

# Predicting \$GME Stock Price Movement Using Sentiment from Reddit r/wallstreetbets

Charlie Wang, Ben Luo

Department of Computer Science, University of Illinois at Chicago  
twang220@uic.edu, luo9@uic.edu

## Abstract

In early 2021, GameStop’s (NYSE: GME) stock underwent a prolonged period of rapid, drastic fluctuations in price resulting from a short squeeze orchestrated by the r/wallstreetbets community on Reddit. In this paper, we present our experiments and findings for predicting the direction of GME’s daily price movement using sentiment from r/wallstreetbets posts. We use VADER for automatic sentiment analysis combined with word2vec and BERT semantic representations, and experiment with a variety of classifier models, including both traditional ML and deep learning approaches.

## 1 Introduction

GameStop (NYSE: GME) is the world’s largest video game retailer, best known in North America for its ubiquitous brick-and-mortar retail locations. Starting in the late 2000’s, the rise of online retail and digital video game sales contributed to the company’s steady decline over the following decade.

However, in late January of 2021, the company’s stock price experienced a sudden and drastic reversal of fortune as amateur retail investors banded together online to buy GME shares en masse in a manic rush, driving up the share price and disrupting the short selling strategies of major Wall Street players. By mid February, GME had risen in price from a starting point of under \$20 USD to a pre-market peak of over \$500 USD per share [Orland, 2021]. As of this writing, the price continues to fluctuate wildly across this range.

According to Business Insider (2021), GME was the most shorted stock on the New York Stock Exchange as of January 2021. Short selling is an investment strategy that speculates on the decline of a stock’s price — i.e., someone who is selling a stock short stands to profit when the price of that stock declines over a certain period of time. A short squeeze occurs when, fearing losses due to upward stock price movement, short sellers rush to buy shares to cover their losses, thus further driving up demand for the stock and forcing its price upward even more. In this case, GME was the target of a short squeeze that appeared to be orchestrated in large part by members of Reddit’s r/wallstreetbets community<sup>1</sup> [Orland, 2021].

<sup>1</sup><https://www.reddit.com/r/wallstreetbets/>

The r/wallstreetbets subreddit began as a meme-heavy offshoot of more serious investment subreddits such as r/stocks, r/investing, or r/pennystocks. Compared to these other communities, r/wallstreetbets is distinguished by the high level of irony that permeates its discourse. On a surface level, members of this subreddit seem to approach investment with a somewhat nihilistic gambling mindset as opposed to the rational, strategic attitude typical of financial contexts. For example, r/wallstreetbets members often post screenshots of their investment portfolio performance showing the disastrous results of reckless speculation. In these threads, fellow users both commiserate and celebrate over these losses, terming them as *loss porn*.

Boasting a membership of over 10 million users as of May 2021, this subreddit has become a formidable community of small-time individual investors in its own right, earning occasional mentions in financial news and investment-related media even before its starring role in the GME frenzy. The community’s struggle against hedge funds, investment brokerages, and even the SEC has helped draw widespread media coverage to these events, which often depicts the situation as a David vs. Goliath underdog story [Orland, 2021].

These historic events surrounding GME are, in many ways, unprecedented and highly irregular. Thus, conventional methods for stock price prediction may generalize poorly to this particular situation. The principal novel contribution of this paper is our approach to data preprocessing, which is specifically aimed at capturing sentiment information from r/wallstreetbets text and was developed in response to linguistic peculiarities we observed in the data harvested from the subreddit.

Given the unique relationship between r/wallstreetbets and the GME turmoil, we hypothesize that general sentiment on r/wallstreetbets is correlated with GME price movement. We attempt to exploit this relationship by predicting the direction of GME’s daily net price movement using sentiment from posts on r/wallstreetbets.

## 2 Related Work

Securities and cryptocurrency market prediction is a popular problem, which is perhaps not surprising given the obvious financial incentive. Prediction using textual sentiment as an input feature is perhaps less explored than prediction primarily based on financial data and market indicators. Li et al. (2014) predict price for individual stocks based on sentiment from

financial news posts. Similarly, Prosky et al. (2018) predict stock price movement based on sentiment from news, additionally incorporating event extraction and entity recognition to enhance sentiment analysis.

However, user-generated text on social media platforms has been shown to differ significantly from standard English, such as that found in news articles [Baldwin *et al.*, 2013]. This presents different challenges to NLP that must be overcome in order to extract useful information from noisy social media text.

A number of past efforts predict stock and cryptocurrency price movement using sentiment information gathered from Twitter. Pagolu et al. (2016) find a strong correlation between Twitter sentiment and the movement of the Dow Jones Industrial Average index. Bing et al. (2014) examine the effect of Twitter sentiment on individual stocks. Kim et al. (2016) predict movement in the value of cryptocurrencies based on Twitter sentiment, and use statistical methods to demonstrate a causal relationship. Valencia et al. (2019) similarly predict cryptocurrency value using Twitter sentiment. These latter two papers employ VADER for automatic sentiment analysis, and include additional non-linguistic features alongside sentiment, such as topic count and reply count. We borrow extensively from this methodology.

To our knowledge, little prior work focuses on sentiment extracted from investment-related posts on Reddit in particular. Twitter text has been shown to differ stylistically in several important ways from more long-form online text, such as that found in forums like Reddit [Baldwin *et al.*, 2013]. Upon analyzing the collected data, we notice a number of characteristics of *r/wallstreetbets* language that present further obstacles to sentiment analysis compared to Twitter language (and even to language on Reddit outside of *r/wallstreetbets* and related subreddits), which we discuss in detail in Section 3.3. Our work most significantly differs from the above prior efforts in the source of our textual sentiment (i.e., *r/wallstreetbets* as opposed to news or Twitter) and the innovations in data preprocessing that are necessitated thereby.

## 3 Data

### 3.1 *r/wallstreetbets* posts

We leverage an existing corpus of *r/wallstreetbets* posts available on Kaggle.com<sup>2</sup>, consisting of all 44,293 posts from September 2020 through March 2021. We initially investigated the Pushshift API, but found that it was missing a significant amount of data over our target time period due to technical issues. We filter the Kaggle data by date to obtain a dataset totaling 35,726 posts from all weekdays (excluding trading holidays) from January 4<sup>th</sup>, 2021 – March 31<sup>st</sup>, 2021.

Apart from standard text preprocessing (e.g., removal of URLs, punctuation, user handle mentions, etc.), notable processing steps include concatenating the post title and body together, as well as replacing emojis with descriptive text tags sandwiched by delimiter symbols. This approach enables us to take advantage of word embeddings, which we discuss later in Section 4.2, to capture semantic information from emojis.

<sup>2</sup><https://www.kaggle.com/gpreda/reddit-wallstreetsbets-posts/>

Table 1: Distribution of *r/wallstreetbets* posts by direction of daily price movement

Price movement direction	# of posts
up	7583 (21.2%)
down	28152 (78.8%)

### 3.2 GME price data

We extract stock price data for the same time period (Jan. 4<sup>th</sup> – Mar. 31<sup>st</sup>) from the RapidAPI API for Yahoo! Finance<sup>3</sup>. We calculate net price movement for each trading day as (*close\_price* – *open\_price*), and from this we determine the direction of price movement (up or down) for each day.

This begs the question: how should posts made outside of trading hours be associated with stock price data? NYSE trading hours are limited to Monday – Friday, 9:30AM – 4:00PM Eastern Time. Limited trading and changes in stock price can take place after-hours. Prior work in stock prediction employs a variety of heuristics for associating events to time series data, dependent largely on the relationship between predictive features and the aspect of stock performance being predicted. If a post is made after market close, is it indicative of that day’s performance, or the price change between market close and the next day’s market open? At its root, the question is essentially whether there is a causal relationship between GME price movement and general discussion on *r/wallstreetbets*.

In this paper, we wish not to assume any such causal relationship, as any investigation of such is outside the scope of our work. However, short of discarding all posts made outside trading hours, any choice would seem to imply some kind of causal assumption. Considering this, we choose the easiest method to implement, which is to simply associate posts with the trading day on which they were made. This carries a soft assumption that on a given day, posts made after market close tend to discuss that day’s stock performance, while posts before market open tend to discuss the upcoming trading day.

Following the simple association scheme detailed above, Table 1 shows the distribution of *r/wallstreetbets* posts by direction of price movement.

### 3.3 Qualitative analysis

In our dataset, 52% of posts contain no body text. These are usually image or video posts, where the only readily parsable text is the post title. Images or videos may contain some text, but multimodal techniques would be needed in order to extract such text. We leave this as a possible direction for future work.

There are few explicit mentions of ‘\$GME’, ‘GME,’ or ‘GameStop’ among post titles and bodies — filtering for these three keywords leaves us with only approximately 1.1k posts out of the original 35k. While the topic of GME and a few other short squeeze targets (e.g., AMC, BB, and NOK) certainly dominate the most popular threads on *r/wallstreetbets* during this time period, users also continue to, as usual, discuss a wide variety of other securities in other posts.

Without sophisticated multimodal techniques combined with named entity recognition, it is difficult to identify the

<sup>3</sup><https://rapidapi.com/apidojo/api/yahoo-finance1>

Post Titles
<i>It's not about the money, it's about sending a message</i>  rocket   fullmoon   fullmoon
<i>If this isn't raw talent, then I don't know what is</i>  gemstone   raisedhand  <i>bought GME at \$493</i> <i>when its 52wk high was \$483</i>  rocket   rocket  <i>If I can hodl, so can u</i>  smilingfacewithsunglasses
<i>ALL YOU \$300+ BAGHOLDING RETARDS PUT YOUR AFFAIRS IN ORDER &amp; BE SURE</i> <i>YOUR BAGS ARE PACKED CUZ THE ROCKET AIN'T STOPPING. BUCKLE THE F*** UP</i>  rocket   rocket   rocket   gemstone   rasinghands   gorilla   briefcase
<i>Rocketing to the Moon tendies with mah ape brain</i>  gorilla   brain   rocket   rocket   rocket   crescentmoon

Figure 1: Example post titles. Emojis are replaced with text tokens (e.g., rocket|, |gorilla|)

many posts that are about GME but do not directly mention GME or associated entities in the text. Figure 2 is an example of one such post. Thus, rather than attempting to isolate sentiment toward GME or other particular entities, in this paper we examine the general sentiment expressed on r/wallstreetbets across all posts, regardless of topic.

Some peculiar rhetorical habits of r/wallstreetbets users may cause difficulties for uninitiated human sentiment annotators, to say nothing of automatic analyzers. We find many examples of esoteric slang and self-referential humor. In particular, the language of r/wallstreetbets users is characterized by heavy use of emojis, often in senses that are unique to the subreddit. These emojis often convey strong positive sentiment. Posters often use profanity and what might normally be construed as aggressive tone to express positive emotions. Examples of the above are illustrated in Figure 1.

We anticipate that these factors may present difficult obstacles to automatic sentiment analysis.

## 4 Methods

### 4.1 Automatic sentiment analysis

For automatic sentiment analysis, we turn to VADER, a popular general sentiment analyzer for social media text. VADER is a rule-based model that relies on a lexicon containing sentiment information (both polarity and intensity) for a wide range of tokens, including slang and emoticons, in a social media context. This approach lends itself well to lightweight implementation and has been shown to work well on typical social media text [Hutto and Gilbert, 2014]. Our choice to use VADER is inspired by related work [Kim *et al.*, 2016] [Valencia *et al.*, 2019].

However, due to the observed peculiarities of the language produced by r/wallstreetbets users discussed previously (Section 3.3), we anticipate difficulties arising from using off-the-shelf VADER.

To gauge the reliability of VADER’s sentiment analysis results on r/wallstreetbets text, we perform a preliminary annotation experiment wherein three human annotators, all of whom possess working familiarity with r/wallstreetbets content and investment lingo, each evaluate the same random sample of 100 post titles for majority sentiment (positive, negative, or neutral).

We observe average agreement between the three human annotators (Cohen’s Kappa) of  $\kappa = 0.7280$ , indicating middling to high agreement — a higher than anticipated result, given the subjective nature of the sentiment annotation task. On

the other hand, average agreement between human annotators and VADER is low at  $\kappa = 0.078$ . This suggests rather poor performance on the part of VADER, as we anticipated.

Caveats to this annotation experiment: First, this experiment solely examines post titles, as some posts contain long bodies, which would lead to a cumbersome annotation task. In many cases, this of course reduces the number of words and amount of context available. Second, due to time constraints, the data used in this evaluation is only partially preprocessed. Most notable here is the absence of emojis, which, as previously discussed, can convey especially strong sentiment on r/wsb. These factors likely impact annotation performance, both human and automatic.

We observe that VADER tends to overpredict neutral sentiment, especially for posts labeled as ‘positive’ by human annotators. Of the sample of 100 post titles, VADER scores 93/100 examples as being highest in neutral sentiment. Of those 93, 30 are labeled by all three human annotators as ‘positive.’

We speculate that this is due to a high proportion of out-of-vocabulary words (recall that post titles are often terse), which VADER generally scores as neutral, as well as in-vocabulary words with r/wallstreetbets-specific senses that convey different sentiment polarity or intensity than they would in a generic social media context. For example, in the post title in Figure 2, “Diamond Hands” is a popular r/wallstreetbets idiom that expresses determination, encouraging the reader to avoid the temptation to sell their GME shares, even in the face of declining share price or the allure of slight profit. Outside of r/wallstreetbets and similar communities, where this usually translates to positive sentiment, this phrase has little meaning.

The sentiment features we extract using VADER take the form of three normalized sentiment scores: positive, neutral, and negative. These are the features we use as inputs in our classification experiments.

### 4.2 Semantic embeddings

Noting poor classification performance from sentiment features alone, we opt to augment sentiment with semantic representations. In our experiments, we compare two different semantic representations as features for classification alongside sentiment: word2vec and pretrained BERT.

First, we generate word2vec embeddings<sup>4</sup> of 100 dimensions from our training set, ignoring any words that appear fewer than two times in the data, since these are likely to be

<sup>4</sup><https://radimrehurek.com/gensim/models/word2vec.html>

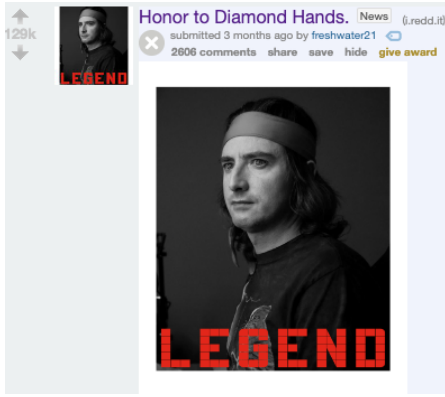


Figure 2: An example post. Pictured: Keith Gill a.k.a. RoaringKitty, widely regarded as a principal architect of the GME short squeeze.

typos. To produce a vector representation for the text of each post’s title + body, we simply average the word embeddings for all words in a post.

For comparison, we extract from a BERT model<sup>5</sup>, pretrained on BookCorpus<sup>6</sup> and Wikipedia text, the last dense layer. From this, we generate the output of the last hidden layer before the output classification layers. This produces an output vector of 768 dimensions to represent the text in each post.

### 4.3 Additional features

Beyond sentiment and semantic representations, we explore other features that could benefit the prediction of price movement. For additional textual features, we extract word count, stopword count, average word length, and emoji count. Next, we incorporate some metadata features not derived from text, namely, upvote score and number of comments. We suspect that these properties may indirectly indicate the level of excitement of the r/wallstreetbets community. We experiment with combining these additional features with the sentiment and semantic features.

## 5 Experiments

For our classification experiments, we employ a stratified train-test split of 80%/20% and a variety of classifier models: support vector machine (SVM or SVC), random forest (RF), decision tree (DT), and multilayer perceptron (MLP).

First, we perform a preliminary experiment using only sentiment to predict the direction of price movement. In our second experiment, we compare the performance of two different semantic representations as input features, as detailed in Section 4.2: word2vec and BERT. For the last experiment, we test combinations of sentiment, semantic representations, and the additional features from Section 4.3.

## 6 Results

### 6.1 Sentiment only

Table 2 shows the classification result using only VADER sentiment features. In the 'F1 up' column, which is the F1

<sup>5</sup><https://huggingface.co/bert-base-uncased/>

<sup>6</sup><https://yknzhu.wixsite.com/mbweb/>

score for the minority class, two classifiers have an F1 score of zero. These classifiers label all test examples as 'down,' showing that VADER sentiment alone is not suitable for the prediction of price movement.

### 6.2 Semantic representations

Following the classification with sentiment features only, we repeat the prior experiment with the addition of word2vec and BERT semantic representations as input features. The features are combined via simple vector concatenation. Table 3 displays results across different classifier models. Word2vec with the decision tree classifier has overall best accuracy, macro average F1 score, and minority class F1 score. However, the word2vec features also produce the worst minority class performance with the SVM classifier.

In addition to using BERT output as a feature, we also experiment with pretrained BERT as a classifier model (Table 3). On the BERT side, the pretrained BERT model ("Original") shows better performance than other classifiers in terms of F1 score (0.6419) and minority class F1 score (0.4580).

When introducing BERT embeddings to supplement VADER sentiment features, we observe a general strengthening of majority class performance and weakening of minority class performance. Overall, utilizing semantic representations as features generally improves performance, although not as much as word2vec.

### 6.3 Sentiment with semantic representations and additional features

Finally, we experiment with combining our two prior features (sentiment and semantic representations), again using simple vector concatenation, with additional features that might benefit performance (detailed previously in Section 4.3). Table 4 shows the results of different feature combinations. Here, "All" refers to VADER sentiment, and the additional features detailed in Section 4.3.

The first row of table 4 shows the result of different combinations of word2vec, sentiment, and extra features. We observe a large improvement after the addition of VADER sentiment. Here, we achieve the overall best result with 0.9905 accuracy and 0.9857 macro average F1 score. The random forest classifier also shows great improvement in this case. This improvement happens on both BERT and word2vec. We provide a visualization of the decision tree nodes (Appendix A) and find that the VADER vectors are included in every branch, which indicates that the sentiment information is actively involved in the decision process.

Adding all of the additional features did not yield significant improvement over sentiment and semantic embeddings.

## 7 Discussion & Future Work

From our results, we are unable to definitively demonstrate a strong relationship between sentiment and price movement. However, our results do show that semantic embeddings, which may capture sentiment information to some degree, are useful for this prediction task. Predicting based on a combination of sentiment and semantic embeddings produces overall

Table 2: VADER sentiment: preliminary classification

VADER				
	ACC	F1 average	F1 down	F1 up
<b>SVC</b>	<b>0.7882</b>	<b>0.4408</b>	<b>0.8815</b>	<b>0</b>
<b>RF</b>	0.777	0.4591	0.8738	0.0444
<b>DT</b>	0.774	0.4676	0.8715	0.0638
<b>MLP</b>	<b>0.7882</b>	<b>0.4408</b>	<b>0.8815</b>	<b>0</b>

Table 3: Semantic representations: word2vec vs. BERT. The "Original" row represents the result of pretrained BERT with the last dense layer output of one dimension as prediction.

	BERT				W2V				
	Accuracy	F1 average	F1 down	F1 up	Accuracy	F1 average	F1 down	F1 up	
<b>ORIGINAL</b>	0.7364	<b>0.6419</b>	0.8258	<b>0.458</b>					
<b>SVC</b>	0.7572	0.6281	0.8473	0.4089	<b>SVC</b>	0.7575	0.4743	0.8602	0.0884
<b>RF</b>	0.7466	0.6377	0.8363	0.4391	<b>RF</b>	0.7662	0.6498	0.8517	0.448
<b>DT</b>	0.7586	0.6243	0.849	0.3996	<b>DT</b>	<b>0.8147</b>	<b>0.6766</b>	<b>0.888</b>	<b>0.4653</b>
<b>MLP</b>	<b>0.7701</b>	0.6316	<b>0.8575</b>	0.4058	<b>MLP</b>	0.7742	0.6273	0.8613	0.3932

Table 4: Word embeddings with sentiment and additional features.

	W2V+VADER				W2V+ALL				
	Accuracy	F1 average	F1 down	F1 up	Accuracy	F1 average	F1 down	F1 up	
<b>SVC</b>	0.4617	0.448	0.535	0.3611	<b>SVC</b>	0.5085	0.4847	0.5954	0.3739
<b>RF</b>	0.8587	0.8228	0.9025	0.743	<b>RF</b>	0.8578	0.8246	0.901	0.7481
<b>DT</b>	<b>0.9905</b>	<b>0.9857</b>	<b>0.994</b>	<b>0.9775</b>	<b>DT</b>	<b>0.991</b>	<b>0.9866</b>	<b>0.9943</b>	<b>0.979</b>
<b>MLP</b>	0.7698	0.6085	0.8598	0.3572	<b>MLP</b>	0.6982	0.5732	0.8042	0.3422

	BERT+VADER				BERT+ALL				
	Accuracy	F1 average	F1 down	F1 up	Accuracy	F1 average	F1 down	F1 up	
<b>SVC</b>	0.7987	0.5363	0.8851	0.1875	<b>SVC</b>	0.8075	0.6702	0.883	0.4574
<b>RF</b>	0.7717	0.6643	0.8541	0.4746	<b>RF</b>	0.7648	0.6566	0.8494	0.4638
<b>DT</b>	<b>0.8222</b>	<b>0.6955</b>	<b>0.8919</b>	<b>0.499</b>	<b>DT</b>	<b>0.8223</b>	<b>0.6962</b>	<b>0.8927</b>	<b>0.4998</b>
<b>MLP</b>	0.7837	0.6297	0.8685	0.3909	<b>MLP</b>	0.7873	0.5379	0.8774	0.1983

best performance, so sentiment does appear to have some positive impact.

One limitation of this work is dataset quality stemming from the poor performance of automatic sentiment analy-

sis. As demonstrated previously, off-the-shelf VADER sentiment analysis produces disappointing results on text from r/wallstreetbets. From our brief subjective analysis, we speculate that the major contributing factors to this are:

- frequent use of sarcasm or irony, especially pejorative terms and profanity to convey positive sentiment
- a preponderance of out-of-vocabulary words, chiefly slang, investment jargon, and emojis
- non-standard usages of many in-vocabulary words specific to the subreddit or the investment domain

We suspect that improved sentiment analysis may yield further improvements in performance. We leave to future work the task of customizing the VADER sentiment lexicon in order to better accommodate r/wallstreetbets language. Our decision to replace unicode emojis with descriptive text tags is made with this future task in mind.

Alternatively, more sophisticated sentiment extraction approaches may bear investigation, such as that explored by Li and Lu (2017), who introduce scoped entity awareness to sentiment analysis. For reasons mentioned in Section 3.3, it is difficult to isolate posts by topic in r/wallstreetbets. Because of this, in this paper we examine general sentiment from the subreddit as a whole, with no regard to the specific object(s) of the sentiment expressed in a given post. Approaches that incorporate entity awareness may be better suited for evaluating sentiment across long texts, such as in the bodies of some r/wallstreetbets posts. For example, a user might express positive sentiment toward GME in part by using disparaging or defiant language directed toward Robinhood or the SEC, two entities that gained notoriety on the subreddit as the short squeeze progressed.

Our data is moderately imbalanced between the two classes, with only 21.2% of posts falling on days with upward net price movement. Resampling methods – undersampling, oversampling, and ensemble sampling – may improve performance on the minority class while maintaining similar performance for non-minority classes. [More, 2016].

In the intervening months since the peak of GME mania, several "meme stock"-focused offshoot subreddits have appeared, including r/gme and r/superstonk. Based on the premise that these subreddits emerged largely as spin-offs of r/wallstreetbets and are likely to share significant portions of r/wallstreetbets's user base, we speculate that going forward, these subreddits may be another useful source of training data for the task of predicting GME price movement as the saga continues to unfold. At the same time, sentiment from r/wallstreetbets posts may become a worse predictor for GME price movement over time as GME-related discussion no longer dominates discourse on the subreddit as it did during Q1 of 2021.

A more sophisticated heuristic for temporally associating market data to social media posts (discussed in Section 3.2) may allow for additional powerful approaches to be employed, including feedback neural networks, such as the GRU and LSTM models used by Minh et al. (2018).

Finally, the direction of change in stock price is a simplistic index for stock performance and likely fails to indicate much nuance in stock market activity, especially in an unusual short squeeze situation. Rather than framing this prediction task as a binary classification problem, future work might benefit from experimenting with more granular labels that correspond to different magnitudes of price movement. Additionally, the

findings of Prosky et al. (2018) suggest that sentiment from social media may, under some conditions, be a better predictor for other market indicators such as volume and liquidity than for price movement.

## References

- [Baldwin et al., 2013] Timothy Baldwin, Paul Cook, Marco Lui, Andrew MacKinlay, and Li Wang. How noisy social media text, how different social media sources? In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 356–364, 2013.
- [Bing et al., 2014] Li Bing, Keith C. C. Chan, and Carol Ou. Public sentiment analysis in twitter data for prediction of a company's stock price movements. In *Proceedings of the 2014 IEEE 11th International Conference on E-Business Engineering*, ICEBE '14, page 232–239, USA, 2014. IEEE Computer Society.
- [bus, 2021] The 10 most heavily-shorted stocks of January 2021, 2021.
- [Hutto and Gilbert, 2014] Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, 2014.
- [Kim et al., 2016] Young Bin Kim, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, and Chang Hun Kim. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PLoS one*, 11(8):e0161197, 2016.
- [Li and Lu, 2017] Hao Li and Wei Lu. Learning latent sentiment scopes for entity-level sentiment analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), Feb. 2017.
- [Li et al., 2014] Xiaodong Li, Haoran Xie, Li Chen, Jianping Wang, and Xiaotie Deng. News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, 69:14–23, 2014.
- [Minh et al., 2018] Dang Lien Minh, Abolghasem Sadeghi-Niaraki, Huynh Duc Huy, Kyungbok Min, and Hyeonjoon Moon. Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. *Ieee Access*, 6:55392–55404, 2018.
- [More, 2016] Ajinkya More. Survey of resampling techniques for improving classification performance in unbalanced datasets, 2016.
- [Orland, 2021] Kyle Orland. The complete moron's guide to gamestop's stock roller coaster, Jan 2021.
- [Pagolu et al., 2016] Venkata Sasank Pagolu, Kamal Nayan Reddy, Ganapati Panda, and Babita Majhi. Sentiment analysis of twitter data for predicting stock market movements. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pages 1345–1350, 2016.

[Prosky *et al.*, 2018] Jordan Prosky, Xingyou Song, Andrew Tan, and Michael Zhao. Sentiment predictability for stocks, 2018.

[Valencia *et al.*, 2019] Franco Valencia, Alfonso Gómez-Espinosa, and Benjamín Valdés-Aguirre. Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. *Entropy*, 21(6):589, 2019.

## **A Appendix: Visualization of decision tree classifier**

The following pages contain a visualization of the decision tree nodes from the experiment in Table 4.

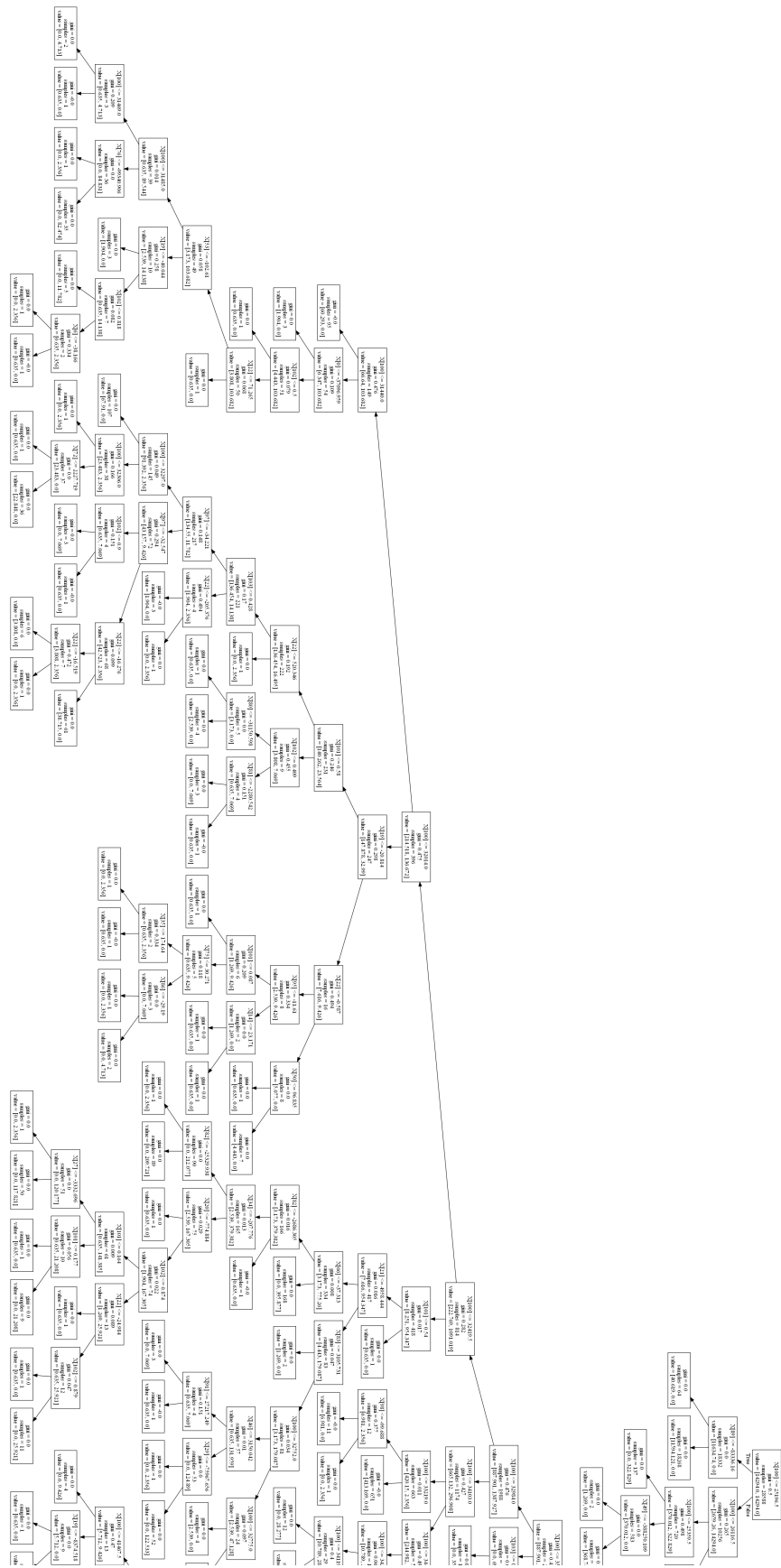


Figure 3: visualization of decision tree part 1 of 2



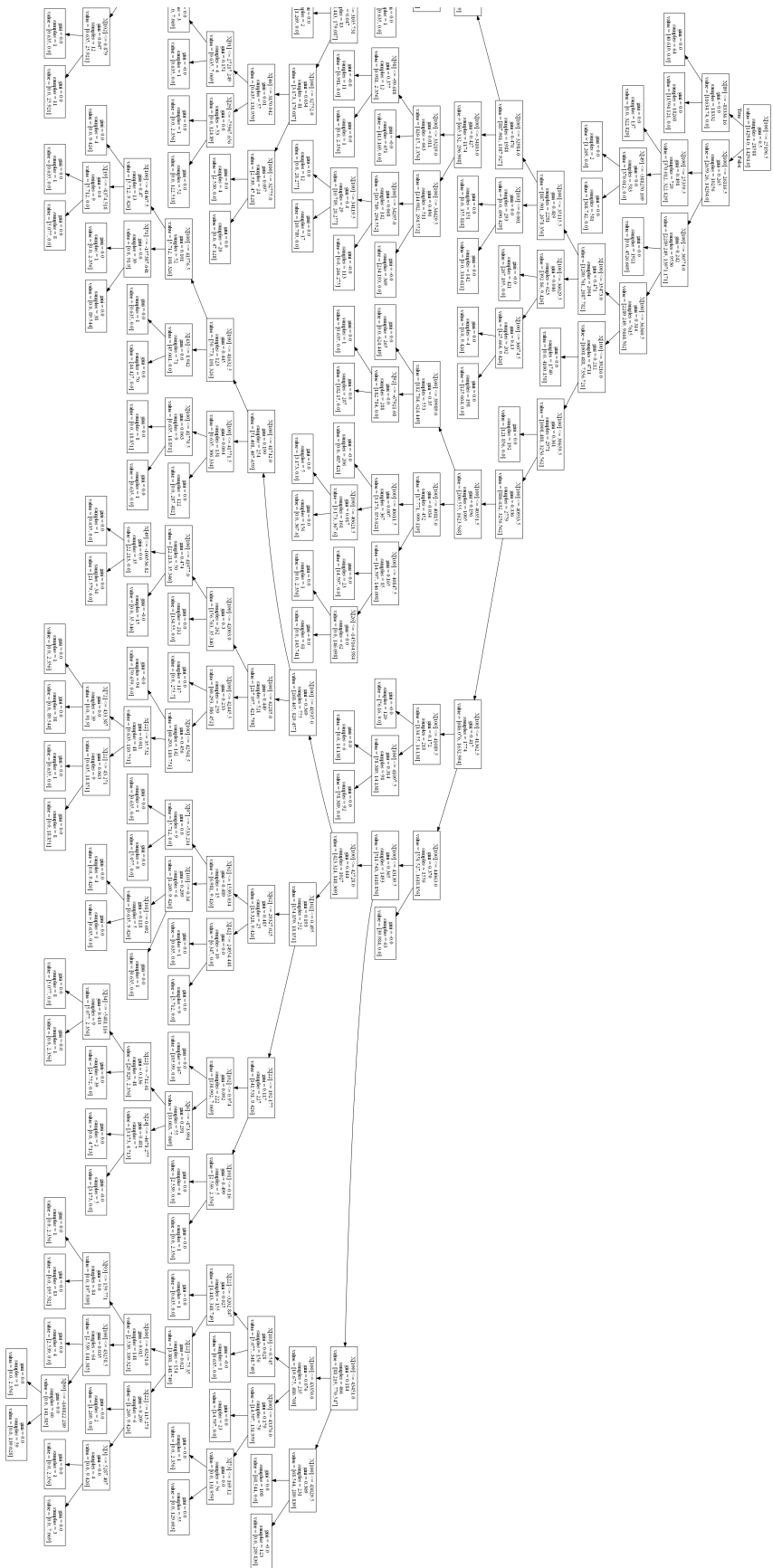


Figure 4: visualization of decision tree part 2 of 2