Problem Statement

Classifying Spam Agency of Singapore would like to create a filter that classifies financial updates and posts as potentially helpful or spam.

Using the following subreddits as examples, we will try to create a model that is able to classify a given body of text:

r/personalfinance

Tends to be more conservative financial advice

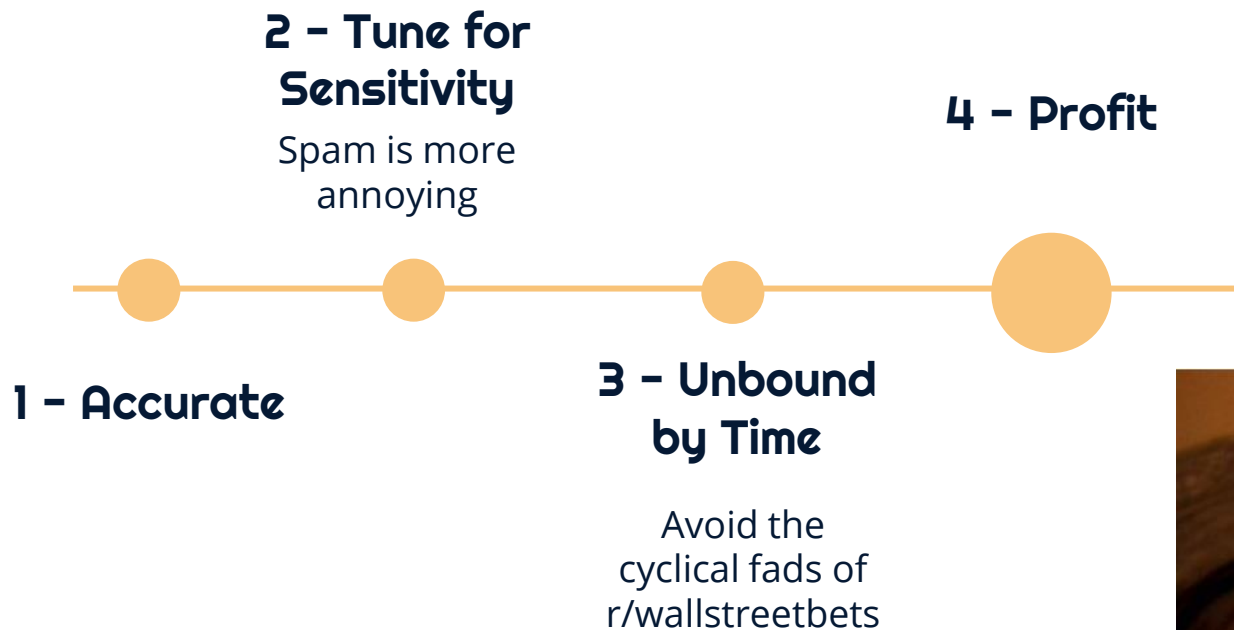Targeted at financially sustainable living or retirement

r/wallstreetbets

More active investing

Driven by the stock of the month

Accuracy, Sensitivity, Longevity

2 – Tune for Sensitivity

Spam is more annoying

4 – Profit

1 – Accurate

3 – Unbound by Time

Avoid the cyclical fads of r/wallstreetbets

ONE DOES NOT SIMPLY

MAKE A MODEL

Models Tested

Classifier

Multinomial Naïve Bayes

Random Forest

Extremely Randomized Trees

ADA Boost

Vectorizer

Count Vectorizer

TF - IDF

Words

Cleaned

Lemmatized

Stemmed

Model Summary

	Naive Bayes						Random Forest		Extremely Randomized Trees		ADA Boost	
Word Vectorizer	Count Vectorizer			TF-IDF			Count Vectorizer	TF-IDF	Count Vectorizer	TF-IDF	Count Vectorizer	TF-IDF
Word Modifications	Unmodified	Lemmatize	Stemmed	Unmodified	Lemmatize	Stemmed	Stemmed	Stemmed	Stemmed	Stemmed	Stemmed	Stemmed
Train Accuracy Score	0.9502	0.9536	0.9540	0.9234	0.9292	0.9296	0.9430	0.9344	0.9445	0.9445	0.8522	0.8804
Test Accuracy Score	0.9497	0.9483	0.9512	0.9153	0.9282	0.9426	0.9512	0.9368	0.9454	0.9440	0.8522	0.9053

Stemmed Words performed better on the initial Naïve Bayes, so they were used on the other tests

The best models in terms of accuracy on the test set had the same score of 0.9512

Naïve Bayes, Count Vectorised, Stemmed Words

Best Model had the following parameters:
- Multinomial Naïve Bayes
- Count Vectorised
- Stemmed Words
- Max Features = 5000
- nGram Range = 1, 1

To tune for sensitivity, the threshold of the probability of the post to be classified as r/wallstreetbets was shifted from 0.5 to 0.1

- Sensitivity increased from **0.9256** to **0.9473**

- Specificity loss from **0.9732** to **0.9679**
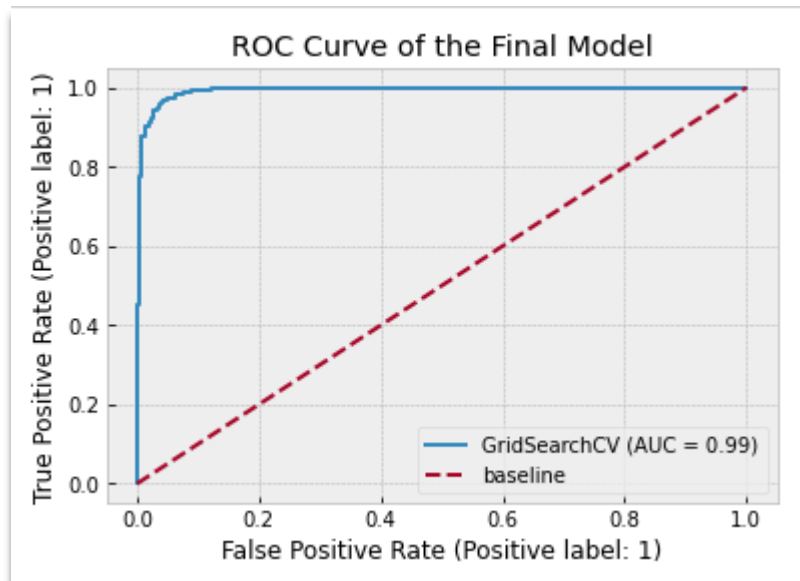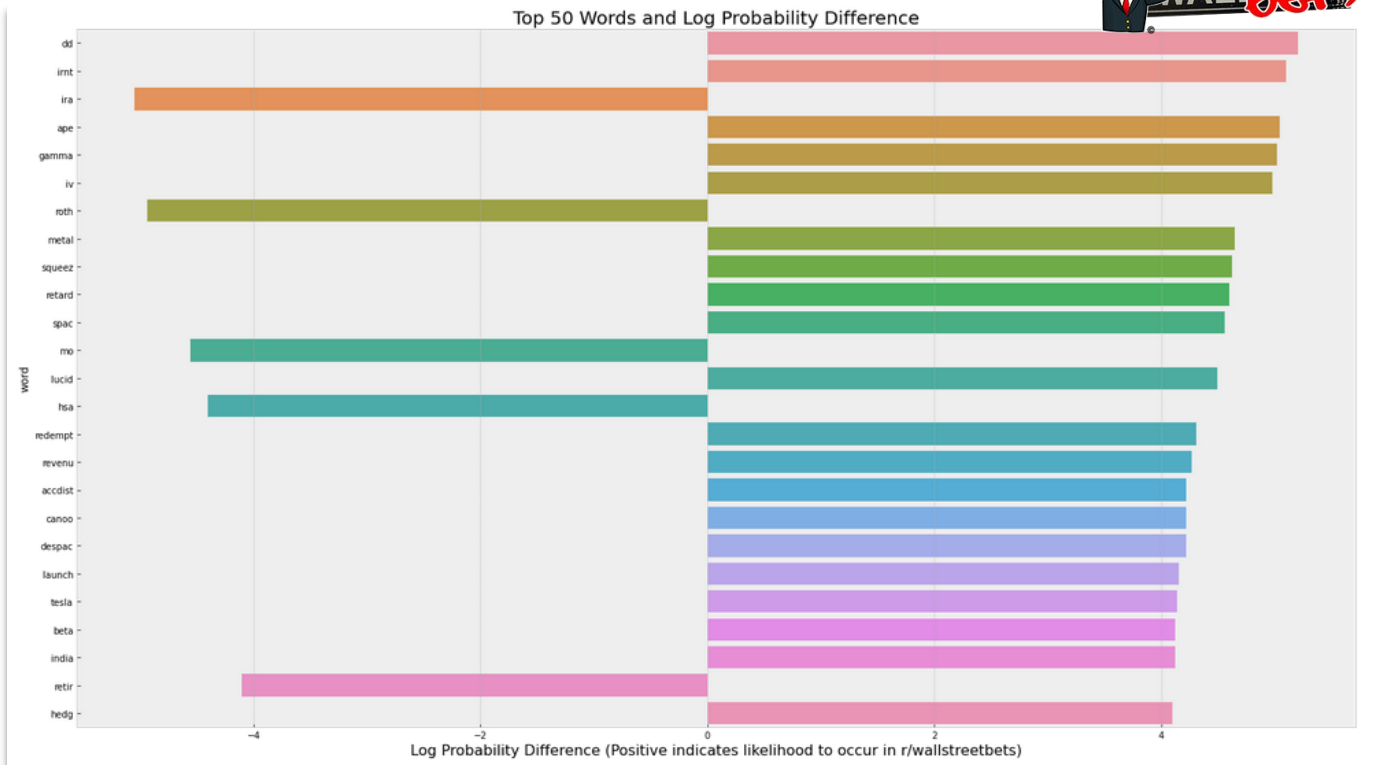

AUC Score = 0.9937

Best Indicators Came From r/wallstreetbets
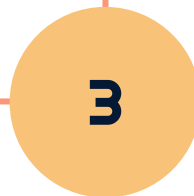
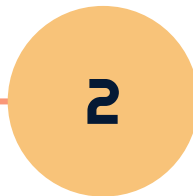The best indications came from r/wallstreetbets

r/wallstreetbets tends to have its own "lingo" of inside jokes and shorthand, e.g. dd – due diligence

r/personalfinance showed a clearer bent towards slow wealth accumulation:
ira, roth, hsa

Top 50 Words and Log Probability Difference

Log Probability Difference (Positive indicates likelihood to occur in r/wallstreetbets)

Takeaways

**Accuracy –
95.8%**

Time
Popular stock tickers
removed in stop words

1

2

3

4

**Sensitivity –
94.7%**

It's larger than the Moon

Success?

Mostly successful, however, 4 posts
that were false negative slipped
through should have been caught and
the model could be further improved

Improving The Model

Titles
Words in the titles could be weighted heavier than the body

Context
Model assumes words are independent of each other, adding a dimension for context would help increase accuracy

Metadata
Post length might be an additional indicator as posts giving advice tend to be longer

THANKS!