Relevant Posts, Happy Users

A presentation to Reddit APAC Senior Management

Leveraging machine learning techniques to improve content moderation and grow user base

By: Team DSI-23 - Kimberley Chen, Andrew Chia, Timothy Chan, Lee Shi Min

Agenda

01 Introduction
and Problem Statement

02 Exploratory Data
Analysis

03 Modelling
and Evaluation

04 Conclusions and
Recommendation

52 MILLION

52 Million Daily Users on Reddit in 2020

303.4 MILLION

303.4 Million posts on Reddit in 2020
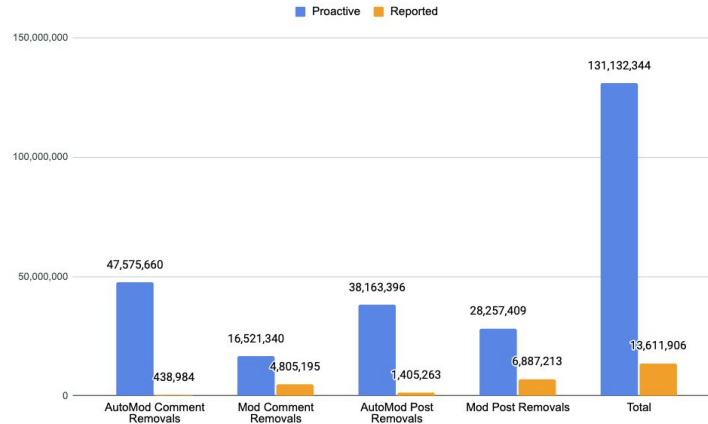
49.2 BILLION

49.2 Billion Upvotes on Reddit in 2020

Content optimisation is a cornerstone...

- High impact on **user retention, engagement and growth**
- > 70,000 volunteer moderators are critical partners to ensure posts are safe, original, authentic
- AutoModerator tool introduced in 2019

... backed by extensive moderation efforts

More than 100 million posts removed in 2020!

Proactive ■ Reported

- 150,000,000
- 100,000,000
- 50,000,000
- 0

AutoMod Comment Removals: 47,575,660 / 438,984
Mod Comment Removals: 16,521,340 / 4,805,195
AutoMod Post Removals: 38,163,396 / 1,405,263
Mod Post Removals: 28,257,409 / 6,887,213
Total: 131,132,344 / 13,611,906

2 types of moderators:

Auto Moderator

AutoModerator is a system built into reddit that allows moderators to define "rules" to be automatically applied to posts in their subreddit.

Human Moderator

Redditors who volunteer their time to help guide and create Reddit's many communities.

Able to report or remove comments or submissions containing certain words or phrases

Able to report or remove comments or submissions containing certain words or phrases

Ban spammers or other redditors who may be breaking community rules from posting or commenting in their community.

Improving moderation efforts and content optimisation

Scope to further curate posts for relevance
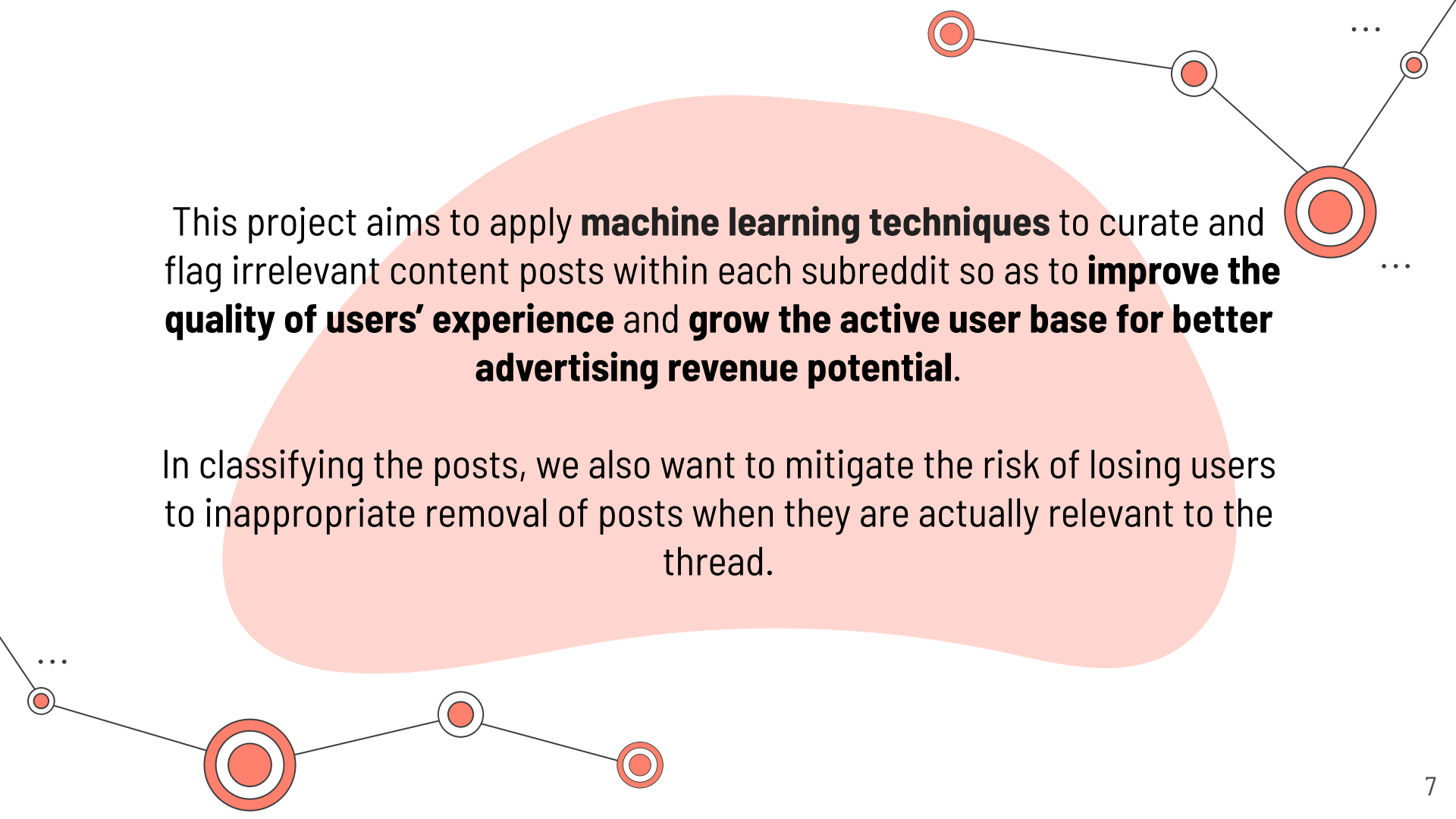
Relevant posts keep users reading and engaged

Easing moderation work

Leverage machine learning tools to help moderators complete key tasks with less time

More moderators stepping up to support communities

A more vibrant reddit community

This project aims to apply **machine learning techniques** to curate and flag irrelevant content posts within each subreddit so as to **improve the quality of users' experience** and **grow the active user base for better advertising revenue potential**.

In classifying the posts, we also want to mitigate the risk of losing users to inappropriate removal of posts when they are actually relevant to the thread.

2 Datasets that were used:

Stocks
2.9 million community members
Created on 27 June 2008

CryptoCurrency
3.3 million community members
Created on March 11 2013

Data Size:
1,700
Time Range:
July 21, Wednesday 2021   (GMT)
July 22, Tuesday 2021   (GMT)
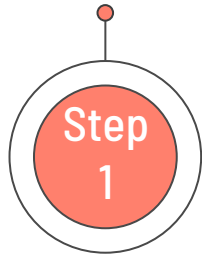
Data Size:
1,700
Time Range:
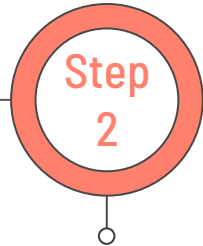July 21, Wednesday 2021   (GMT)
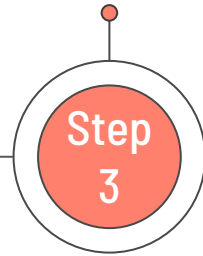July 17, Saturday 2021   (GMT)

Outline of our Workflow

Web Scraping

EDA

Insights and Recommendations

Step 1

Step 2

Step 3

Step 4

Step 5

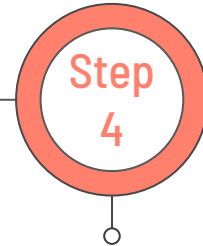**Cleaning of the Data**

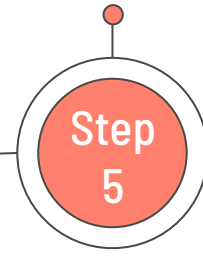**Modelling**

Data Cleaning / Data Processing

Data Cleaning

- Filtering null values and [removed]
- Remove duplicate entries

Natural Language Processing Techniques

- Removing punctuations
- Tokenizing
- Remove stopwords
- Done lemmatization (more accurate compared to stemming)
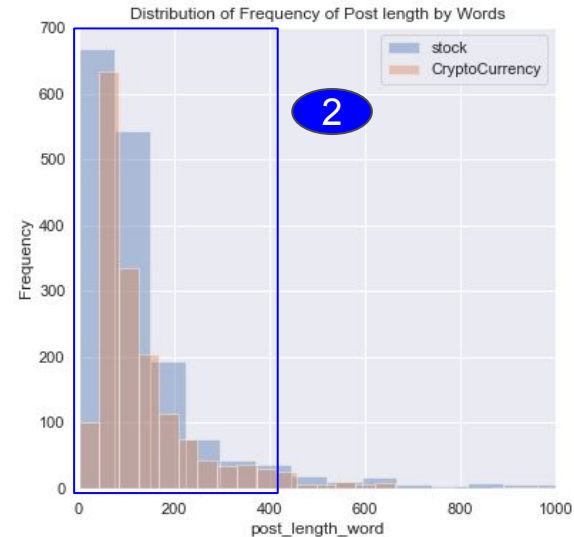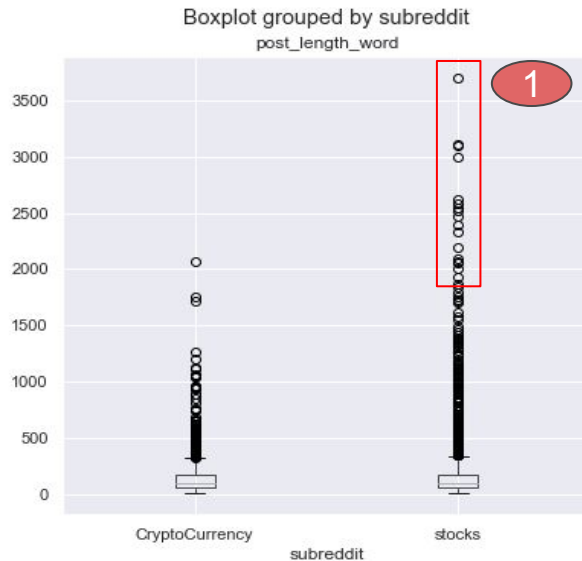
Vectorization

- Count Vectorization
- N-Gram Vectorization
- TF-IDF Vectorization

Exploratory Data Analysis

To understand more about 2 datasets
from **r/stocks** and **r/CryptoCurrency**

Length of subreddit posts by Word Count

1 **Stocks** subreddit have <u>higher mean</u> due to presence of more post >2000 words.

2 Based on distribution, most post from <u>both</u> subreddits have < 400 words.

Boxplot grouped by subreddit
post_length_word

Distribution of Frequency of Post length by Words

Most related words to "stocks" & "Crypto"

By using <u>Word Vectorization</u>, following words are top 10 related words with 'stocks' and 'crypto'

1 However, words belonging to Cryptocurrency category appears in words related to stocks.

Therefore, this highlights the potential challenge we will faced in classification modelling later.

Stocks

	Word	Word Vector Coefficient
0	dip	0.964579
1	buying	0.963334
2	sell	0.945972
3	btc	0.945943
4	buy	0.940262
5	selling	0.935968
6	eth	0.932741
7	hold	0.919836
8	spacs	0.919339
9	back	0.919201

Crypto

|	Word	Word Vector Coefficient
0	investing	0.991024
1	hey	0.98157
2	everyone	0.980778
3	curious	0.980551
4	trying	0.9802
5	someone	0.979036
6	guy	0.978475
7	hope	0.976843
8	wondering	0.97567
9	best	0.974884

1-word appear frequently in 'Stocks'

Most words are not a surprise:

- **'company'**: Stocks is a company's equity.
- **'market'**: Place that we buy/sell stocks.
- **'price'**: Indicator on how a stock performs.
- **'share'**: Synonym to stock.

Some Insights

- There are some words that shows optimism towards the stock market Words: **'buy', 'growth','good'**

Frequency of 1-word appearance appeared in subreddit "Stocks"
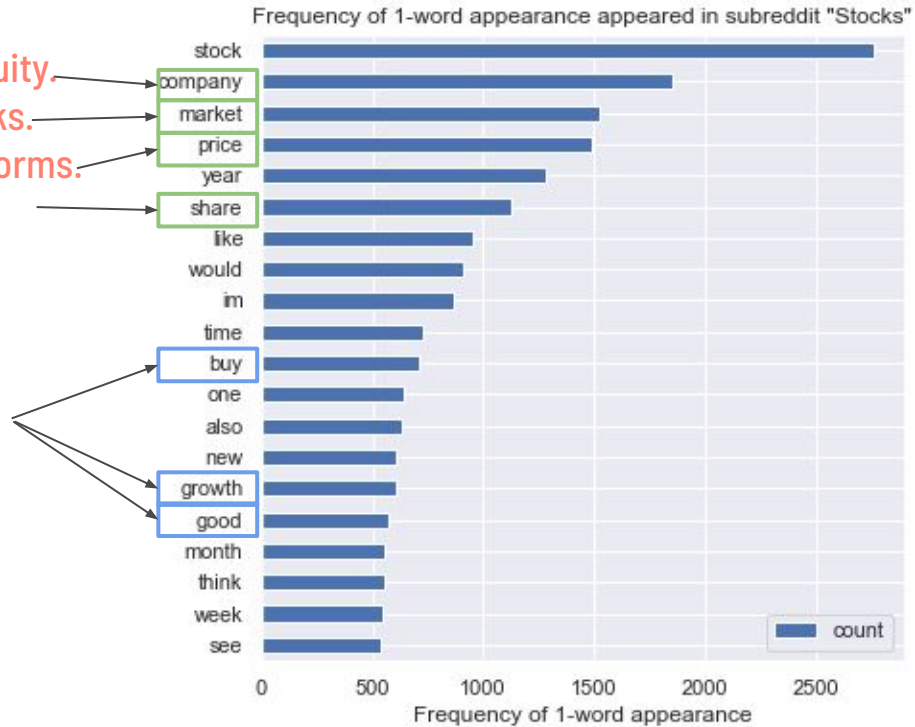
1-word appear frequently in 'CryptoCurrency'
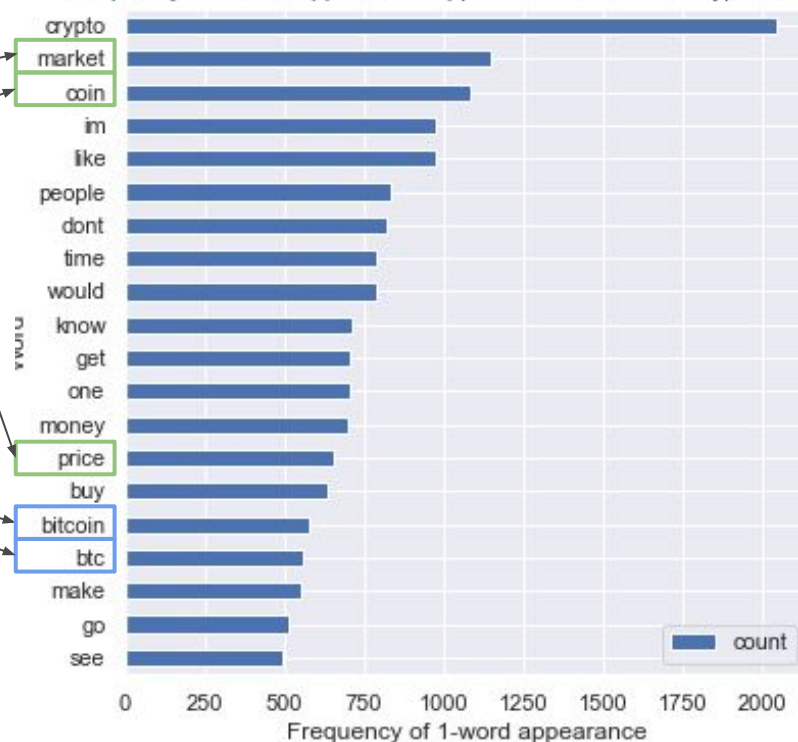
Most words are not a surprise:

- **'market'**: Place that we buy/sell crypto.
- **'coin'**: Synonym to cryptocurrency.
- **'price'**: Indicator on how a cryptocurrency performs.

Some Insights

The most discussed cryptocurrency is Bitcoin

- Words: **'bitcoin', 'btc'**

Frequency of 1-word appearance appeared in subreddit "CryptoCurrency"

2-words appear frequently in 'Stocks'

Most words are not surprise:

- **'price target'**: while buying/selling stocks, people set a price target to rip the gain or to stop loss.
- **'market cap'**: total dollar market value of a company's outstanding shares of stock.

Frequency of 2-word combination appeared in subreddit "Stocks"
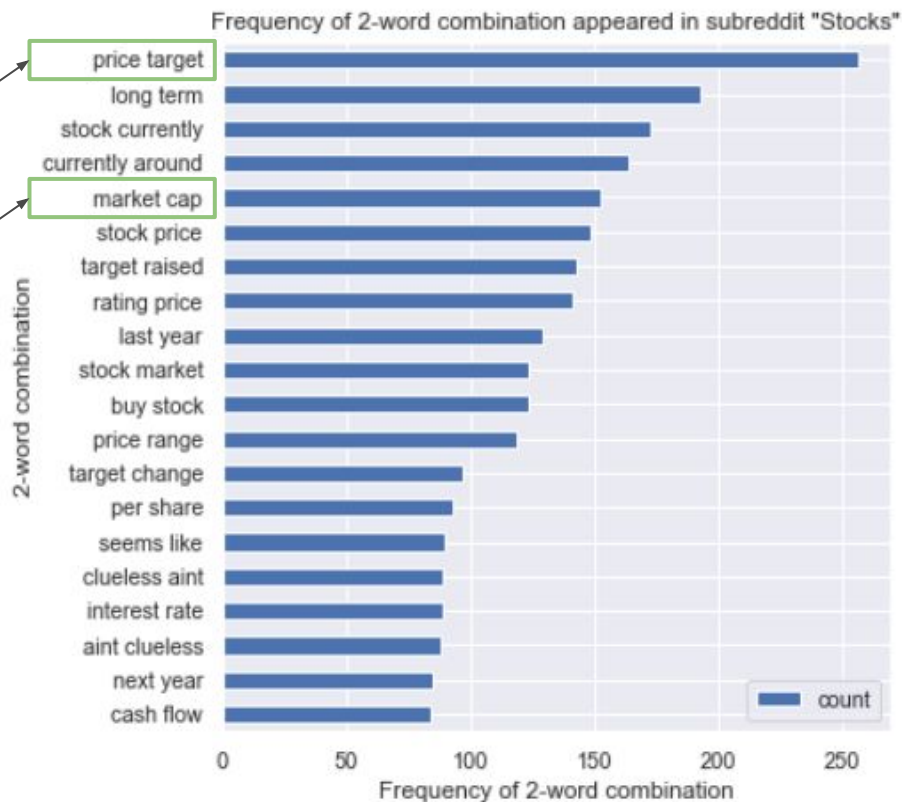
2-words appear frequently in 'CryptoCurrency'

Some Insights

1 Some words indicating the more discussed cryptocurrency
"gtpoplt gtpoplt" - Gifto, PlanetPay
"btc eth" - Bitcoin, Ethereum

2 Mixed sentiments in crypto market due to high volatility
"bull run" + "bull market"
v.s.
"bear market" + "buy dip"

Frequency of 2-word combination appeared in subreddit "CryptoCurrency"
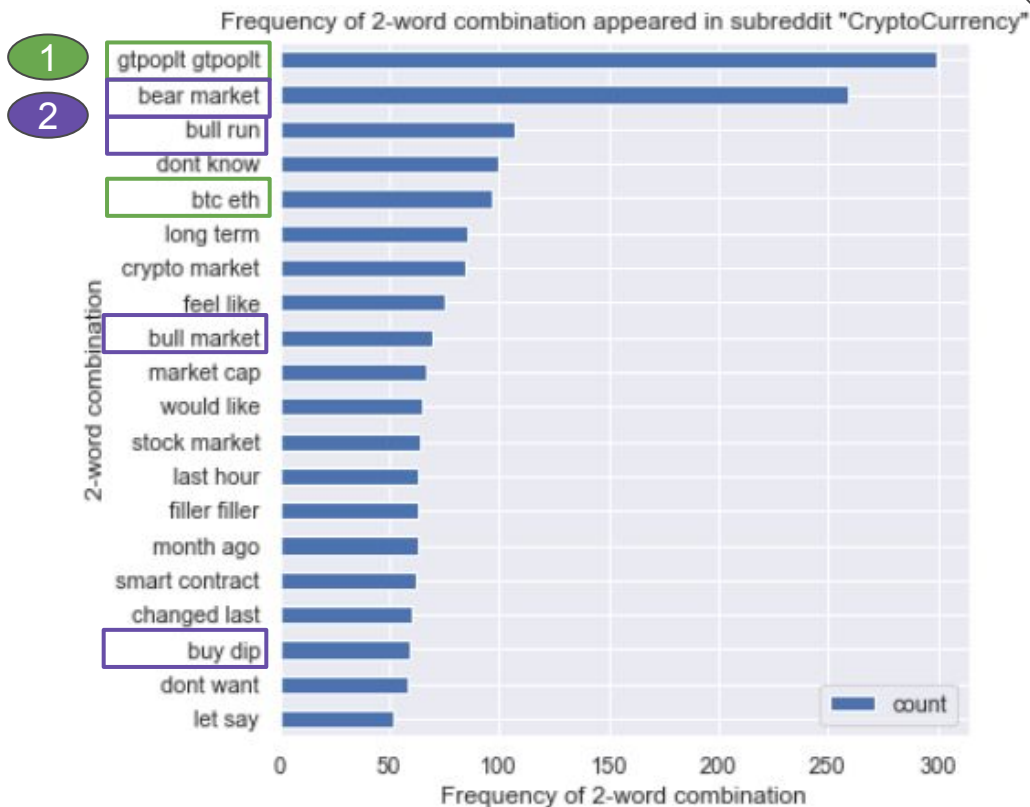
Heatmap for Feature Selection
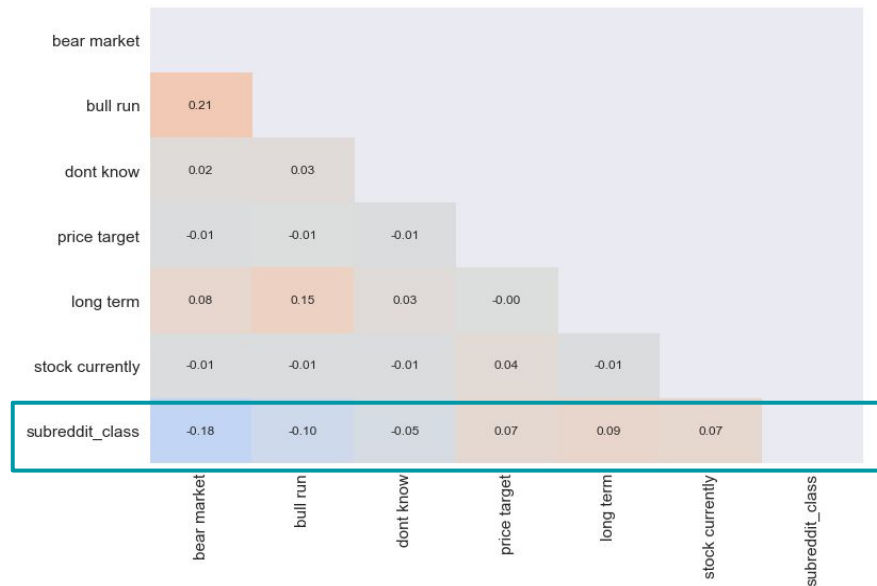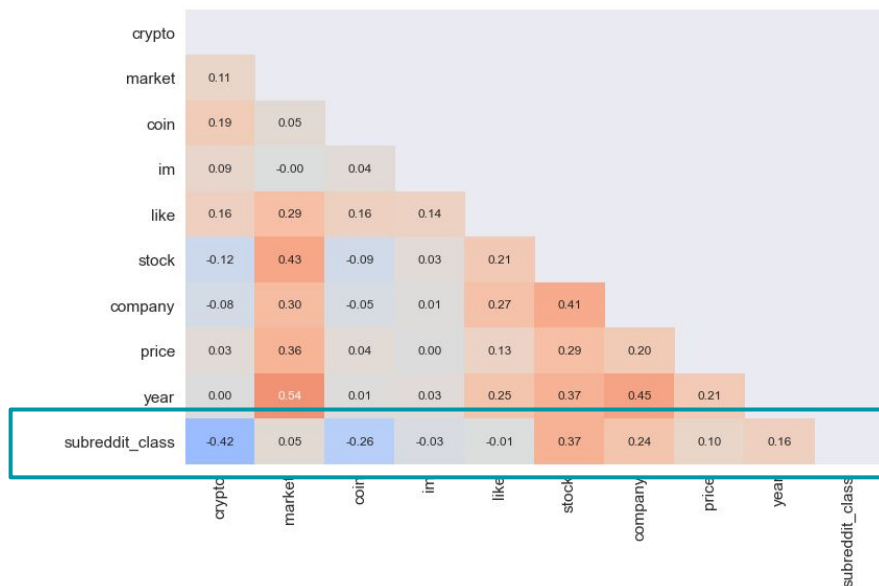
- Single words show a higher correlation than word pairs to the outcome of subreddit class.
- Hence, only Count and TF–IDF vectorizers are used to convert the words to numeric values before model fitting.
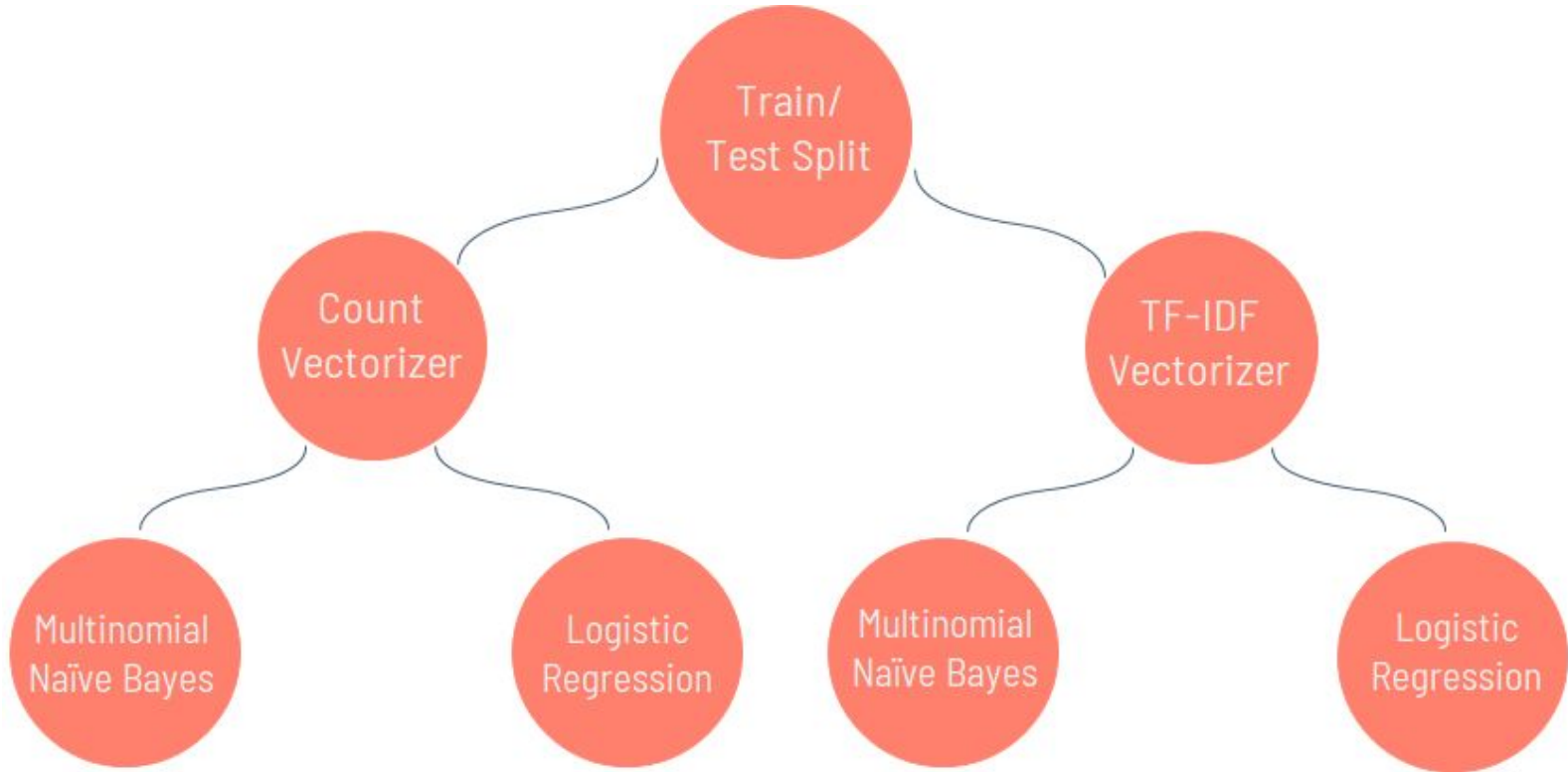
Model Selection for Naive Bayes

Count Vectorizer

NB Model	Train Accuracy	Test Accuracy
Bernoulli Naïve Bayes	76.2%	72.1%
Multinomial Naïve Bayes	96.5%	93.0%
Gaussian Naïve Bayes	97.4%	80.4%

TF-IDF Vectorizer

NB Model	Train Accuracy	Test Accuracy
Bernoulli Naïve Bayes	50.0%	50.0%
Multinomial Naïve Bayes	98.3%	95.0%
Gaussian Naïve Bayes	97.5%	78.7%

Model Preparation

Model Evaluation

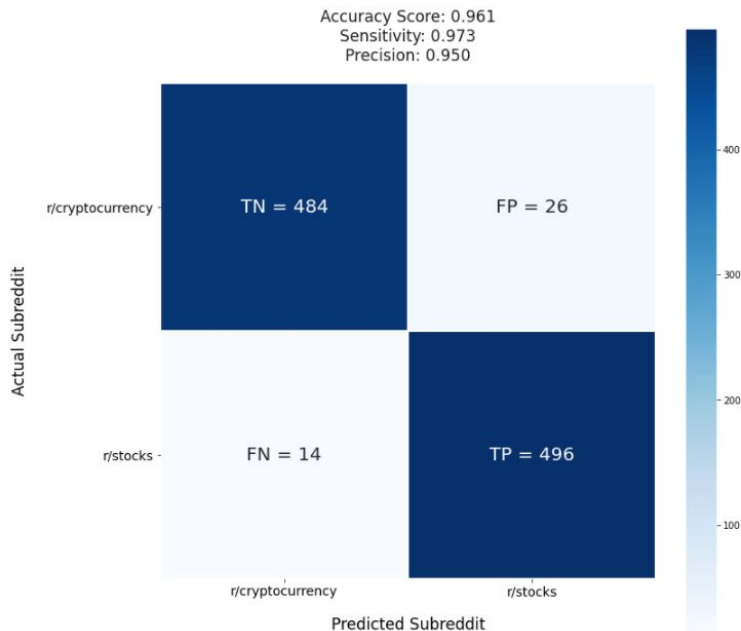| Vectorizer	Classification Model	Accuracy	Sensitivity (Recall)
Count	Multinomial Naïve Bayes	93.0%	87.3%
Count	Logistic Regression	95.6%	95.7%
TF-IDF	Multinomial Naïve Bayes	95.0%	91.1%
TF-IDF	Logistic Regression	96.1%	97.3%

Our model performance is strong

01 <u>Accurate</u> classification **> 96%** of the time; 980 out of 1020 predictions

**02** <u>Sensitive</u> classification ~**97%** relevant post detected

- Out of 510 relevant posts, only 14 will be mis-labelled (~3%)

Confusion Matrix for TF-IDF Logistic Regression Model

Accuracy Score: 0.961
Sensitivity: 0.973
Precision: 0.950

Actual Subreddit

r/cryptocurrency — TN = 484 | FP = 26

r/stocks — FN = 14 | TP = 496

r/cryptocurrency   r/stocks

Predicted Subreddit

Some keywords are strong predictors of a post's relevance to r/stocks or r/CryptoCurrency

Keywords for the classification exercise
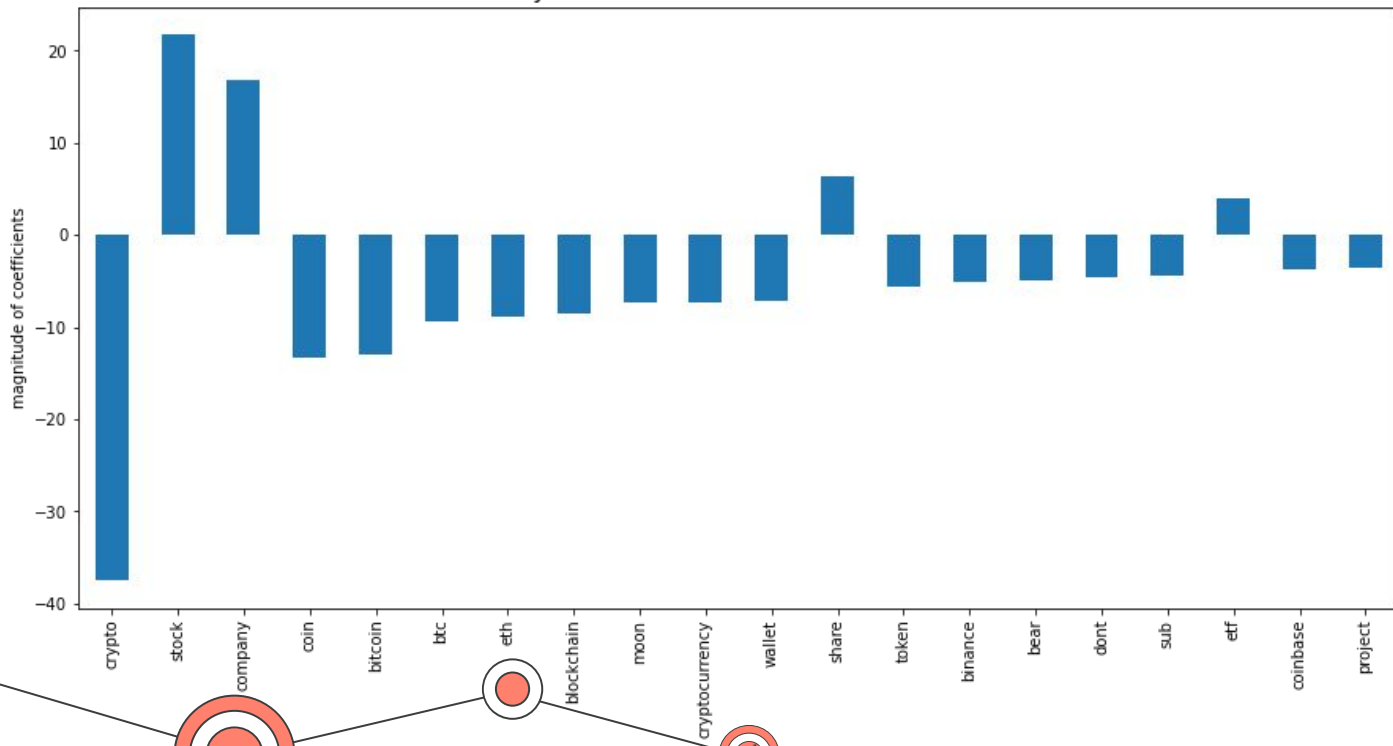
Keywords for r/stocks

Stock, company, share, etf

Keywords for r/CryptoCurrency

Crypto, coin, bitcoin, btc, eth

23

Recommendations

01 Deploy the model for testing with live data

Given the strong performance of the model (96% accuracy score and 97% sensitivity score), we can benefit from deploying the model for pilot testing on the reddit platform.

02 Flag posts classified as irrelevant for moderators' assessment

There is a risk of incurring user frustrations if relevant posts are removed as a result of the model misclassification. To minimize wrongful post(s) removal, we can use the model's recommendation to flag posts for moderators' consideration.

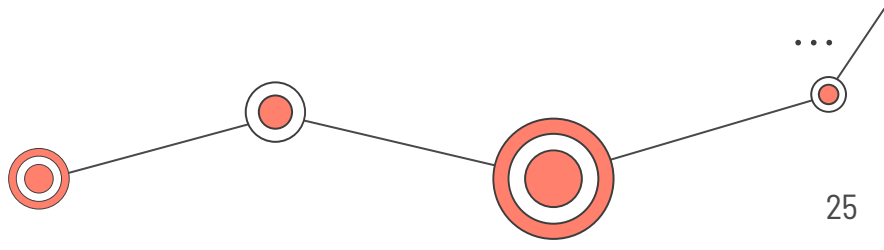## 03 Share keywords insights with the moderators

This can help moderators set rules for content posting using the AutoModerator to ensure irrelevant topics will not be posted within the community.

Future Steps

1. **Collect data samples across a longer time (e.g. the past year) for training and model optimisation** to pre-empt risk that the topics of discussion and keywords may change across different times of the year

2. **Train the data using irrelevant posts from <u>more than one</u> source**

3. **Train and further optimise the model for production across more subreddit communities**

25

Thanks!

Do you have any questions?