



# **PROJECT 3:** **NLP WITH** **REDDIT**

# HELLO!

**I'm Varun!**

I like doing cool stuff with NLP!





1.

# Our Goals

If it ain't broke...



**Can we build a classifier capable of distinguishing between subreddits based on the language used in each?**

# Our Two Subreddits

## r/Investing

Discussions about public markets, corporate finance, federal reserve, government spending, investment portfolios.

## r/PersonalFinance

Discussions about paychecks, bills, personal debt, mortgages, careers, 401k, making ends meet, etc.

# My Initial Impressions

## r/Investing

- ❑ More focused on firms and their decisions, how they affect investors
- ❑ Discussions about daily news
- ❑ Larger scale

## r/PersonalFinance

- ❑ Individual stories, each post often paints a picture
- ❑ More about managing personal finances, taxes, debt
- ❑ A little more sad to read through



2.

# Diving into the Data

Scraping and preliminary observations



# **40,000 posts**

20,000 scraped from each sub

# **11,888 empty selftext**

8,253 r/Investing; 3,634 r/PersonalFinance

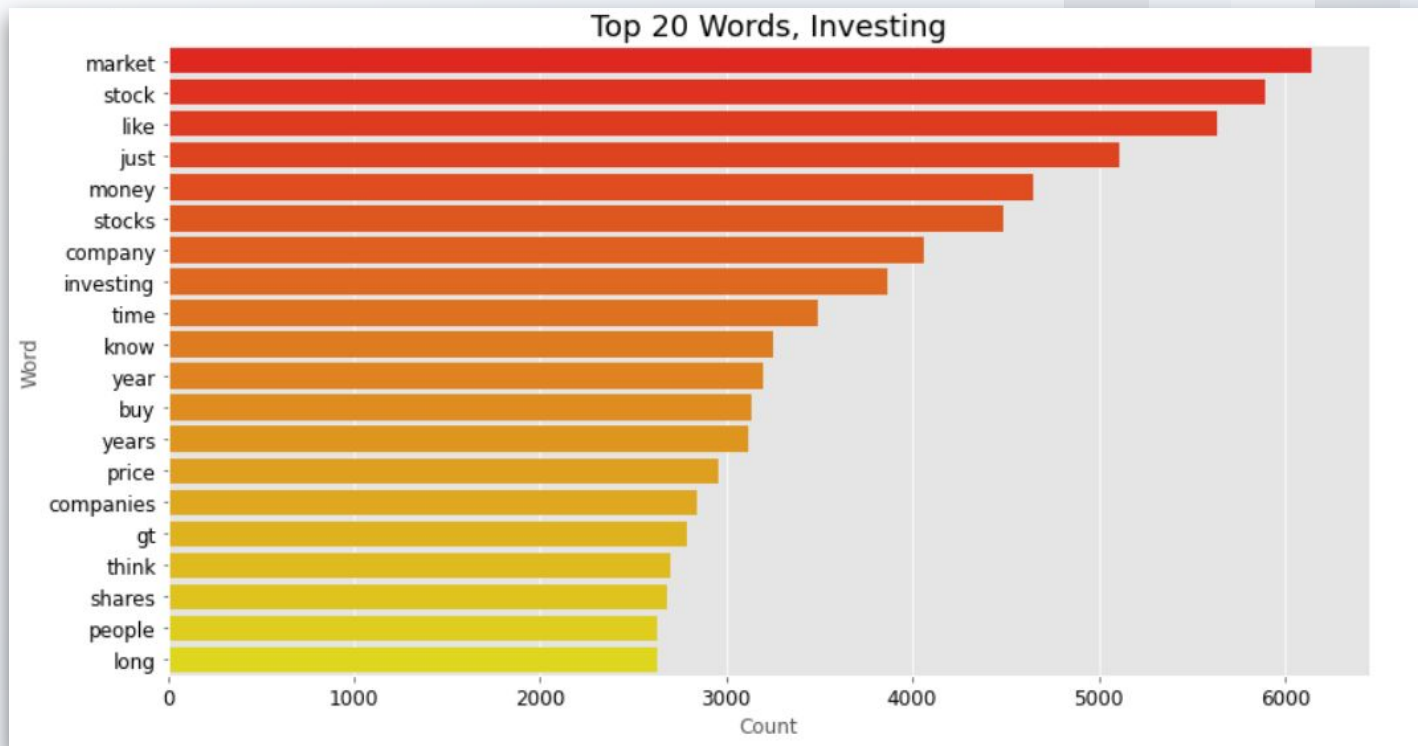


# **0 empty titles**

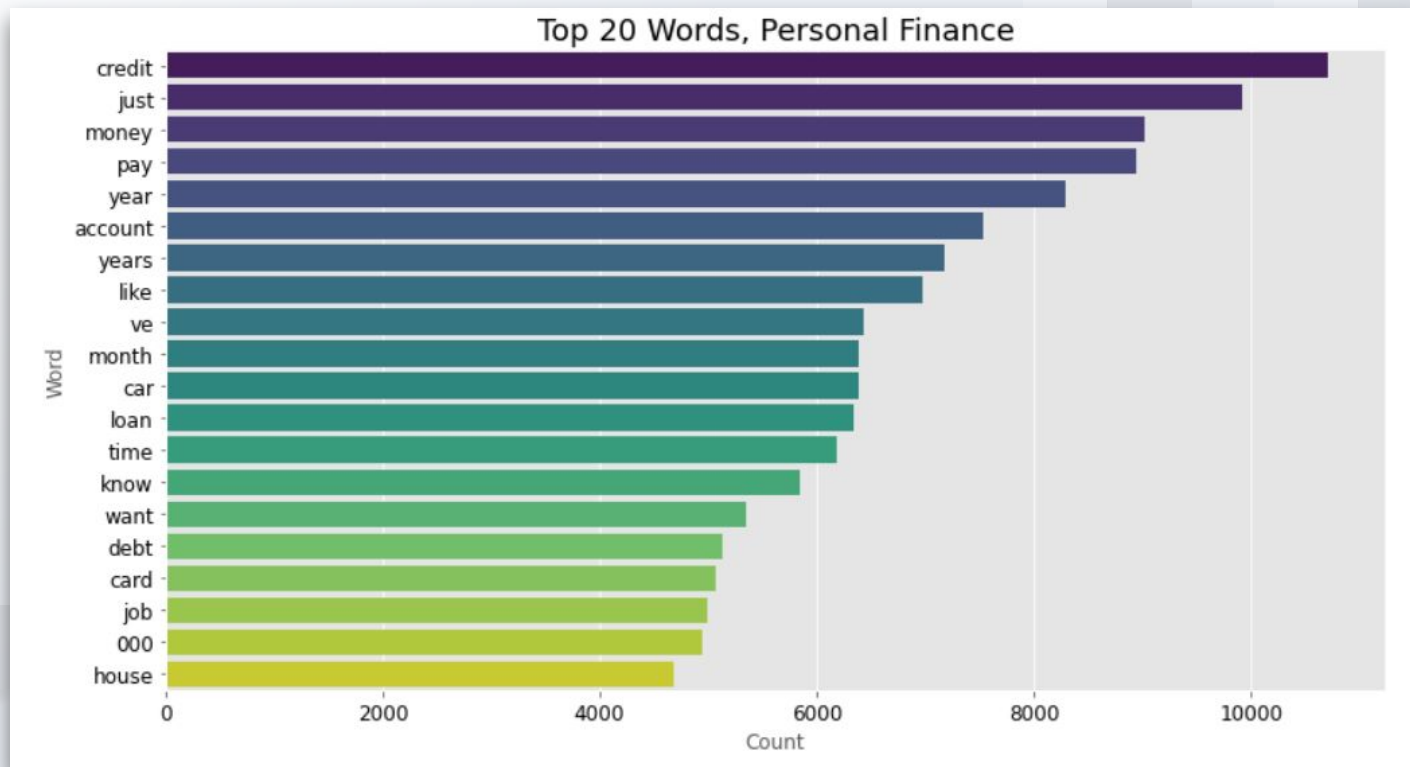
Which is why it helped to use both



# TOP 20 MOST COMMON WORDS IN r/INVESTING



## TOP 20 MOST COMMON WORDS IN r/PERSONALFINANCE





3.

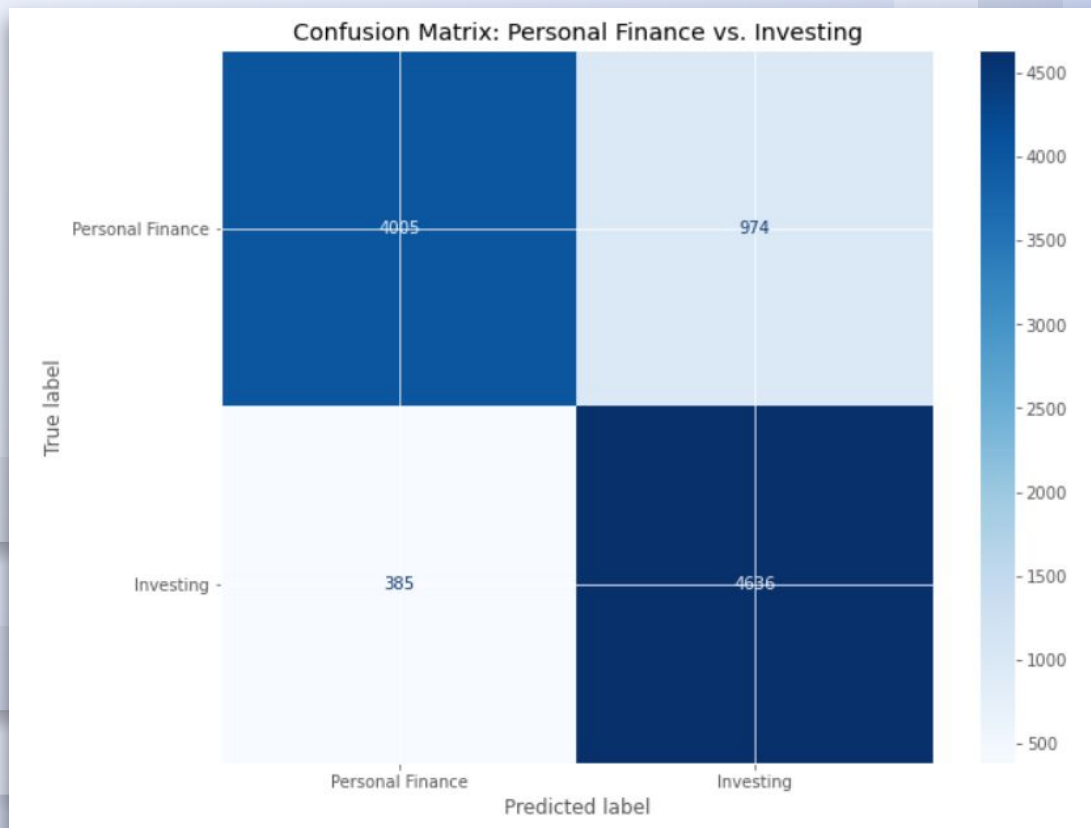
# Modeling

Logistic, KNN, Forest, oh my...

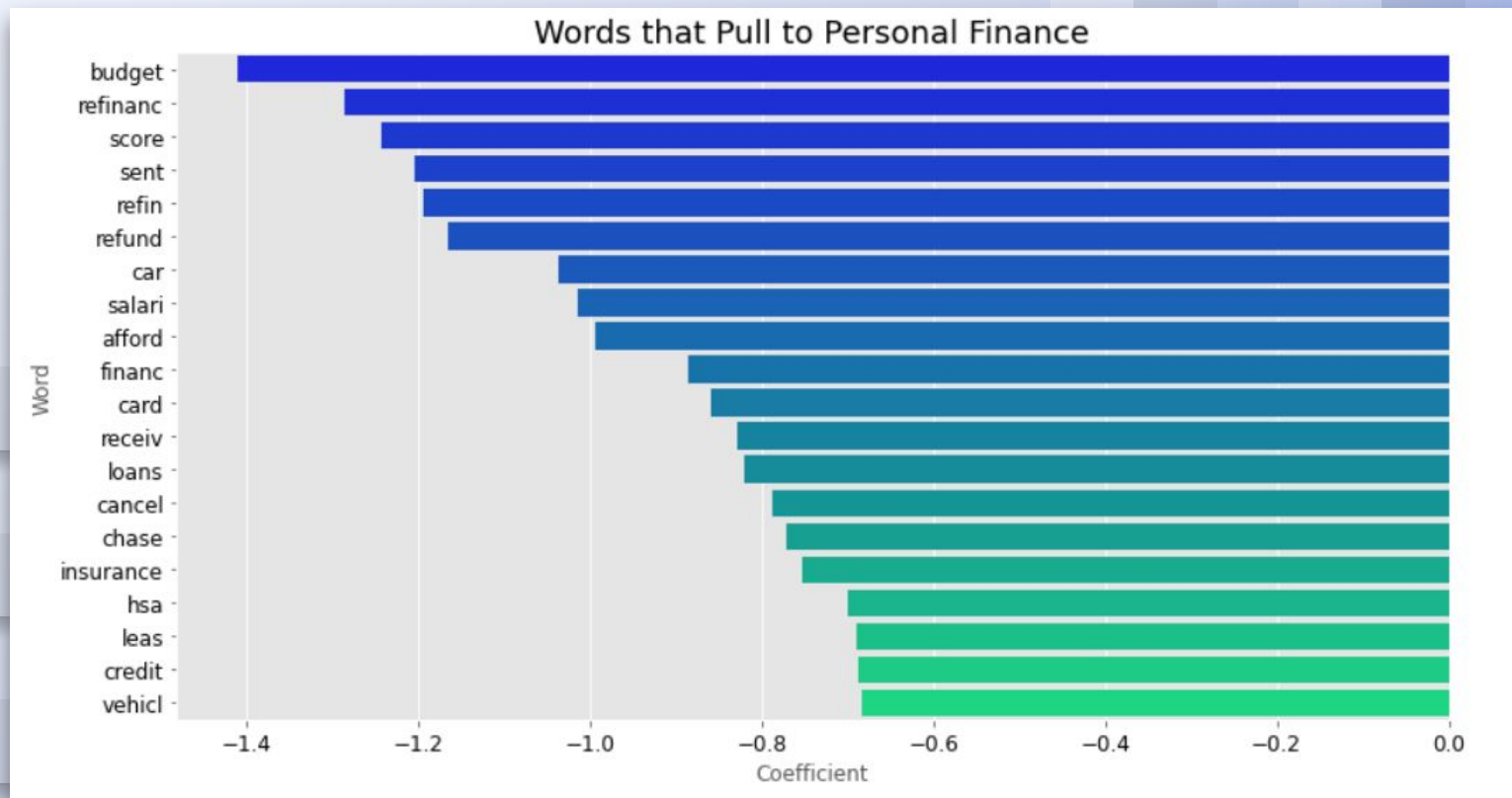
## Some of My Classifiers

Type	Important Params	Train Score	Test Score
<b>Logistic/CVect</b>	<b>C = 0.4</b>	<b>0.8588</b>	<b>0.8515</b>
Logistic/TFIDF	C = 0.4	0.8567	0.8485
NBayes/TFIDF	Alpha = 0.1	0.8267	0.8173
RForest/CVect	Dep = 7, Leaf = 2	0.7932	0.7903
KNN/CVect	K = 9	0.7892	0.7709

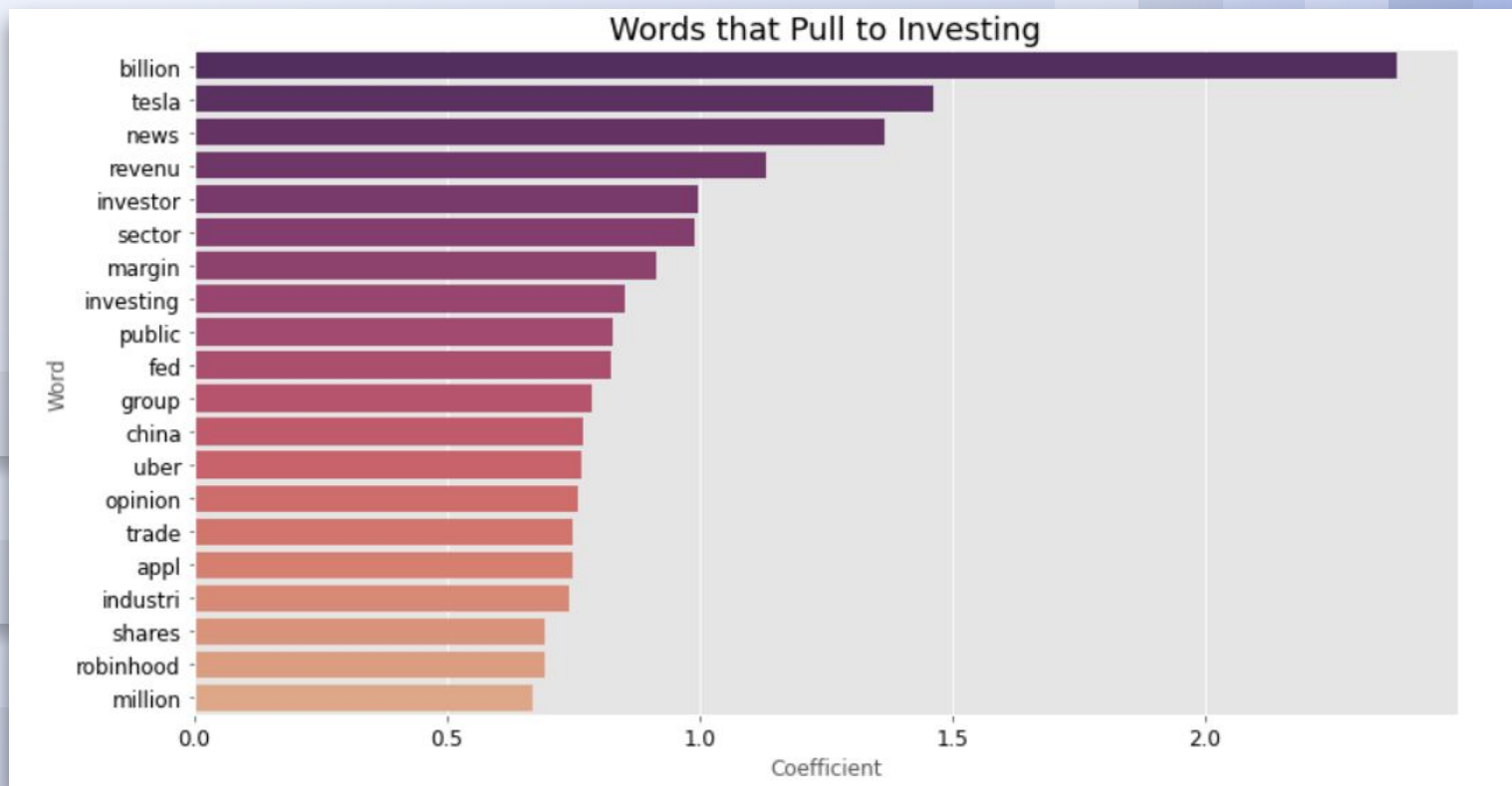
# CONFUSION MATRIX FOR OUR BEST MODEL: **LOGISTIC REGRESSION**



## HIGHEST WORD COEFFICIENTS TOWARDS **PERSONAL FINANCE**



# HIGHEST WORD COEFFICIENTS TOWARDS **INVESTING**





4.

# Conclusions

What can we say about what people are saying?



# Selected Takeaways

## Scale

Investing talks more about current events, the Fed, larger sums of money, macro-economics. Personal Finance tends to be more specific.

## Overlap

There are definitely posts in Personal Finance that could have justifiably ended up in Investing, some but less of the reverse.

## WHAT ARE PEOPLE TALKING ABOUT IN **r/PERSONALFINANCE**?

Lots of **job** shifting and questions about it, a good number of people asking about **savings**, credit health.



## WHAT ARE PEOPLE TALKING ABOUT IN **r/INVESTING**?

Lots of discussion about **Tesla**: they've been performing well recently after overcoming Model 3 production issues. Even during the pandemic, Tesla maintains a strong market in China.



# Live Predictions!



# THANKS!

**Any questions?**