



Text Classification Within Subreddits

A comparison between r/WallStreetBets and r/Investing



Problem Statement

Our client, Ademala Research—a hedge-fund of relaxed moral composure-wished to move capital markets by using spam bots to post spurious investment news and information on personal finance subreddits.

We used an NLP classification model to conduct a preliminary study of the language, phrases, and idioms used in two popular subreddits, r/WallStreetBets and r/Investing, so that Ademala had the linguistic tools required to build highly-tuned and nefarious spam-bots.

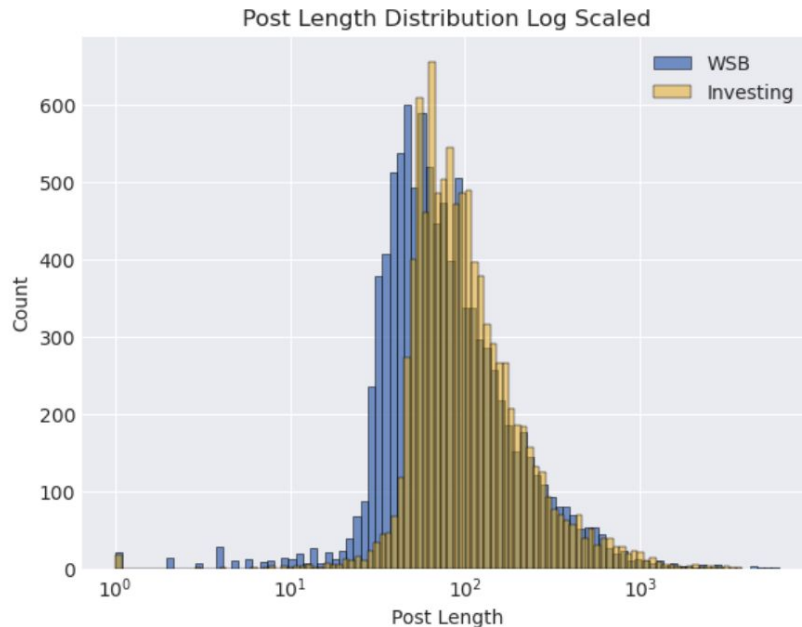
Data

r/WallStreetBets

- > 13M members
- 10,000 posts collected
- 6,000 unique authors
- Average post length is 120 words

r/investing

- > 2 M members
- 10,000 posts collected
- 7,000 unique authors
- Average post length is 140 words





Posted by u/flopsie123 5 years ago 🗨️ 🔄



I somehow made \$110k this morning and I'm still not totally sure how

STRENGTH

Levered 460:1

Yesterday at 4:10pm EST I sold the 266.50/266.00 put spread for a 2 cent credit. 1,000 of them. I guess at 4:14 or 4:15 before the option expired SPY must've dipped below 266.50 because I woke up the next morning before the market opened and saw my balance was up about \$140k. I thought it was a glitch with the Etrade app at first but then looked at my portfolio and saw what happened. For some reason I was assigned only 863 of the 1000 put contracts I sold. That's 86,300 shares of SPY which is about \$23,000,000 worth. I only had \$50,000 in my account so I started panicking and wondering if Etrade would liquidate my account or something, so I put in a limit order of 268 (SPY was trading around 268.12 at the time), hoping the sale would go through immediately. After putting in the order a message popped up saying my order would be put through when the market opened, which I don't understand because I thought you could buy and sell stocks/ETFs during premarket hours. So anyway I was stuck staring at my phone's screen until 9:30 when the market opened. Immediately about half my shares sold for \$268 each. The market seemed to be dropping so I lowered my limit for the rest to 267.60 and the rest of them sold for that price. After all that my balance was about \$112k higher than it was yesterday. Does anyone know how much interest Etrade will charge me for holding 23 million dollars of SPY for these few hours? Proof: <https://imgur.com/a/onEjU>



387 Comments



Award



Share



Save





15



r/investing · Posted by u/12kkarmagotbanned 9 hours ago



Has any research been done on if different types of bonds increase the Safe Withdrawal Rate (SWR)?

When retired, do different kinds of bonds change the SWR? I'd imagine the default assumption is Intermediate US Treasury Bonds, like VGIT.

Does adding corporates increased the SWR? How about changing the bond duration? Long term? Short term? Does adding global bonds change it?

Any research on this?



5 Comments



Award

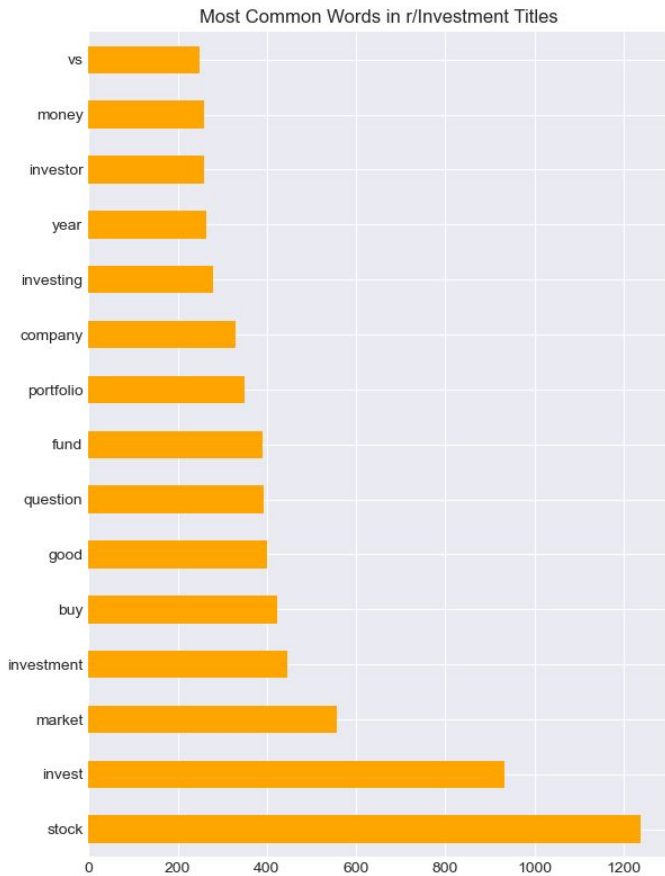
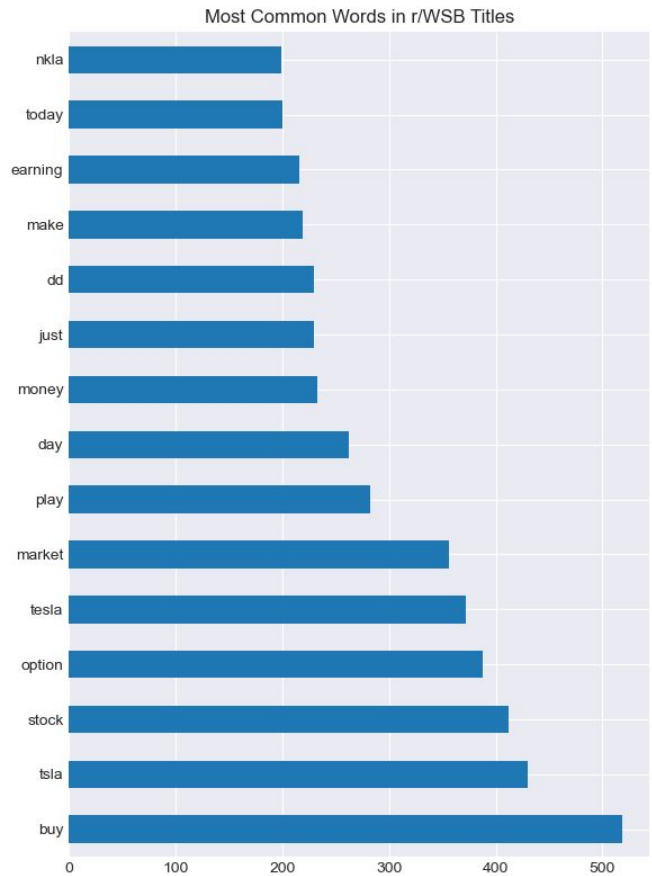


Share



Save





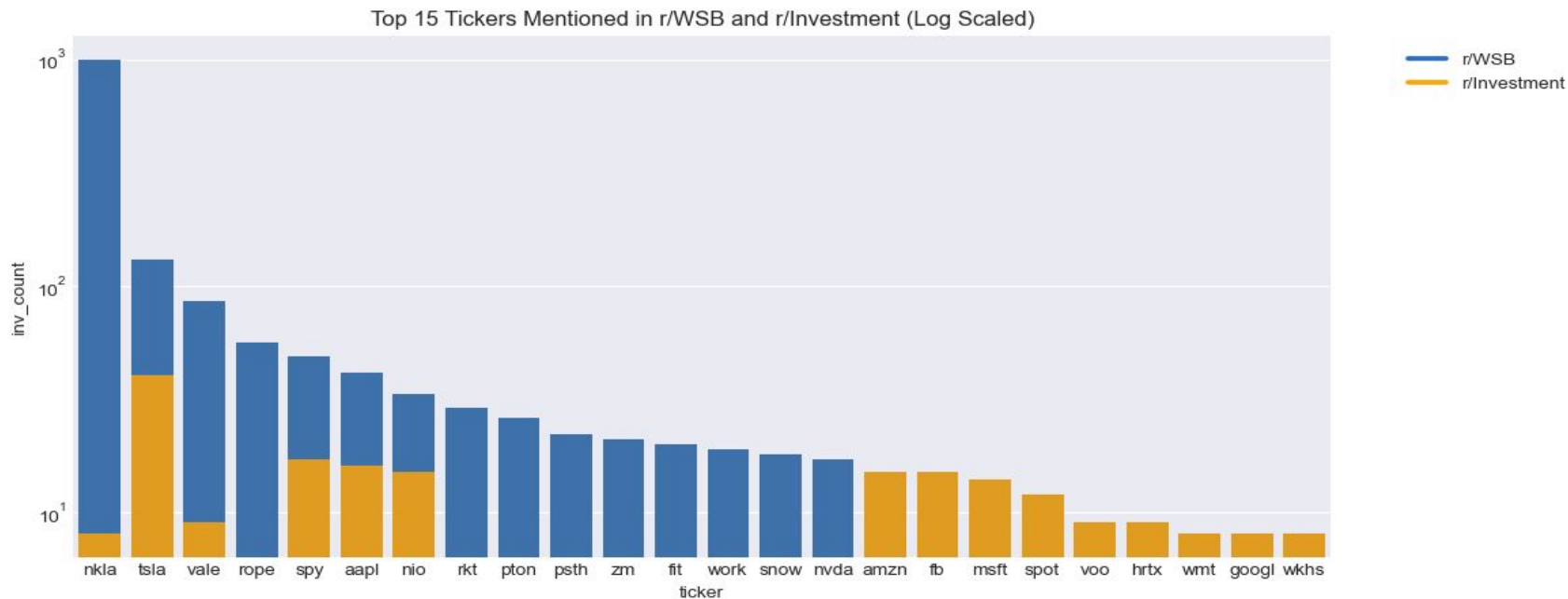
r/WSB Most Common Words:

- Tesla
- Option
- DD
- Play

r/Inv. Most Common Words

- Invest
- Market
- Portfolio
- Fund

Most Frequently Mentioned Securities

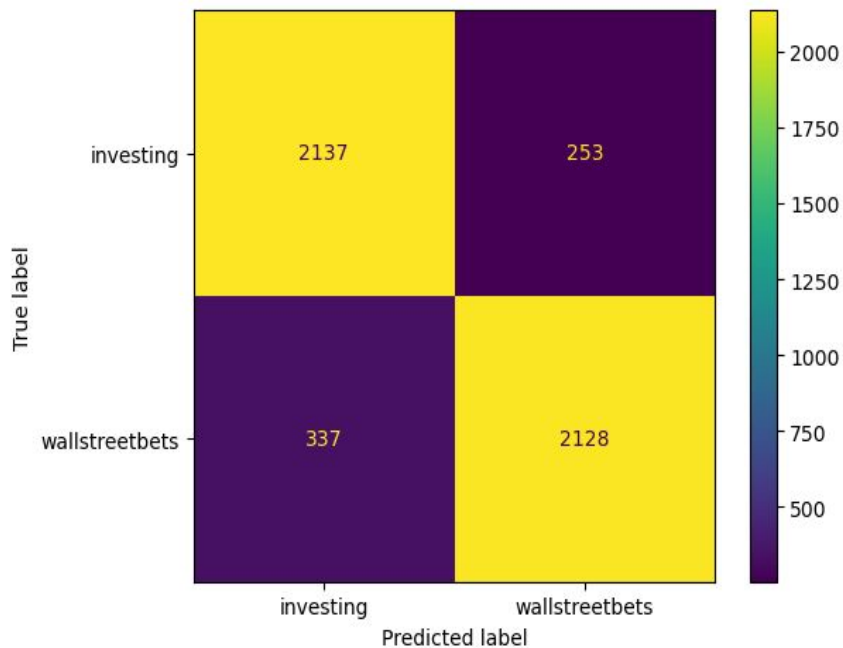




Building a Language Classification Model

1. Tried Logistic Regression, KNN, Random Forest and Extra Trees as well as Naive Bayes evaluators, and multiple transformers. In total 20,750 fits were made.
2. Final Model uses Lemmatization, a TF-IDF Vectorizer, and a Logistic Regression Classifier.

Model Results



Accuracy Score of .88 on unseen data

- .89 Accuracy for r/WSB
- .86 Accuracy for r/Investing



Conclusions

1. Logistic Regression was the best performing model
2. The TF-IDF vectorizer was best
3. We achieve better classification accuracy using title and selftext from subreddit posts rather than one or the other.
4. Our classification model breaks down when either subreddit uses general investment language.

Spam-Bot Recommendations

Depends on the Subreddit:

For r/WSB:

1. Use the words option or Tesla
2. Use Emojis
3. Be Vulgar

For r/Investing

1. Don't do any of the above.