. This approach promotes robust learning of price trajectories that generalize beyond short-term autocorrelations, aligning the model's predictive objectives with the economic characteristics of real-world stock movements.

Building upon the temporal prediction framework, three progressively structured modeling strategies are employed, each designed to strengthen the central argument that market sentiment, particularly retail investor sentiment expressed via Reddit, played a pivotal role in driving GME stock price movements. In the first stage, a baseline is constructed using only historical stock price data. The model receives a one-hour context of past price observations and is tasked with predicting the stock price one hour after this context, with an intentional one-hour gap in between. This design ensures that any predictive power arises from an understanding of broader price trends rather than short-term continuity, setting a foundational benchmark for subsequent stages.

In the second stage, the model is enriched by introducing Reddit activity from the r/wallstreetbets subreddit during the one-hour gap. This setup maintains the same structure of stock price context and prediction target, but now supplements the missing interval with exogenous data reflecting public engagement through the total number of comments and scores (net of upvotes and downvotes). These engagement indicators serve as a proxy for real-time investor attention and discourse, allowing the model to infer possible shifts in stock price in the absence of price data itself during the one-hour window.

The third stage further refines the integration of Reddit data by replacing the engagement-based proxies with semantic sentiment scores extracted from Reddit posts using a state-of-the-art large language model. Rather than relying solely on volume or popularity signals, the model now incorporates the actual emotional and narrative content of the discourse—whether it is bullish (positive), bearish (negative), or neutral towards GME. By sequentially transitioning from price-only data, to activity-based proxies, and finally to sentiment-rich textual representations, the modeling strategy is intentionally designed to test and build upon the hypothesis that retail investor sentiment, as expressed in real time on social media, not only correlates with but can predict stock price fluctuations.

3.2.2 Temporal Model

As mentioned earlier, a Long Short-Term Memory (LSTM) network, a type of recurrent neural network (RNN), is employed due to its suitability for modeling sequential data such as stock prices. The key advantage of LSTMs over traditional RNNs is their ability to capture long-range dependencies in the data, meaning that they can remember important past events while forgetting irrelevant ones. This property is crucial when modeling stock price movements, as financial markets often display patterns that span over varying time scales. In this framework, the LSTM learns from the historical GME stock prices over a one-hour window, making it capable of identifying patterns in

price movements that are not immediately obvious from short-term fluctuations.

LSTM Structure The LSTM is structured around three gates—*input gate*, *forget gate*, and *output gate*—which control the flow of information through the network. Each gate has an associated weight matrix that is updated during training to optimize the model's performance. The mathematical operations for these gates are as follows:

1. **Forget Gate**: This gate determines what information from the previous time step should be discarded. The forget gate is computed as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

where: - f_t is the forget gate's output at time step t, - W_f is the weight matrix associated with the forget gate, - h_{t-1} is the previous hidden state, - x_t is the current input (i.e., the stock price at time t), - b_f is the bias term for the forget gate, - σ is the sigmoid activation function, which outputs a value between 0 and 1, indicating the degree of forgetting.

2. **Input Gate**: This gate determines which new information will be added to the cell state. The input gate is computed as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

where: - i_t is the input gate's output at time step t, - W_i is the weight matrix for the input gate, - b_i is the bias term for the input gate.

3. **Cell State Update**: The new information to be added is generated by the candidate cell state, denoted by $\tilde{C}_t$, and is computed as:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

where tanh is the hyperbolic tangent function, which ensures that the candidate cell state is bounded between -1 and 1.

4. **Output Gate**: The output gate decides what the next hidden state h_t will be. The hidden state is passed to the next time step and also to the final output. The output gate is computed as:

$$h_t = o_t \cdot \tanh(C_t)$$

where: - o_t is the output gate's value at time step t, - C_t is the updated cell state, - h_t is the output of the LSTM at time step t, which will be used in the next time step or as the final prediction.

The update to the cell state is given by:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

where C_{t-1} is the previous cell state, and C_t is the updated cell state at time t.

Loss Function The model is trained to predict the GME stock price one hour ahead based on the historical data. To evaluate the accuracy of its predictions, the Mean Squared Error (MSE) loss function is employed, a standard choice for regression tasks. The MSE loss function penalizes large deviations between the predicted and actual stock prices, thereby encouraging the model to make more accurate predictions. Mathematically, the MSE is defined as:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where: - N is the number of data points in the batch, - y_i is the true stock price at time i, - $\hat{y}_i$ is the predicted stock price at time i.

Minimizing this loss function ensures that the model is learning to predict the stock prices in a way that reduces the error between the predicted and actual values.

Training Procedure During training, the LSTM network learns the weights of the gates and the cell state updates by adjusting the parameters to minimize the MSE loss. This is achieved using backpropagation through time (BPTT), a variant of the backpropagation algorithm specifically designed for recurrent neural networks. BPTT works by computing the gradients of the loss with respect to the model's parameters and updating the weights in the direction that reduces the error. Over multiple iterations, the model learns to predict stock prices based on past price information, while also considering the broader trends that influence market behavior.

In the subsequent modeling stages, the input data is expanded to include Reddit post data, which is hypothesized to be a significant factor influencing GME stock price movements. By introducing sentiment proxies—starting with Reddit activity and engagement data and then replacing them with sentiment-rich textual data—the model is designed to capture the influence of social media sentiment on stock price dynamics, thereby enhancing predictive performance.

3.2.3 Extracting sentiments

Sentiment analysis of social media discourse on GameStop (GME) stock was conducted using a structured Chain-of-Thought (CoT) prompting approach to ensure transparency, interpretability, and replicability. The method employed the current state-of-the-art Gemma 3 (12B parameters) language model [10], explicitly guiding it to assess sentiment specific to GME while filtering out unrelated emotions, humor, or off-topic content. Gemma-3 was selected due to its competitive performance among open-weight instruction-tuned models and its strong generalization capabilities across downstream tasks, including text classification. Its relatively smaller size compared to some proprietary models makes it more computationally accessible while still benefiting from Google's instruction tuning and optimization strategies for helpfulness and reliability. The model's output combined general reasoning with seperate category-specific justifications—positive, negative, and neutral—providing clear, evidence-based sentiment attributions. This layered explanation enhances methodological clarity, supports error analysis, and mitigates risks of misinterpretation in complex

[10] `https://huggingface.co/google/gemma-3-12b-it`

or ambiguous posts.

The model was prompted to act as an expert financial sentiment analyst, scoring sentiment towards GME across three sentiment dimensions: Positive, Negative and Neutral, each rated on a scale from 0 to 10. A score of 0 indicated the complete absence of a particular sentiment, while a score of 10 reflected its maximum presence. Positive sentiment captured explicitly favorable, optimistic, or supportive language toward GME; negative sentiment reflected unfavorable, critical, or pessimistic commentary; and neutral sentiment measured the extent to which a post was factual, unemotional, ambiguous, or devoid of a clear opinion, with high scores indicating strong neutrality. Sentiment scores were assigned only when posts included clear emotional or opinionated language specifically directed toward GME. Mere mentions without discernible attitude were labeled fully neutral.

Posts expressing frustration about price declines were interpreted as implicitly bullish, reflecting continued emotional investment in the stock. Conversely, criticism directed toward hedge funds, market makers, or brokerage platforms—in the context of the GME short squeeze—was treated as supportive of GME and thus assigned a positive sentiment score. Given the prevalence of irony and coded language in retail investor forums (e.g., Reddit's r/WallStreetBets), the model was instructed to interpret sarcasm and cultural idioms (e.g., "diamond hands," "the squeeze," "Gamestonks") based on contextually grounded community norms.

In addition to sentiment evaluation, all explicit references to stock tickers or company names within each post were identified and recorded. Details of the prompt and output of sample posts are included in the Appendix A.2 and A.3.

Once the sentiment labels and the scores are extracted, a new variable named NET sentiment is made by taking the difference between Positive and Negative sentiment score and averaging the value of the net scores over the 5 min period.

Figure 4 and Figure 5 show the total daily sentiment score (volume * intensity) and average daily sentiment scores respectively. The time series plots reveal distinct shifts across the pre-squeeze, squeeze, and post-squeeze periods. Prior to January 13, 2021, sentiment was moderately positive, with average positive scores consistently above 6 and minimal negative sentiment, indicating early optimism within the Reddit community. During the squeeze period (January 13–27), average positive sentiment peaked, and net sentiment reached its highest levels, reflecting heightened enthusiasm and coordinated bullish sentiment. In the post-squeeze phase, while positive sentiment declined slightly, negative sentiment remained low, and neutral sentiment increased, suggesting that even post-squeeze, users remained mostly loyal or cautious rather than bearish. Overall, sentiment remained predominantly positive throughout, closely mirroring the timeline and collective dynamics of the GME short squeeze.

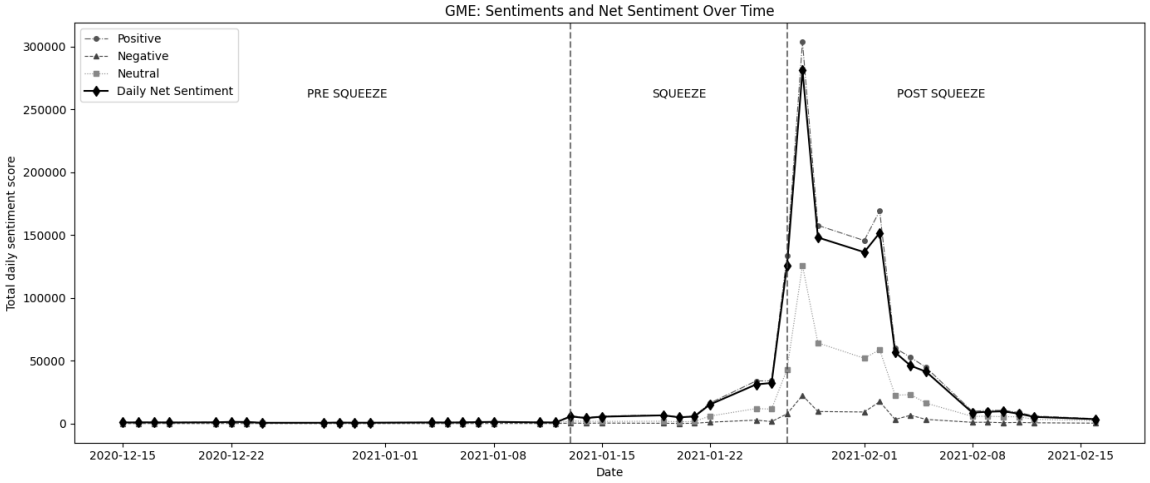

Figure 4: *Sentiment timeseries for posts explicitly mentioning GME stock. The figure captures how much enthusiasm / pessimism showed up that day (volume × intensity).Daily Net sentiment is calculated here as a difference between daily Positive and daily Negative score.*

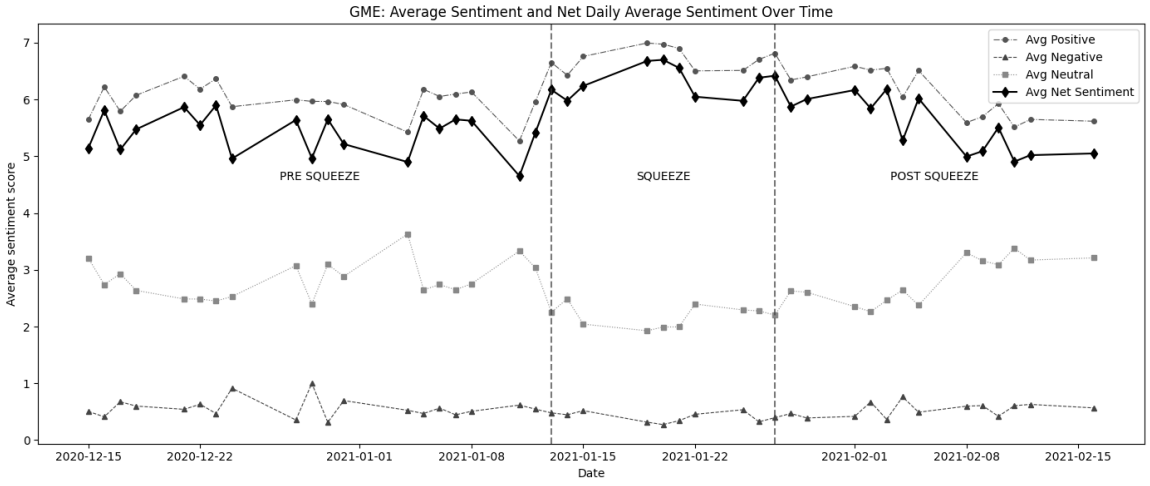

Figure 5: *Sentiment timeseries for posts explicitly mentioning GME stock. The figure captures how much How strongly the average post felt, independent of volume.Daily average net sentiment is calculated here as a difference between daily avg Positive and daily avg Negative score.*

3.2.4 Reliability check

To assess the reliability and accuracy of the sentiments extracted from the r/WSB dataset, a sample of 600 posts was annotated by five human annotators.[11]

To ensure high-quality annotations, the following procedure was implemented: (1) Annotators were first screened based on their fundamental knowledge of the stock market and were required to score above 80% to qualify. (2) Each qualified annotator then participated in a sentiment rating training session—conducted either online or in person—which included practice on 20 example posts. After completing the training, annotators were instructed to manually evaluate 600 posts using a standardized annotation guide. For each post, they identified the dominant emotion and independently rated emotional intensity on three scales—Positive, Negative, and Neutral—ranging from 0 (none) to 10 (very strong). Spam, incomplete sentences, bots comments and those unrelated to Gamestop were flagged. Sentiment assessments were made from the perspective of the Reddit user's attitude toward Gamestop. Notably, posts expressing anger toward hedge funds or trading platforms were interpreted as supportive of Gamestop (i.e., pro-GME) and annotated as exhibiting positive sentiment toward GME. This approach enabled the construction of temporally-resolved sentiment timelines reflecting public sentiment toward GME over the relevant period. (3) A crowd-based aggregation approach was used to determine dominant sentiment labels. Final sentiment ratings were derived by computing the mean of the individual intensity scores across annotators for each sentiment dimension. Recalibrated scores yielded high inter-rater reliability: Fleiss's Kappa = 0.66; ICC = 0.959 (Positive), 0.932 (Negative), and 0.908 (Neutral). Additional information, including annotator evaluation and confusion matrices illustrating individual annotator label classification, is provided in Appendix Table 5 and Figure A.2.

Subsequently, multiple sentiment analysis models were evaluated against the human-annotated dataset, which served as the basis for the accuracy assessment. These models (VADER, finBERT-tone [12] , FinGPT (Llama2 13B, LoRA) [13], Gemma-3 were selected based on their established use in the prior literature or their specific focus on financial sentiment text analysis for short pieces of texts such as social media posts. Apart from VADER which uses rule-based sentiment scoring, all other models used are large language models. finBERT-tone is a BERT model fine-tuned on 10,000 manually annotated (positive, negative, neutral) sentences from analyst reports. FinGPT v3 are LLMs finetuned with the LoRA method on the News and Tweets sentiment analysis dataset and are touted to outperform BloombergGPT in financial sentiment analysis Zhang et al. (2023a), Zhang et al. (2023b). Moreover, unlike base LLMs trained solely for next-word prediction, instruction-tuned models like finGPT and Gemma-3 are explicitly optimized to follow human instructions. Their con-

[11] This sample size was determined using the standard statistical formula for population sampling ($n = [z^2 \times p(1\text{-}p)] \div e^2$), where z=1.96 (95% confidence level), p=0.5 (maximum variance assumption), and e=0.05 (5% margin of error). Due to the imbalanced proportion of sentiments, with positive sentiment being overrepresented in the initial sample, the annotation set was expanded to 600 posts to ensure adequate representation across all sentiment categories while maintaining statistical significance.

[12] `https://huggingface.co/yiyanghkust/finbert-tone`

[13] `https://huggingface.co/oliverwang15/FinGPT_v33_Llama2_13B_Sentiment_Instruction_LoRA_FT_8bit`

sistent and predictable outputs also make them more reliable for controlled experimental settings. Zhang et al. (2023c).

Taking into account the inherent imbalance of the data set with respect to the positive sentiment for GameStop, comprehensive accuracy metrics—including the F1 score, precision, recall, and overall accuracy—are calculated and reported in Table 2, along with the results of the per-sentiment class (positive, negative, neutral). Among the models tested, the Gemma-3 model, comprising 12 billion parameters, exhibited superior performance relative to alternative sentiment annotation methods used previously to analyze Reddit posts.

	VADER	finBERT-tone	FinGPT v3	Gemma-3
Accuracy	0.60	0.48	0.45	0.80
F1 score	0.46	0.40	0.41	0.68
Precision	0.48	0.48	0.53	0.81
Recall	0.48	0.47	0.47	0.62

Table 2: *Overall performance comparison of sentiment classification models based on accuracy and macro-averaged metrics. Gemma-3 outperforms the other models across all metrics*

4 Findings and Discussions

4.1 Empirical Results

Using a Long Short-Term Memory (LSTM)-based temporal prediction framework, the closing prices of GME stock were modeled at a 5-minute aggregation level. The model was trained on non-autoregressive input spanning from December 25, 2020, to January 19, 2021. This training window was selected to provide a sufficient number of data points to learn underlying stock movement patterns, including the initial upward momentum that began after January 13, 2021. To ensure consistent training dynamics and enable fair comparison across all experimental stages, the model's hyperparameters were fixed as follows: sequence length = 12, batch size = 32, learning rate = 0.00001, and number of epochs = 600. These settings allowed the model to reach a minimum in validation loss before overfitting effects emerged, ensuring stable and interpretable training behavior. Each stage experiment was run 30 times with different seeds for reliability.

Model predictions were evaluated across two distinct market phases: the "squeeze" period (January 20–27, 2021), hereafter referred to as *DS*, which was characterized by a sharp and sustained increase in the stock price; and the "post-squeeze" period (January 28 – February 3, 2021), referred to as *AS*, marked by heightened volatility and erratic price movements. This segmentation enables a comparative analysis of the model's predictive performance under contrasting market dynamics. Evaluating performance across both periods, each of similar duration, ensures that the effects of

Model	Class	Precision	Recall	F1-score
VADER	Positive	0.80	0.71	0.75
	Negative	0.12	0.31	0.18
	Neutral	0.52	0.41	0.46
finBERT- tone	Positive	0.91	0.38	0.54
	Negative	0.20	0.20	0.20
	Neutral	0.33	0.81	0.47
FinGPT v3	Positive	0.89	0.33	0.48
	Negative	0.40	0.22	0.29
	Neutral	0.30	0.86	0.45
Gemma-3	Positive	0.80	0.99	0.88
	Negative	0.79	0.42	0.55
	Neutral	0.86	0.46	0.60

Table 3: *The table presents per-class precision, recall, and F1-score for each sentiment class (Positive, Negative, Neutral) across different models (VADER, finBERT-tone, FinGPT v3, and Gemma-3) used for sentiment classification. Among the evaluated models, Gemma-3 consistently outperforms others across all sentiment classes, particularly in the Positive class with a high recall (0.99) and strong F1-score (0.88), indicating excellent sensitivity and balanced performance. While VADER shows moderate balance in performance, it struggles significantly with Negative sentiments (F1-score: 0.18). finBERT-tone and FinGPT v3 achieve high recall for Neutral class but suffer from low precision, resulting in moderate F1-scores. In contrast, Gemma-3 delivers more stable and higher scores across the board, demonstrating its robustness in sentiment classification tasks.*

extreme market behavior—especially increased volatility— are adequately captured and assessed.

Figure 6 presents the results of the first stage, where only stock price data were utilized. A 12-step sliding window and a one-hour gap were employed in the forecasting setup. During the *DS* period, the model achieved a Mean Squared Error (MSE) of 192.122 ± 36.020, Mean Absolute Error (MAE) of 7.668 ± 0.739, and a coefficient of determination (R^2) of 0.727 ± 0.051. In contrast, during the *AS* period, model performance deteriorated significantly, yielding an MSE of 343.397 ± 110.419, MAE of 12.643 ± 1.970, and R^2 of 0.317 ± 0.220. This decline in predictive accuracy in the post-squeeze period can largely be attributed to the elevated market volatility and the model's prior exposure primarily to relatively stable price patterns during training.

In the second stage of the analysis, Reddit activity data was integrated into the predictive framework, maintaining the 5-minute aggregation interval. One feature was extracted from Reddit discussions: the total number of comments received during the same interval (volume). This feature was incorporated alongside stock price data as input to forecast GME's closing price at the 5-minute level over the previously defined time periods. During the *DS* period, the inclusion of Reddit activity significantly enhanced model performance. The non-autoregressive model achieved an MSE

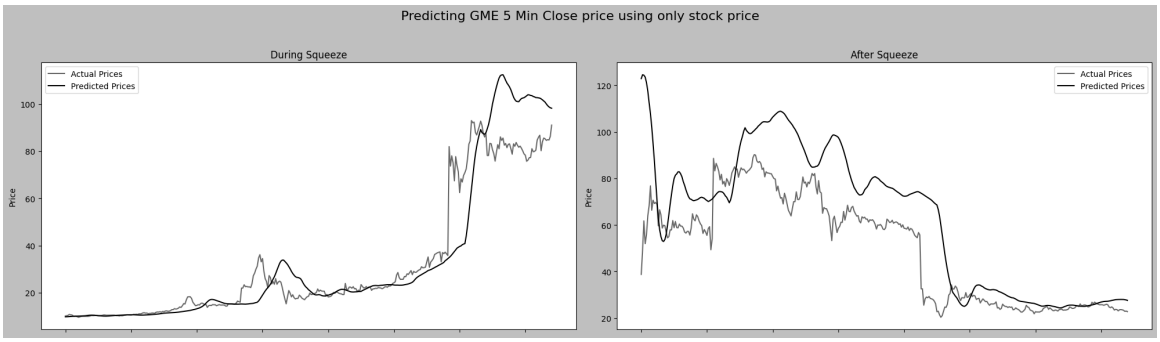

Figure 6: *Stage 1: Predicted vs. Actual GME 5-Minute Close Prices During and After the Short Squeeze : The plots compare the predicted prices (black line) and actual prices (gray line) of GameStop (GME) at 5-minute intervals. The left panel shows the period during squeeze, i.e. (Jan 20-27), while the right panel covers the period after the squeeze (Jan 28- 3). Predictions were generated using a model based solely on past stock prices taking one hour of data as input to forecast a price one hour ahead the end point. The predicted prices have been shifted 12 time steps for visual clarity.*

of 161.570 ± 25.610, MAE of 7.114 ± 0.622, and R^2 of 0.771 ± 0.036. Notably, while performance during *AS* period improved relative to the first stage— achieving an MSE of 254.253 ± 58.900, MAE of 10.933 ± 1.299, and R^2 of 0.494 ± 0.117 — it still lagged behind the model's performance in the *DS* period due to price fluctuations and heightened Reddit activity during this period. Overall, these results underscore the predictive value of Reddit engagement metrics, suggesting that real-time social media activity may contribute meaningfully to the short-term dynamics of stock prices. Figure 7 presents the second-stage results, highlighting consistent performance improvements across both market conditions. Replacing total comment counts with scores (max scores) yields similar results. (Appendix A.3)

In the third stage, the model incorporated both stock price information and NET sentiment features derived from Gemma-3, aggregated at a 5-minute level with a one-hour lag. This sentiment score was obtained from Reddit discussions and reflects the overall polarity of user posts. The multivariate input was again used to predict GME's 5-minute closing prices across the same two periods.During the *DS* period, the non-autoregressive model achieved improved performance from Stage 2, best so far, with an MSE of 134.755 ± 14.794, MAE of 6.558 ± 0.422, and R^2 of 0.808 ± 0.021. In the *AS* period the model achieved its best performance to date, yielding an MSE of 195.691 ± 27.485, MAE of 9.430 ± 0.769, and R^2 of 0.610 ± 0.055. These results, presented in Figure 8, highlight the added predictive value of sentiment-based features, particularly in capturing investor sentiment and behavioral signals that are not reflected in price or activity data alone.

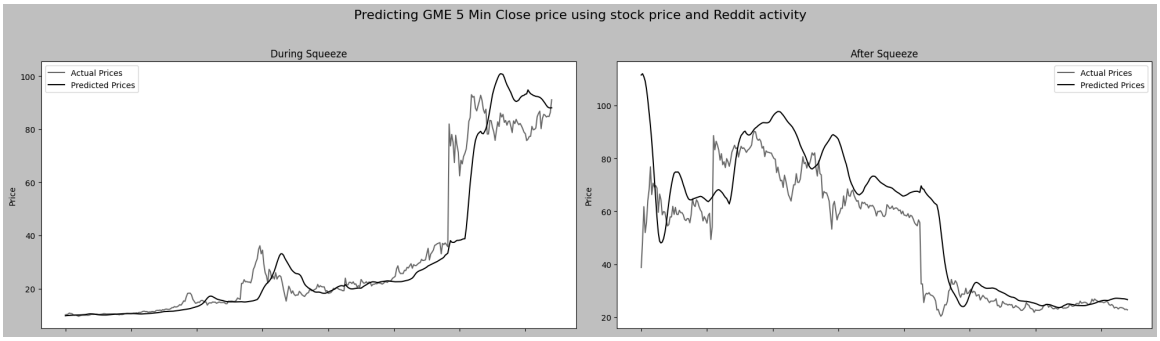

Figure 7: *Stage2: Predicted vs. Actual GME 5-Minute Close Prices During and After the Short Squeeze : The plots compare the predicted prices (black line) and actual prices (gray line) of GameStop (GME) at 5-minute intervals. The left panel shows the period during squeeze, i.e. (Jan 20-27), while the right panel covers the period after the squeeze (Jan 28- 3). Predictions were generated using a model that first ingested one hour of historical stock prices, followed by one hour of Reddit activity data, to forecast the next price point. The predicted prices have been shifted 12 time steps for visual clarity.*

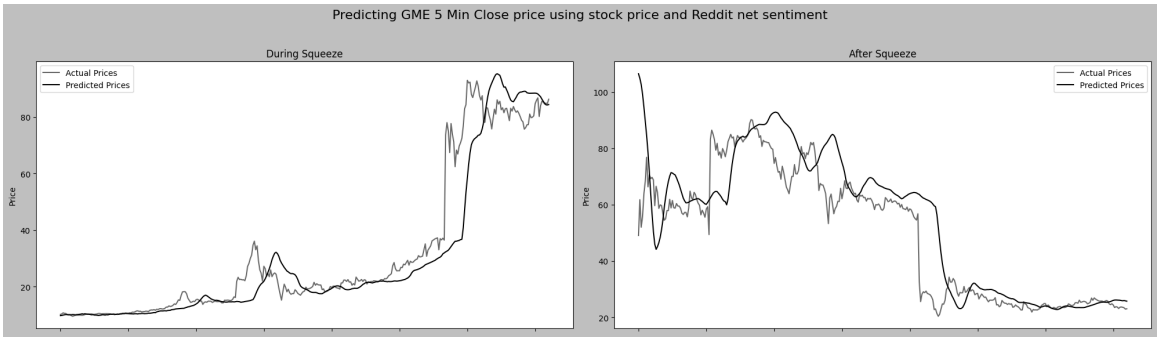

Figure 8: *Stage3: Predicted vs. Actual GME 5-Minute Close Prices During and After the Short Squeeze : The plots compare the predicted prices (black line) and actual prices (gray line) of GameStop (GME) at 5-minute intervals. The left panel shows the period during squeeze, i.e. (Jan 20-27), while the right panel covers the period after the squeeze (Jan 28- 3). Predictions were generated using a model that first ingested one hour of historical stock prices, followed by one hour of Reddit net sentiment data, to forecast the next price point. The predicted prices have been shifted 12 time steps for visual clarity.*

4.2 Discussion

The findings align closely with Kim et al. (2023), who demonstrate that "coordinated social informedness" among retail investors significantly influenced GME's price during the short squeeze. The above results, which show improved predictive performance with Reddit activity and sentiment data, strengthen the argument that social systems — not just individual motives—drive speculative asset dynamics.

The improvements in model performance observed in Stage 2 also align with theories of herding behavior, wherein investors mimic the actions of others rather than rely on independent assessments. During the squeeze period, the predictive value of Reddit activity features—such as the number of comments and scores — indicates that collective investor engagement contributed to synchronized trading behavior. The model's enhanced accuracy suggests that these activity metrics captured retail investors' coordinated responses to popular narratives circulating on social platforms. This herding dynamic is particularly pronounced in environments characterized by uncertainty and limited access to fundamental valuation metrics.

Further theoretical support is drawn from the Information Cascade Theory, Bikhchandani et al. (1998), which explains how individuals, observing the actions of others, may disregard their own information in favor of the dominant observed behavior. The temporal structure of Reddit activity, where a few high-scoring posts and viral discussions preceded large-scale trading surges, exemplifies this mechanism. The predictive power of early social activity signals (captured through one-hour lag windows) in Stage 2 confirms that Reddit discourse may have initiated cascades that drove collective retail investment behaviors. This phenomenon reinforces the idea that investor decision-making during speculative events can propagate through networks via informational mimicry rather than independent analysis.

From a behavioral standpoint, the reduced predictive performance during the AS period in Stage 2, despite the inclusion of Reddit activity features, can be attributed to a breakdown in the coherence and informativeness of crowd behavior. During the DS period, Reddit discussions were likely more unified and goal-oriented, revolving around a collective objective to drive up GME's stock price through a coordinated short squeeze. This alignment created strong, directional signals that were easier for the model to capture and learn from. In contrast, the AS period was marked by fragmented sentiment, emotional volatility, and divergent decision-making—conditions that weaken the predictive value of raw engagement metrics. The sentiment on Reddit likely became fragmented and reactionary—shifting from optimism and unity to uncertainty, profit-taking, fear of loss, and confusion. The content and engagement metrics may have remained high, but the underlying sentiment and intent became more volatile and noisy, reducing the signal quality for prediction.This underscores the limitations of relying on engagement metrics alone in high-volatility phases and the importance of incorporating sentiment-level signals, as explored in Stage 3.

Consistent with the above argument, the model achieved its high performance during the squeeze period (DS) in Stage 3. The NET sentiment, which was predominantly positive during this phase (as shown in Figure 8), contributed to heightened market enthusiasm and momentum, providing strong predictive and directional signals. In contrast, the persistence of positive sentiment following the imposition of trading restrictions—evident in the "hold the line" narrative—led to a temporary decoupling between sentiment and price action. This lag in sentiment adjustment reduced the model's predictive power in the initial phase of the post-squeeze (AS) period. Figure 4 shows continued bullishness or optimism for GME even after the restrictions in trading by brokerage platforms.

4.3 Placebo tests

The study assumes that Reddit activity, particularly within the r/WSB subreddit, played a significant role in the GameStop (GME) stock bubble and the associated volatility. This claim is supported by previous studies, such as Fernandez-Perez et al. (2024), Malz (2021), Long et al. (2023). To further investigate this claim, Granger Causality tests were performed on the stock returns of stocks similar to GameStop, based on the past six months of stock data prior to the peak of the GME bubble in January (i.e., July to December). According to the Bloomberg Industry Classification System (BICS), seven stocks were identified: EVGO, LESL, EYE, FCFS, FRPT, SBH, and ODP. The BICS offers a comprehensive and granular classification system for publicly traded and private companies, as well as bonds, enhancing the accuracy of financial analysis and indexing Phillips and Ormsby (2016).

Additionally, stocks sharing the same first two digits of the Standard Industrial Classification (SIC) code as GME, corresponding to the Retail Trade Division, were considered. Based on Bloomberg Terminal data, this process yielded an additional seven stocks: MRDH, CONNQ, SVSN, CFGX, HGGG, KLGG, and BBY.

For the matching process, market capitalization (Mcap), trading volume, and volatility were used as criteria. Market capitalization was calculated as the product of trading volume and the average of the opening and closing prices:

$$\text{Mcap} = \text{Volume} \times (0.5 \times (\text{Open} + \text{Close}))$$

The daily return is calculated by the percentage change in the closing price from one day to the next:

$$\text{Return} = \frac{\text{Close}_t - \text{Close}_{t-1}}{\text{Close}_{t-1}}$$

The intraday volatility is calculated using the Garman-Klass formula:

$$\text{Volatility} = \sqrt{0.5 \times (\log(\text{High}) - \log(\text{Low}))^2 - (2\log(2) - 1) \times (\log(\text{Close}) - \log(\text{Open}))^2}$$

The matching process involved calculating the smoothed values (using a 3-day moving average) for Mcap, Return, and Volatility for both GME and the other stocks. The total absolute differences between GME and each of the other stocks were summed, with the top three closest stocks identified as (ODP Corp) ODP, (Best Buy and Co) BBY, and (Firstcash Holdings Inc) FCFS. Analysis also includes the Russell 2000 Index, of which GME is a constituent.

Finally, pairwise Granger Causality tests were conducted at five min and the daily level to determine

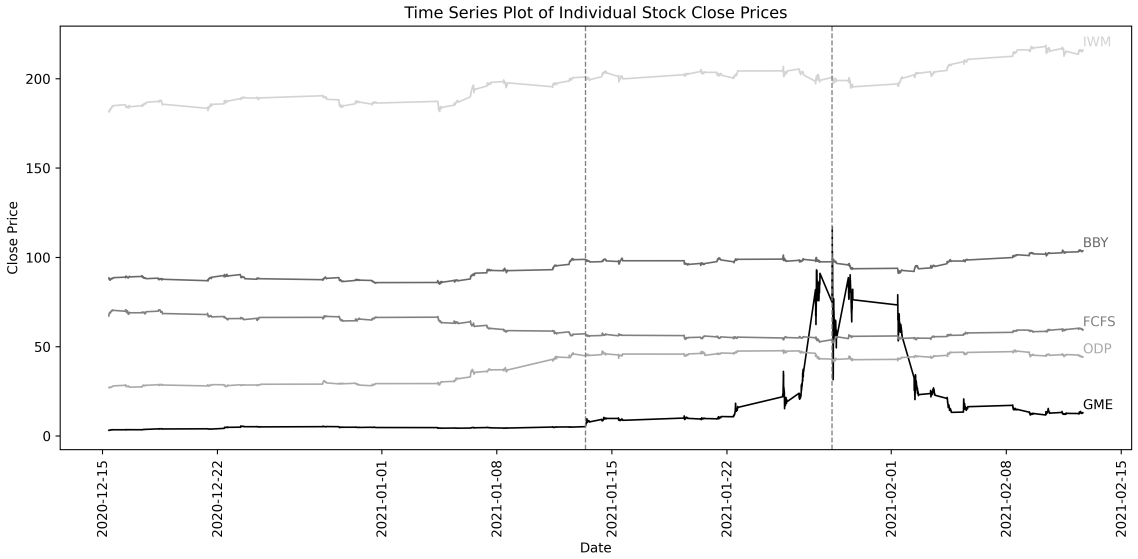

Figure 9: *Daily Closing Prices of GME-Matched Stocks Based on Propensity Score Matching: This figure displays the daily closing prices of stocks matched to GameStop (GME) using propensity score matching. The matching was based on smoothed 3-day moving averages of market capitalization, return, and volatility. The IWM Russell 2000 Index, of which GME is a constituent, is also shown for reference.*

whether the sum of comments and maximum scores in r/wallstreetbets Granger cause stock returns and vice versa. After performing the (Kwiatkowski-Phillips-Schmidt-Shin) KPSS and (Augmented Dickey-Fuller) ADF tests, it was found that the closing price, sum of comments, and scores were non-stationary. As a result, first differences were taken to make them stationary. Since close-to-close returns inherently represent first differences of closing prices,Granger tests were conducted on stock returns. All returns were found to be stationary.

Table 4 presents the Granger causality results where Reddit activity (comments and scores) are considered as a predictor for stock returns and vice versa. The table highlights instances where Reddit activity Granger-causes stock returns, with significance levels indicated for each relationship.

Stocks	sum comments		max score	
	Daily	5 min	Daily	5 min
GME_{ret}	0.003 ***	0.000 ***	0.004 ***	0.010 **
IWM_{ret}	0.051 *	0.082	0.100	0.030 **
ODP_{ret}	0.819	0.001 ***	0.082 *	0.362
$FCFS_{ret}$	0.046 **	0.175	0.705	0.837
BBY_{ret}	0.101	0.010	0.001 ***	0.228

Table 4: *This table presents the Granger causality results based on stock returns, where the null hypothesis assumes that the column variable does not Granger cause the row variable. Three alternative stocks were chosen based on propensity matching, taking into account factors like market capitalization, trading volume, and volatility to match with GameStop stock. These alternatives, ranked in descending order of matching scores, include ODP (ODP Corp), BBY (Best Buy Co.), FCFS (Firstcash Holdings Inc.), along with the IWM Russell 2000 index. To keep it concise, only the first column of the Granger causality matrix is shown, focusing on how Reddit activity influences stock returns. Data at both the five-minute and daily frequencies were used, covering the period from December 15, 2020, to February 28, 2021. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level*

As observed, the results show statistical significance for GameStop (GME) returns, but not for the other stocks. The results also hold in the reverse direction of causality: GME returns significantly Granger-cause Reddit activity — both comment volume and max scores ($p = 0.000^{***}$) — at both the 5-minute and daily levels. This bidirectional relationship is observed exclusively for GME and not for the other stocks. Notably, the IWM returns are marginally significant at the daily level, but this significance is not consistent at the five-minute frequency. This discrepancy may be attributed to heightened concerns related to the COVID-19 pandemic and the resulting volatility in the oil market at the same time as the squeeze, which affected both the market index (SPX) and the IWM index.

5 Conclusion

This paper examines the predictive power of Reddit-derived sentiment on GameStop's intraday price movements during a period of extreme market volatility. Focusing on posts from the r/WallStreetBets subreddit between December 15, 2020, and February 15, 2021, it introduces a novel framework for extracting and evaluating sentiment using prompt-based natural language processing, with sentiment scores validated through a human-centered approach. The methodology offers a robust tool for researchers analyzing investment-related sentiment from Reddit and other social media platforms. Furthermore, its adaptability extends beyond finance, demonstrating broader applicability for sentiment analysis in other domains.

This paper also contributes to the understanding of the meme stock phenomenon by demon-

strating that social media discussions and investor sentiment possess predictive power over stock price movements, particularly in volatile market conditions—beyond what is captured by price data alone. While prior literature has acknowledged the role of sentiment during the GameStop short squeeze, this study offers a more robust and accurate methodology for extracting sentiment and provides stronger empirical evidence of how coordinated retail investor behavior can influence market dynamics. These findings are particularly relevant to policymakers, academics, and both institutional and retail investors, as they underscore the emergence of social media as a de facto financial advisory platform—albeit one that operates without regulatory safeguards, especially during market downturns.

The results identify predictive power of Reddit activity and discussion popularity —measured by the total number of comments received on a post and max net score of a post (difference in upvotes and downvotes) on 5-minute interval GME closing prices. Additionally, retail investor sentiment aggregated at 5-minute intervals exhibits even stronger predictive power, as evidenced by higher R^2 values and lower mean absolute errors. Notably, temporal analysis shows that these relationships are more robust during bullish market phases, where positive sentiment aligns with upward price movements. In contrast, during bearish periods, the predictive influence of sentiment weakens— possibly due to a lag in sentiment adjustment or the persistence of bullish sentiment among retail investors despite declining prices, leading to a decoupling between sentiment and market behavior. These insights carry practical implications: for financial practitioners, the integration of social media sentiment into forecasting models can enhance risk management, particularly in meme-driven or highly volatile asset classes; for regulators, sentiment analytics may serve as an early-warning mechanism for detecting speculative bubbles or episodes of irrational exuberance within digital trading environments.

Directions for future research While this study focuses on sentiment analysis, it does not delve into more nuanced emotional states, which are finer-grained and may offer deeper insights into investor behavior. In particular, the persistence or "stickiness" of certain emotions—such as hope, fear, or euphoria—during periods of price decline remains unexplored. Future research could benefit from advancements in natural language processing that enable more precise emotion classification. Additionally, this analysis is limited to the case of GameStop during a highly volatile, crisis-like period. Further studies could investigate whether similar predictive relationships hold in more stable, non-bubble market conditions and examine the broader role social media plays in influencing investor sentiment across different contexts. Also studies can extend the analysis by considering other social and multimedia platforms like Facebook and other stocks using similar explainable AI models.

References

Khaled A Althelaya, El-Sayed M El-Alfy, and Salahadin Mohammed. Evaluation of bidirectional lstm for short-and long-term stock market prediction. In *2018 9th international conference on information and communication systems (ICICS)*, pages 151–156. IEEE, 2018.

Anamika Anamika and Sowmya Subramaniam. Do news headlines matter in the cryptocurrency market? *Applied Economics*, 54(54):6322–6338, 2022.

Abhinav Anand and Jalaj Pathak. The role of reddit in the gamestop short squeeze. *Economics Letters*, 211:110249, 2022.

Abdulahi D Athur. *Effect of behavioural biases on investment decisions of individual investors in Kenya*. PhD thesis, University of Nairobi, 2014.

Adam Atkins, Mahesan Niranjan, and Enrico Gerding. Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science*, 4(2):120–137, 2018.

Gautam Banerji, Julia Chung, and Christopher McCabe. Gamestop mania reveals power shift on wall street—and the pros are reeling. *The Wall Street Journal*, 2021.

Brad M Barber and Terrance Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2): 785–818, 2008.

Utpal Bhattacharya, Neal Galpin, Rina Ray, and Xiaoyun Yu. The role of the media in the internet ipo bubble. *Journal of Financial and Quantitative Analysis*, 44(3):657–682, 2009.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of economic perspectives*, 12(3):151–170, 1998.

Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8, 2011.

Gareth Campbell, John D Turner, and Clive B Walker. The role of the media in a bubble. *Explorations in Economic History*, 49(4):461–481, 2012.

Mbonigaba Celestin and N Vanitha. How social media is disrupting traditional finance: The rise of retail investors. In *7th International Conference on Modern Research Trends in Arts, Science, Engineering & Technology*, pages 183–191, 2021.

Usman W Chohan. Yolo capitalism. 2022.

J Anthony Cookson, Joseph E Engelberg, and William Mullins. Echo chambers. *The Review of Financial Studies*, 36(2):450–500, 2023.

Shaen Corbet, Charles Larkin, Brian M Lucey, Andrew Meegan, and Larisa Yarovaya. The impact of macroeconomic news on bitcoin returns. *The European Journal of Finance*, 26(14):1396–1416, 2020.

Shaen Corbet, Greg Hou, Yang Hu, and Les Oxley. We reddit in a forum: The influence of messaging boards on firm stability. *Available at SSRN 3776445*, 2021.

Hang Dong and Javier Gil-Bazo. Sentiment stocks. *International Review of Financial Analysis*, 72: 101573, 2020.

Alexander Dyck and Luigi Zingales. The bubble and the media. *Corporate governance and capital flows in a global economy*, 3:83–104, 2003.

Adrian Fernandez-Perez, Ivan Indriawan, and Marta Khomyn. Emotions and stock returns during the gamestop bubble. *Financial Review*, 2024.

Antonino Ferraro and Giancarlo Sperlì. How does user-generated content on social media affect stock predictions? a case study on gamestop. *Online Social Networks and Media*, 43:100293, 2024.

Henry David Gale. " buy gamestop!": The need to rethink the approach to market manipulation in a wallstreetbets world. *Iowa L. Rev.*, 108:1923, 2022.

Diego Garcia. Sentiment during recessions. *The journal of finance*, 68(3):1267–1300, 2013.

Chen Gu and Alexander Kurov. Informational role of social media: Evidence from twitter sentiment. *Journal of Banking & Finance*, 121:105969, 2020.

Danqi Hu, Charles M Jones, Siyang Li, Valerie Zhang, and Xiaoyan Zhang. The rise of reddit: How social media affects retail investors and short-sellers' roles in price discovery. *Available at SSRN 3807655*, 2021.

Kathryn Kadous, Molly Mercer, and Yuepin Zhou. Why do investors rely on low-quality investment advice? experimental evidence from social media platforms. *Behavioral Research in Accounting*, pages 1–19, 2022.

Gillian Kant, Ivan Zhelyazkov, Anton Thielmann, Christoph Weisser, Michael Schlee, Christoph Ehrling, Benjamin Säfken, and Thomas Kneib. One-way ticket to the moon? an nlp-based insight on the phenomenon of small-scale neo-broker trading. *Social Network Analysis and Mining*, 14 (1):121, 2024.

Muskan Kaur, Taruna Kalra, Sakshi Malik, and Anuj Kapoor. Significance of social networking media for influencing the investor behaviour in stock market. *Digital India: Reflections and Practice*, pages 83–98, 2018.

Kwansoo Kim, Sang-Yong Tom Lee, and Robert J Kauffman. Social informedness and investor sentiment in the gamestop short squeeze. *Electronic Markets*, 33(1):23, 2023.

Brandon Kochkodin. How wallstreetbets pushed gamestop shares to the moon. *Bloomberg*, 2021.

Satish Kumar and Nisha Goyal. Evidence on rationality and behavioural biases in investment decision making. *Qualitative Research in Financial Markets*, 8(4):270–287, 2016.

Jennifer S Lerner and Dacher Keltner. Fear, anger, and risk. *Journal of personality and social psychology*, 81(1):146, 2001.

Chao Liang, Linchun Tang, Yan Li, and Yu Wei. Which sentiment index is more informative to forecast stock market volatility? evidence from china. *International Review of Financial Analysis*, 71:101552, 2020.

Suwan Long, Brian Lucey, Ying Xie, and Larisa Yarovaya. "i just like the stock": The role of reddit sentiment in the gamestop share rally. *Financial Review*, 58(1):19–37, 2023.

Štefan Lyócsa, Eduard Baumöhl, and Tomáš Vỳrost. Yolo trading: Riding with the herd during the gamestop episode. *Finance Research Letters*, 46:102359, 2022.

Geetika Madaan and Sanjeet Singh. An analysis of behavioral biases in investment decision-making. *International Journal of Financial Research*, 10(4):55–67, 2019.

Allan M Malz. The gamestop episode: What happened and what does it mean? *Journal of Applied Corporate Finance*, 33(4):87–97, 2021.

Anna Mancini, Antonio Desiderio, Riccardo Di Clemente, and Giulio Cimini. Self-induced consensus of reddit users to characterise the gamestop short squeeze. *Scientific reports*, 12(1):13780, 2022.

Terrance Odean. Are investors reluctant to realize their losses? *The Journal of finance*, 53(5): 1775–1798, 1998.

Ryan L Phillips and Rita Ormsby. Industry classification schemes: An analysis and review. *Journal of Business & Finance Librarianship*, 21(1):1–25, 2016.

Nathaniel Popper and Kellen Browning. The 'roaring kitty' rally: How a reddit user and his friends roiled the markets. *The New York Times*, January 29 2021.

Premium Puzzle, SHLOMO BENARTZI, and RICHARD H THALER. Myopic loss aversion and the equity. *The Quarterly Journal of Economics*, 110(1):73–92, 1995.

Raja Rehan and Imran Umer. Behavioural biases and investor decisions. *Market Forces*, 12(2), 2017.

Thomas Renault. Intraday online investor sentiment and return patterns in the us stock market. *Journal of Banking & Finance*, 84:25–40, 2017.

Kristina M Rennekamp and Patrick D Witz. Linguistic formality and audience engagement: investors' reactions to characteristics of social media disclosures. *Contemporary Accounting Research*, 38(3):1748–1781, 2021.

Maik Schmeling. Investor sentiment and stock returns: Some international evidence. *Journal of empirical finance*, 16(3):394–408, 2009.

Sreelekshmy Selvin, R Vinayakumar, EA Gopalakrishnan, Vijay Krishna Menon, and KP Soman. Stock price prediction using lstm, rnn and cnn-sliding window model. In *2017 international conference on advances in computing, communications and informatics (icacci)*, pages 1643–1647. IEEE, 2017.

Valentina Semenova and Julian Winkler. Reddit's self-organised bull runs: Social contagion and asset prices. 2021.

Robert J Shiller. Market volatility and investor behavior. *The American Economic Review*, 80(2): 58–62, 1990.

Andrei Shleifer and Lawrence H Summers. The noise trader approach to finance. *Journal of Economic perspectives*, 4(2):19–33, 1990.

Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. A comparison of arima and lstm in forecasting time series. In *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, pages 1394–1401. Ieee, 2018.

Kacey Tollefson. Coming of age: Young investors and the rise in riskier investments. *RANGE: Undergraduate Research Journal (2023)*, 2023.

Charlie Wang and Ben Luo. Predicting $GME stock price movement using sentiment from Reddit r/wallstreetbets. In Chung-Chi Chen, Hen-Hsen Huang, Hiroya Takamura, and Hsin-Hsi Chen, editors, *Proceedings of the Third Workshop on Financial Technology and Natural Language Processing*, pages 22–30, Online, 19 August 2021. -. URL `https://aclanthology.org/2021.finnlp-1.4/`.

Sonja Warkulat and Matthias Pelster. Social media attention and retail investor behavior: Evidence from r/wallstreetbets. *International Review of Financial Analysis*, 96:103721, 2024.

H Denis Wu, Michael W McCracken, and Shinichi Saito. Economic communication in the 'lost decade' news coverage and the japanese recession. *Gazette (Leiden, Netherlands)*, 66(2):133–149, 2004.

Yi Yang, Mark Christopher Siy UY, and Allen Huang. Finbert: A pretrained language model for financial communications, 2020.

Mouna Youssef and Sami Sobhi Waked. Herding behavior in the cryptocurrency market during covid-19 pandemic: The role of media coverage. *The North American Journal of Economics and Finance*, 62:101752, 2022.

Boyu Zhang, Hongyang Yang, and Xiao-Yang Liu. Instruct-fingpt: Financial sentiment analysis by instruction tuning of general-purpose large language models. *FinLLM Symposium at IJCAI 2023*, 2023a.

Boyu Zhang, Hongyang Yang, tianyu Zhou, Ali Babar, and Xiao-Yang Liu. Enhancing financial sentiment analysis via retrieval augmented large language models. *ACM International Conference on AI in Finance (ICAIF)*, 2023b.

Shengyu Zhang, Linfeng Dong, Xiaoya Li, Sen Zhang, Xiaofei Sun, Shuhe Wang, Jiwei Li, Runyi Hu, Tianwei Zhang, Fei Wu, et al. Instruction tuning for large language models: A survey. *arXiv preprint arXiv:2308.10792*, 2023c.

Xiaolong Zheng, Hu Tian, Zhe Wan, Xiao Wang, Daniel Dajun Zeng, and Fei-Yue Wang. Game starts at gamestop: Characterizing the collective behaviors and social dynamics in the short squeeze episode. *IEEE Transactions on Computational Social Systems*, 9(1):45–58, 2021.

A Appendix

A.1 Comment and score spikes during squeeze

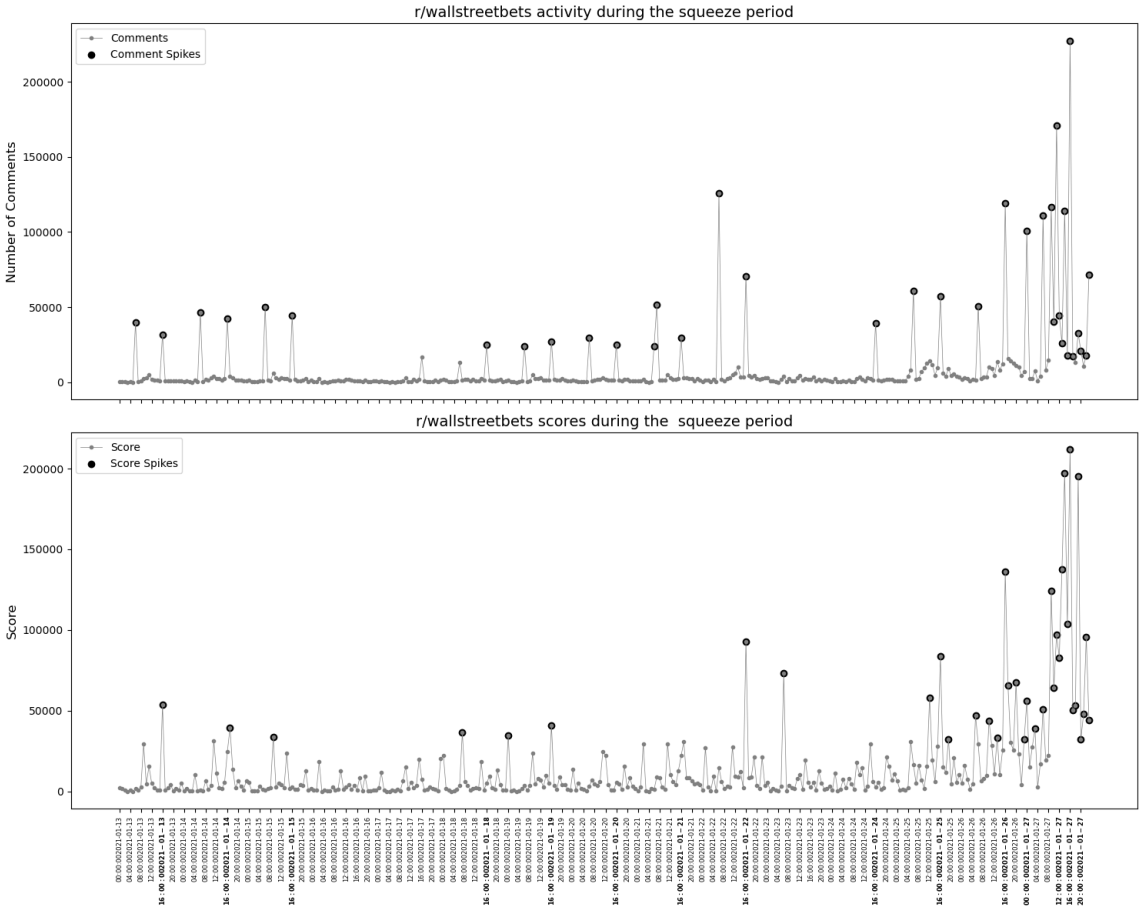

Figure A.1: *The figure illustrates the hourly distribution of comments and post scores on r/WallStreetBets during the squeeze period, spanning January 13 to January 27, 2021. Notably, the most significant spikes in both metrics occur during regular market hours (9:30 AM to 4:00 PM), with pronounced surges near market close. Activity on the forum intensifies markedly near the peak of the squeeze, reflecting heightened engagement and interest during this critical phase.*

A.2 Sentiment Analysis Prompt

```
You are an expert financial sentiment analyzer.

Analyze the following social media post, which is related to a stock.
```

Your task is to determine the sentiment of the post specifically towards the GME stock (Gamestop). Do not evaluate general mood, jokes, or unrelated content.

Please rate the sentiment towards GME only, across three categories:
- Positive sentiment (0 to 10): Where 0 = no positivity and 10 = extremely positive
- Negative sentiment (0 to 10): Where 0 = no negativity and 10 = extremely negative
- Neutral sentiment (0 to 10): Where 0 = no neutrality and 10 = extremely neutral, meaning factual, unemotional, unclear, or ambiguous in tone.

Important rules:
- Do not assume sentiment unless the text includes clear emotional or opinionated language.
- A post must include some evidence of attitude (e.g., cheering, complaints, frustration, sarcasm, bullish/bearish tone) to be rated as positive or negative.
- If the post simply mentions GME or other stocks without expressing any opinion or emotion, the sentiment is Neutral (Neutral = 10).
- If the user appears frustrated about a price drop, that implies they are emotionally invested and bullish, so rate Positive high, Negative low.
- If the post criticizes brokers, institutions, hedge funds in the context of the GME squeeze, rate it Positive, as it supports GME by opposing its opponents.
- If GME is not mentioned explicitly or indirectly, set Neutral = 10, Positive = 0, Negative = 0.
- If a post is sarcastic or ironic, infer sentiment based on common Reddit/WSB culture and context.

Only evaluate sentiment in relation to GME (Gamestop). You may treat aliases like Gamestonks, the squeeze, etc. as referring to GME when contextually clear.

List all stocks mentioned (even if not GME) in the post as strings.

Output must be in the following exact JSON format:

```
{
    "General Reasoning": "",
    "Why Positive": "",
    "Positive": <integer 0-10>,
    "Why Negative": "",
    "Negative": <integer 0-10>,
    "Why Neutral": "",
```

```
    "Neutral": <integer 0-10>,
    "Stocks Mentioned": [list of stock names as strings, comma separated]
}

Here is the text to analyze:
{text}
Assistant:
```

A.3 Gemma 3 (12B) Output

Post Text:

> GME people, hold strong! Don't sell yet. Power to the people! Let's eat cake! [cake emojis] [rocket emojis]

Gemma's Response:

```
{
  "General Reasoning": "The post explicitly encourages holding GME shares
      and uses celebratory language ('Power to the people!', 'Let's eat cake
      !', rockets). This demonstrates strong support and optimism for GME's
      future.",
  "Why Positive": "The post directly urges holders to 'hold strong' and not
      sell, coupled with enthusiastic phrases like 'Power to the people!'
      and 'Let's eat cake!' along with rocket emojis, all indicating a
      bullish sentiment.",
  "Positive": 9,
  "Why Negative": "There is no negative sentiment expressed towards GME.",
  "Negative": 0,
  "Why Neutral": "The post is not neutral; it expresses a clear opinion and
      encouragement.",
  "Neutral": 0,
  "Stocks Mentioned": ["GME"]
}
```

Post Text:

> According to S3 short interest in GME is down to 53% – it's just a ponzi now boys.

Gemma's Response:

```
{
  "General Reasoning": "The post references GME's short interest and then
      uses the term 'ponzi' which is highly derogatory and negative in a
      financial context. While acknowledging a decrease in short interest,
      the overall tone is critical and dismissive of GME's current state.",
```

```
  "Why Positive": "There is no positive sentiment expressed towards GME. The
      post is overwhelmingly negative.",
  "Positive": 0,
  "Why Negative": "The post directly calls GME a 'ponzi' which is a severe
     and negative accusation. This indicates a strong belief that GME's
     value is unsustainable and based on a fraudulent scheme.",
  "Negative": 9,
  "Why Neutral": "The mention of short interest data (53%) is factual but
     presented within a negative context, so it doesn't contribute to
     neutrality.",
  "Neutral": 1,
  "Stocks Mentioned": ["GME"]
}
```

Post Text:

Should I buy more GME or go for AMC or BB? New-ish to stocks and no idea what to do, looking for advice.

Gemma's Response:

```
{
  "General Reasoning": "The post is asking for advice on which stock to buy
     (GME, AMC, or BB). The user explicitly states they are 'new-ish to
     stocks and no idea what to do,' indicating a lack of strong conviction
      or opinion regarding GME specifically. The question itself is neutral
     , seeking guidance rather than expressing a positive or negative
     sentiment.",
  "Why Positive": "The user is considering buying GME, which could be
     interpreted as a slight positive inclination, but it's balanced by
     considering other stocks.",
  "Positive": 2,
  "Why Negative": "There's no negative sentiment expressed towards GME. The
     user is simply exploring options.",
  "Negative": 0,
  "Why Neutral": "The primary tone is one of seeking advice and exploring
     options. The user doesn't express any strong feelings about GME,
     making it largely neutral.",
  "Neutral": 9,
  "Stocks Mentioned": ["GME", "AMC", "BB"]
}
```

Annotator	Sentiment	Precision	Recall	F1 Score
A	Positive	0.92	0.91	0.91
	Negative	0.76	0.75	0.76
	Neutral	0.74	0.74	0.74
	Weighted Average	0.86	0.85	0.85
B	Positive	0.95	0.89	0.92
	Negative	0.68	0.85	0.75
	Neutral	0.77	0.82	0.79
	Weighted Average	0.88	0.87	0.87
C	Positive	0.93	0.90	0.91
	Negative	0.88	0.73	0.80
	Neutral	0.75	0.86	0.80
	Weighted Average	0.88	0.87	0.87
D	Positive	0.93	0.94	0.94
	Negative	0.66	0.83	0.74
	Neutral	0.85	0.75	0.80
	Weighted Average	0.89	0.88	0.88
E	Positive	0.98	0.95	0.96
	Negative	0.79	0.94	0.86
	Neutral	0.91	0.88	0.89
	Weighted Average	0.94	0.93	0.94

Table 5: Precision, Recall, and F1 Score per Sentiment Class for Each Annotator

A.4 Annotator Performance Evaluation Using Ground Truth

A.5 Stage 2 results using max scores instead of sum of comments at 5 min level

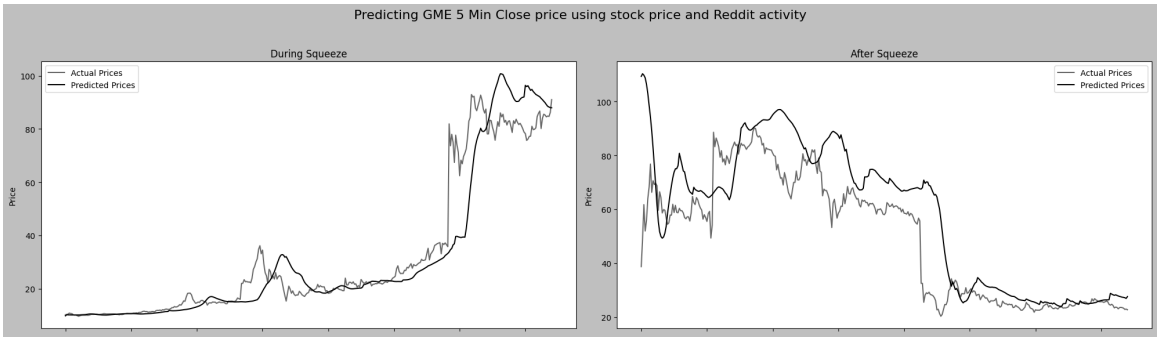

Figure A.3: *Stage 2: Predicted vs. Actual GME 5-Minute Closing Prices During and After the Short Squeeze — The plots compare predicted (black line) and actual (gray line) GME closing prices at 5-minute intervals. The left panel shows the squeeze period (Jan 20–27), and the right panel covers the post-squeeze period (Jan 28–Feb 3). Predictions were generated using a non-autoregressive model with one hour of historical prices and one hour of max score data as input, forecasting the next price point. Predicted values are shifted forward by 12 time steps for clarity. Performance metrics: during squeeze (Jan 21–28), MSE = 96.54, MAE = 5.63, R^2 = 0.863; after squeeze, MSE = 234.11, MAE = 10.42, R^2 = 0.642.*

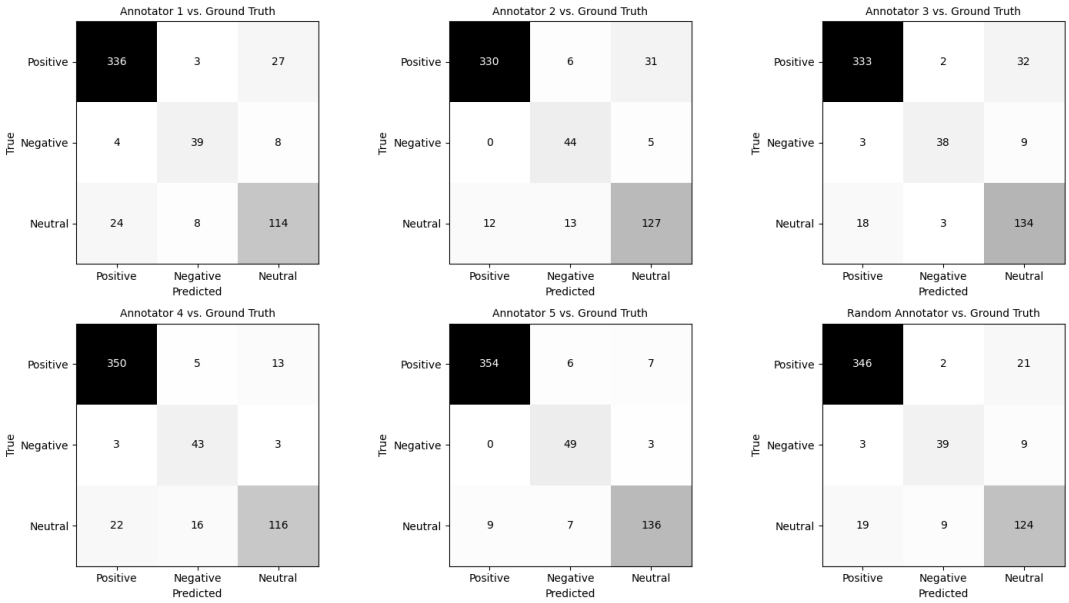

Figure A.2: *Confusion matrices comparing each annotator's sentiment labels to the ground truth. Rows indicate true labels; columns indicate predicted labels. The sixth matrix shows results from a randomly selected annotator per instance.*