

Distinguishing — **Meme Stocks** from Investments

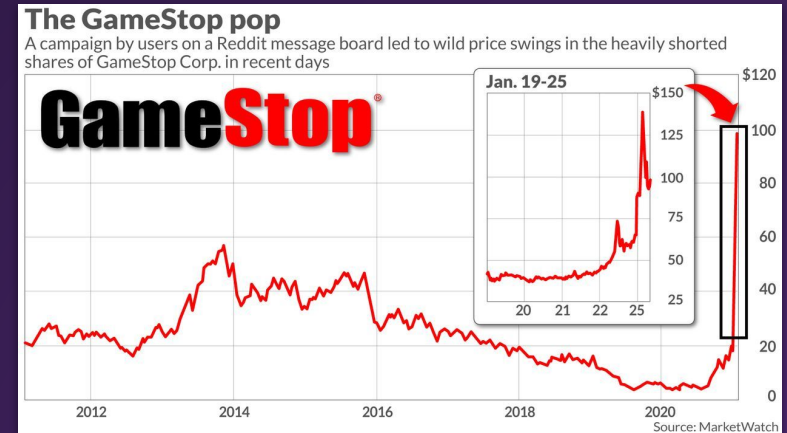
Ashley Poon
April 1, 2021



— The story of GME

In late January 2021, Gamestop's stock prices rose dramatically despite the company's shares being shorted at high rates.

Why?



— The story of GME

A handful of r/WallStreetBets users convinced Redditors en masse to buy up a bunch of GameStop stock, therefore increasing its value.



— Problem Statement

GME was an example of a high risk investment that resulted in large profits for some and large losses for others. However the risks were magnified as the price of the share was dictated by the whims of Redditors rather than actual market value.

Our goal is to **help every day investors find less risky, long-term investments** backed by financial and business results. More specifically, we want to help investors **distinguish investment advice from r/Stocks vs r/WallStreetBets.**

— Data Collection

Reddit API

Leveraged Pushshift Reddit API to collect comments from r/stocks and r/wallstreetbets.

Automation

Created a function that iterates the collection of 100 comments at a time per subreddit, using a sleep timer of 5 seconds.

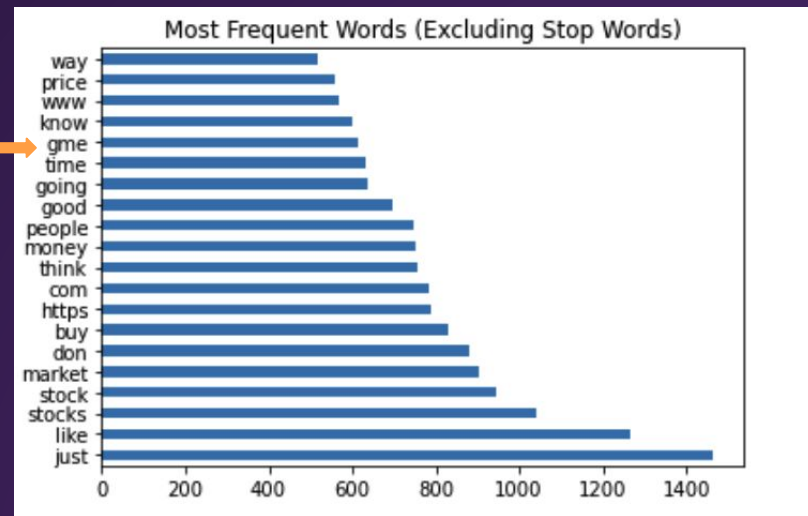
With each iteration, we found the minimum epoch time and started there for the next iteration, and so on until 20K comments were collected.

Compilation

Combined all pulls into a single csv file to use for cleaning and modeling.

— Cleaning & EDA

- Dropped all rows where comments were removed
- Checked the balance of classes (aka baseline accuracy):
 - 51% stocks / 49% wsb
- Binarized subreddits:
 - 0 for stocks / 1 for wsb
- Reviewed top words, excluding stop words, ie. "just", "like", "stocks", and "stock"



ROADMAP TO FINAL MODEL

BEST MODEL

Naive Bayes:

Bernoulli - 73% / 73%
Multinomial - 82% / 77%

1

Logistic Regression:

80% / 76%

3

Decision Trees:

Single - 68% / 66%
Bagging - 96% / 70%
Random Forest - 67% / 66%

5

2

BernoulliNB:

Lemmatizing - 74% / 73%
Stemming - 74% / 73%

4

Support Vector Machines:

80% / 75%

6

Boosting:

77% / 74%

Train / Test Scores

— Logistic Regression

High Accuracy Score

77%

Accuracy

One of the highest out of all the models.

Minimal Overfitting

5%

Difference

Between test score of 76% and train score of 80%.

Optimizes for Recall

82%

Recall

Achieves goal of minimizing false negatives (predicting r/stock when actually r/wsb).

— Conclusion

Confident that this **model is performing near the upper limit** given the **similar content** of both forums and the **bag-of-words modeling**.

Many of the misclassified comments are more **general or vague** and do not necessarily lean towards one subreddit or the other.

comment	actual	preds
Totally up to you! Think for yourself and decided if it's a good investment.	1	0
Block the wind while I roast this bone	0	1
The man!!!!!!!!!!	0	1
I'm in the same boat as you. Possibly going to scoop up some weekly calls to hopefully capture some of the rebound gains. 🙏	0	1
Fair enough	0	1
Doxxx	0	1
Yea, they crush it every Friday to hide the selloffs, it's been a pretty weird pattern.	1	0
none.	0	1
I'm planning on holding onto my positions. May average down if they get hit too hard....	0	1
+1 for CRSR!	0	1
Since the divorce they just can't go back to chilli's.	0	1
Oh, the irony of calling people sheep...	0	1
This is the way.	0	1

Conclusion

For the **false negatives** (predicting stocks when actually wsb), there are comments that could have been accurately classified if the model were able to pick up on sarcasm, which is a common tone in wsb.

Additionally, there are a few usages of wsb terminology that the model missed (ie. GUH, diamond hands).

comment	actual	preds
All the DD I ever need. All Might with diamond hands is too Perfect 💎👊	1	0
AMC on sale again ? Thanks you guys.	1	0
GUH 11 point down on the high already	1	0
I guess giant ship freed has been priced in huh...	1	0
I hear the weather in zero gravity is divine this time of year	1	0
oh boy, what a nice day, it'd be a real shame if somebody covered their positions, I'd \"hate\" that	1	0
PLTR is a long term play and always has been. You shouldn't have been in it in the first place if you weren't planning on holding for a few years. So yeah take the loss if you want to.	1	0
Short my portfolio, guaranteed returns	1	0
Still bleeding? Gotcha. Same time next week?	1	0
Then the entire basis of the stock market is a sham and it'll blow up anyway some other day.	1	0
There once was a boy from Bulgaria Who shorted GME out of hysteria He missed a step, made a bad bet And now he drinks piss in the area **ceo of melvin capital fromm bulgaria**	1	0
THIS JUST IN REDDIT TURNS THE S&P 500 INTO A MEME STOCK	1	0

— Conclusion

For the **false positives** (predicting wsb when actually stocks), the model misclassified several comments with strong or sarcastic language as wsb which is what a human might have guessed as well.

Additionally, some comments using common wsb terms (ie. GME, apes) were also misclassified.

comment	actual	preds
"We don't do that here" And don't buy gme.	0	1
> and then the algos race to the bottom. And Apes race to the moon 🏠💎👉👈🚀	0	1
Blue chip pump and dump? Now I've seen it all!!!	0	1
Gme is different that's an actual meme stock 😊	0	1
Holy sh t +45.82%, congrats on the tendies!	0	1
I hadn't even *started* repaying my student loans when the university started hitting me up for alumus donations.	0	1
I will buy whatever I want as long as it makes me profit. Lol Keep downvoting apes. 😊😊	0	1
stonks	0	1
The Irish are sick of everyone's sh t	0	1
This stock aged like milk	0	1
yOu wOUIDnT dOwNIOaD a CaR	0	1
Yup, they went all in on GME F ickers just couldn't help themselves when they saw the "bankruptcy lottery ticket"	0	1

— Recommendation

To improve the model, we can use **neural networks and sequence modeling** in NLP to help pick up on **nuances and sarcasm** in the comments. This would also help improve classification of **posts that use wsb's lingo**.

— **THANKS!**

Any questions?



— References

<https://time.com/5933242/gamestop-stock-gme/>

<https://www.wallstreetbets.shop/blogs/news/dissecting-the-unique-lingo-and-terminology-used-in-the-subreddit-r-wallstreetbets>