



When you are just starting to invest, there are many places to go for stock news, investment tips, market information..

But where do you start? Do you read everything? Is there a difference between subreddits?

Should you read r/Stocks or r/Investing or both?

Comparing Subreddits

r/Stocks

- "Any post related to stocks"
- Stock suggestions
- Trading strategies
- Options strategies
- Lots of questions

r/Investing

- No description whatsoever
- Similar to r/Stocks
- ETFs, brokers, commodities, etc



Title Words

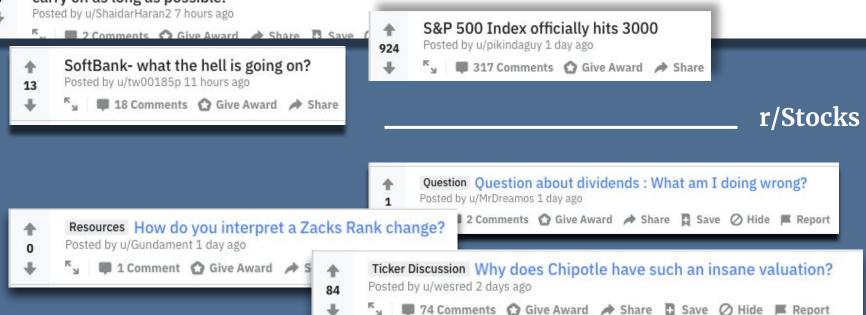


You have to hold 3 (and only 3) stocks for the next 20 years. Which do you choose and why?

Posted by u/RednRich 10 hours ago

Blue 88 Comments Give Award Share Save Hide Report

Powell testimony day 2 - Says U.S. economy was hit by 'confidence shock', has only partly recovered. Wants expansion to carry on as long as possible.

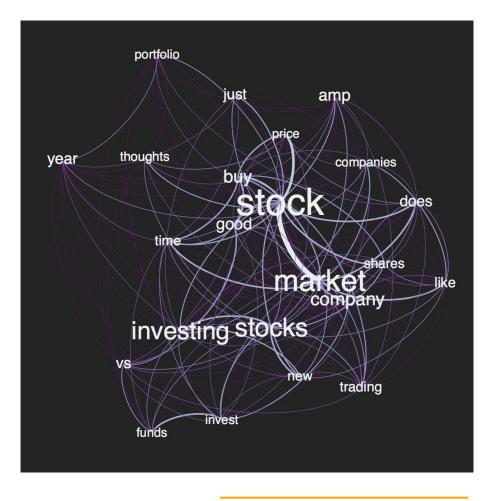


Network Graph

An analysis of the co-occurence of words from the subreddits.

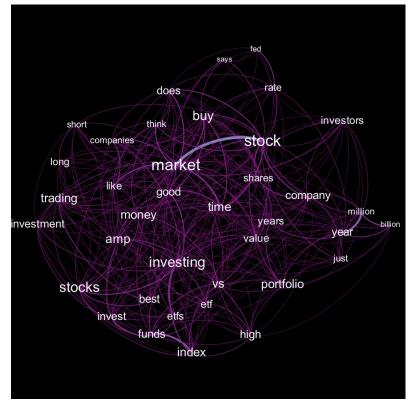
The lines represent the co-occurence of words in the dataset.

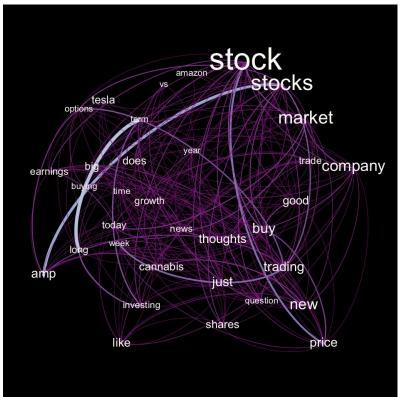
The size of the text represents the amount of times word is seen in dataset.



r/Investing

r/Stocks





Most frequent words

Total of 1,687 posts

1,784 words from r/Stocks

2,272 words from r/Investing

Total frequency 9,397

An overlap in stock and stocks

People seeking for advice and insight of others

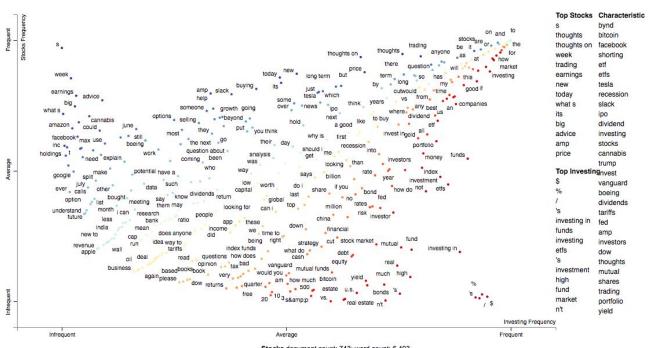
r/Stocks	Frequency	r/Investing	Frequency
stock	85	market	81
stocks	82	stock	79
thoughts	39	investing	74
trading	35	stocks	55
buy	28	advice	49

Scatter Text

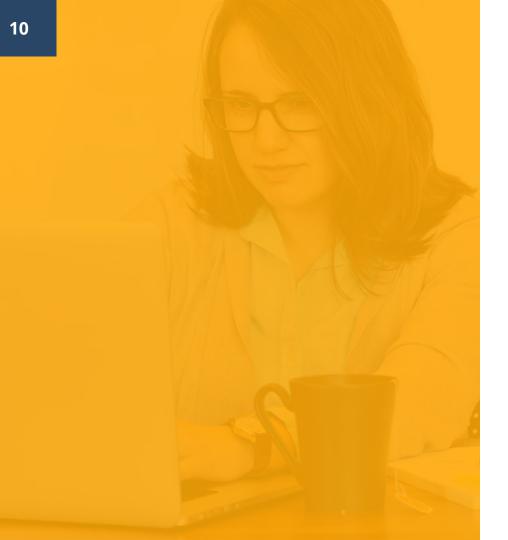
The relationship between terms in the subreddits

Top terms for both subreddits

Topics discussed in the subreddits



Stocks document count: 743; word count: 6,493 investing document count: 944; word count: 9,403



The Model

The purpose of the model is to help new investors decide if there is a large enough difference between topics discussed to warrant reading both subs

Baseline	0 (r/Stocks) = 44.04	1 (r/Investing) = 55.95
Test Data Scores	CountVectorize	TF-IDF
Logistic regression	63.51	62.56
Multinomial Bayes	63.51	64.22
Decision Trees	60.19	59.72
Bagging	61.61	60.19

Confusion Matrix

Specificity: 38.17

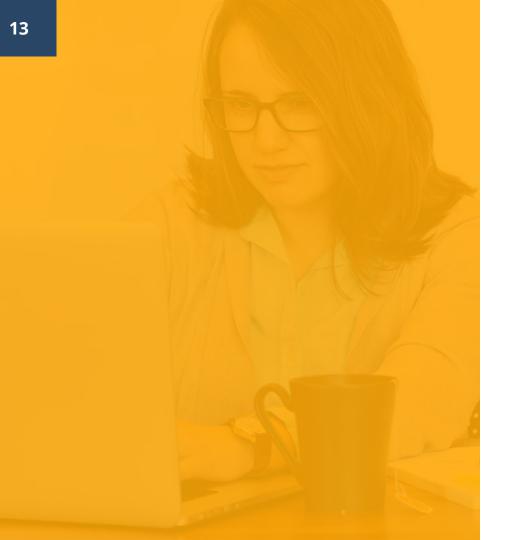
Sensitivity: 84.75

Accuracy: 64.22

More sensitive than specific, better at predicting r/Investing posts

Multinomial Bayes TF IDF

	Actual Stocks	Actual Investing
Predicted Stocks	71	115
Predicted Investing	36	200



Next Steps

So... which one do you read?

Scores:

Accuracy score between models were similar, not much better than baseline.

Model cannot accurately separate posts based on title alone.

Topics discussed in both subreddits are very similar, pick your poison!

Limitations:

Only inspected titles, no context or commentary

Missing breadth of conversation

Not enough testing of different parameters, model too slow

