Analytics Report: Fundraising Trends in Web3

By Shaan Barca & Shaimay Shah

Table of Contents

Introduction

The task assigned to us was to submit a data analysis report discussing the fundraising trends in web3. To do so, we decided to try and answer two fundamental questions:

1. What are the fundraising patterns over the last 10-12 years?
2. How can the money be allocated more efficiently?

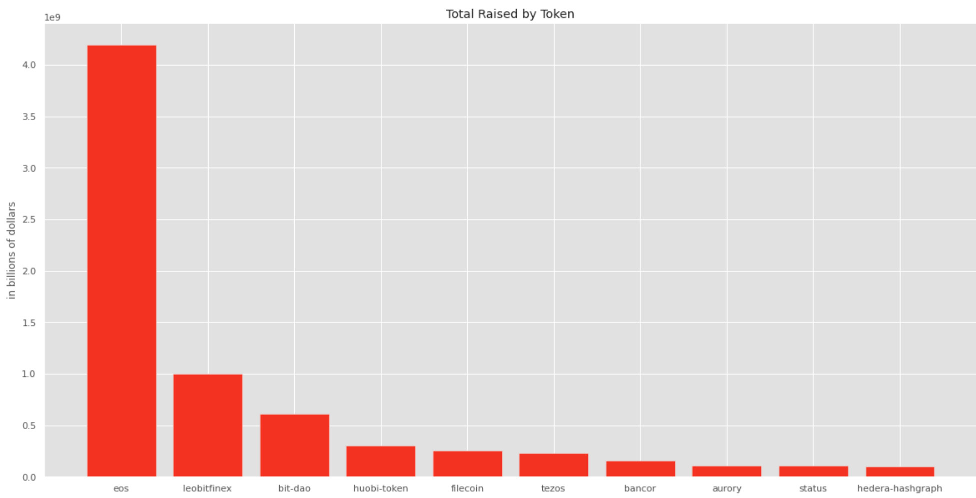
The second question, while not explicitly a part of the task, was an important question we wanted to answer to educate ourselves, and hopefully you, about the kind of tokens and business models to look out for. Of course, the dataset could only tell us what the previous trends were so we used a proxy question — what determines a token's future market cap? To answer this, we built a machine learning model utilizing Random Forest and XGBoost to isolate the actual determiners of future market cap. With the whole market around web3 being extremely vulnerable to hype and FOMO-based investments, we knew that the model wouldn't be as accurate as we'd like it to be but nonetheless we got some interesting results. Before moving on to the analysis part, let's take a look at the kind of data we used.

Data Used

The data we used was collected from Messari's API by Cohurin-san (thank you very much!). The data itself is divided into two datasets — one for qualitative information about tokens and another one for the fundraising details for some tokens. The qualitative information dataset had 500 rows while the fundraising one had 203 rows. Unfortunately, both datasets were filled with an abundance of missing (NaN) values so we had approximately half the dataset to analyze. This is one of the major limitations of our analyses but there are some interesting insights from this limited dataset as well.

Exploratory Data Analysis

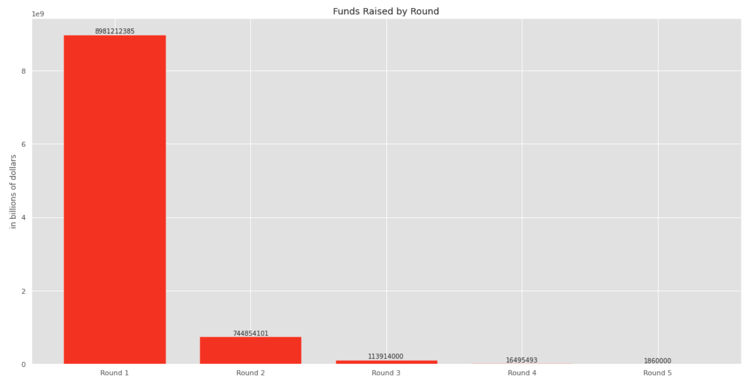
The next few charts focus on the exploratory data analysis we did with regards to the first question.

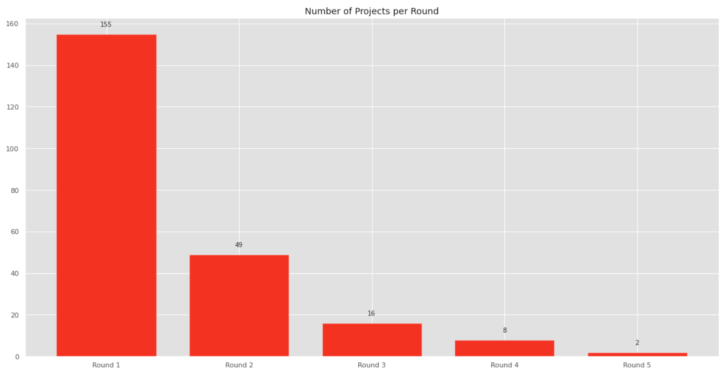


TOP 10 TOKENS BY FUNDING

The most funded token in this dataset is EOS with a little above $4 billion raised through an Initial Coin Offering in 2018 by Block.one supported by investors such as Peter Thiel and Alan Howard. This sets quite a harrowing tone for the rest of the analysis because EOS has ultimately failed — and accused of wash-trading to artificially bolster traded volume and increase its price [1]. The second most funded token — leobitfinex, or Bitfinex — is a cryptocurrency exchange, currently ranked 8th on coinmarketcap by daily volume [2]. However, their journey has not been without major problems either with them having suffered a loss of Bitcoin worth $72 million (worth $3.6

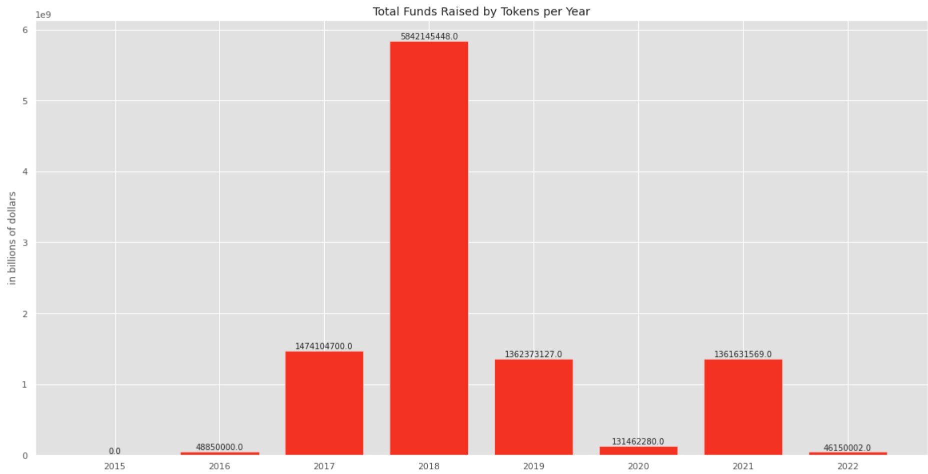
billion in February 2022) due to a major hack. They were also fined $18.5 million by the New York Attorney General's office for concealing losses made by the exchange by using funds gathered for USDT (USD Tether). Some of the other tokens in the top 10 most funded ones also have similar stories, indicating that there is a massive failure in the allocation of funding, while also supporting the theory that most of the investments made in tokens are hype and FOMO-based ones.





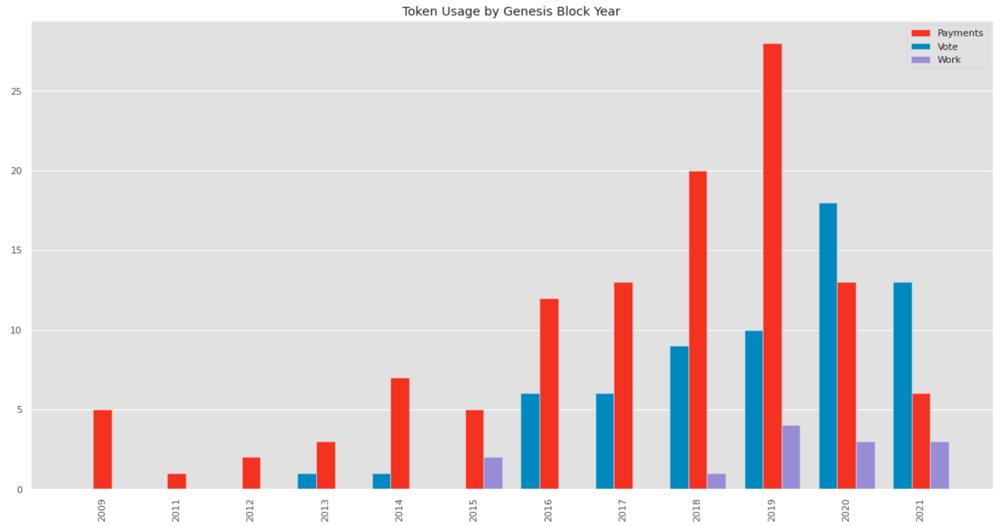
FUNDS RAISED BY ROUND (LEFT) AND NUMBER OF PROJECTS PER ROUND (RIGHT)

Next, we looked at how much was raised per round. As is evident on the charts, the first round is where most, if not all, the funding is concentrated. A guess as to why would be because a lot of tokens don't opt to go for a second round or are bankrupt before the second round. This is supported by the chart on the left showing the steep drop off of tokens in first and second rounds of funding (from 155 projects to just 49).



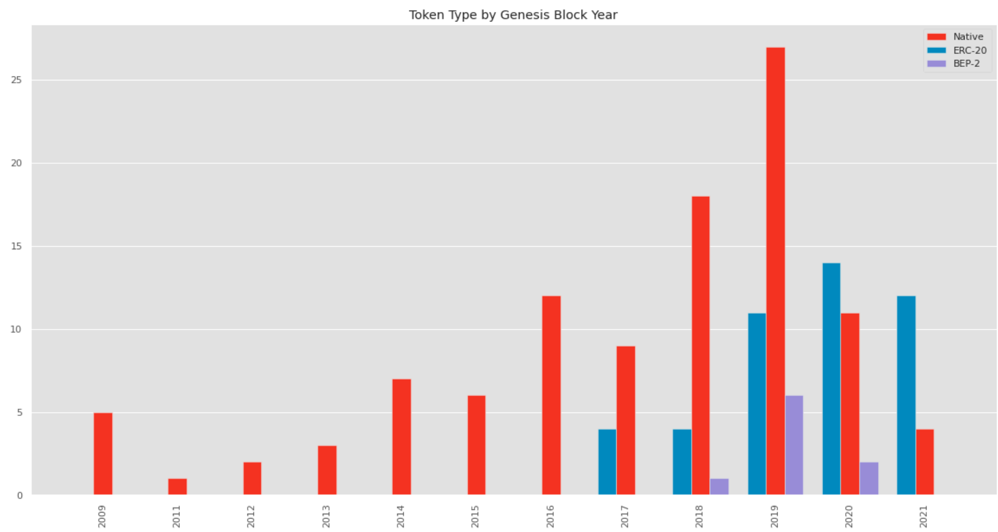
TOTAL FUNDS RAISED PER YEAR

When it comes to the amount of funding received by year, it seems to us that our data is quite incomplete so it becomes hard to arrive at a conclusion from it. One of the noticeable outliers is 2018 where, due to the $4 billion Bitfinex ICO, the amount of money is almost 4 times the next closest year. The impact of COVID-19 can also be seen in 2020 with only $131 million being funded (according to our dataset). The funding received by token projects in 2021 were ten times that of 2020 and it seems that 2022 is also on a similar trajectory.



TOKEN USAGE BY GENESIS BLOCK YEAR

One of the more interesting charts is the token usage chart over the years. While the dataset had 6 different token usages defined, we opted to include the three most common ones in our analysis — payments, voting and work. Looking at the trend of token usage could serve as one of the indicators of where the funds are being allocated and which kinds of projects have investor confidence. Moreover, it can also indicate where the web3 industry is moving — and this shows that in recent years, there has been a shift in tokens being created solely for usage in payments to more DAO-based ones where each member has a vote in the future of the organization. Could this also mean that the industry thinks that they have found a suitable solution to the payments problem (through an Ethereum-based model) or has the industry abandoned the “next financial system” tagline that Bitcoin and earlier cryptocurrencies were based on?



TOKEN TYPE BY GENESIS BLOCK YEAR

Funnily, our question is partially answered in the next chart showing the three most prevalent token standards that tokens launch as (ERC-20, BEP-2 and a native token). Until 2019, it seems that most new tokens were launching with their own standards (i.e. they were native) after which ERC-20 seems to have taken off. This coincides with the change in the token usage chart as well with the number of new tokens launching with the official usage of payments was at its peak but dropped to about half that in 2020 with the usage of voting taking off. It does support the theory that the industry has arrived at the conclusion that at least for now, the ERC-20 standard seems to be the best token creation standard.

Model

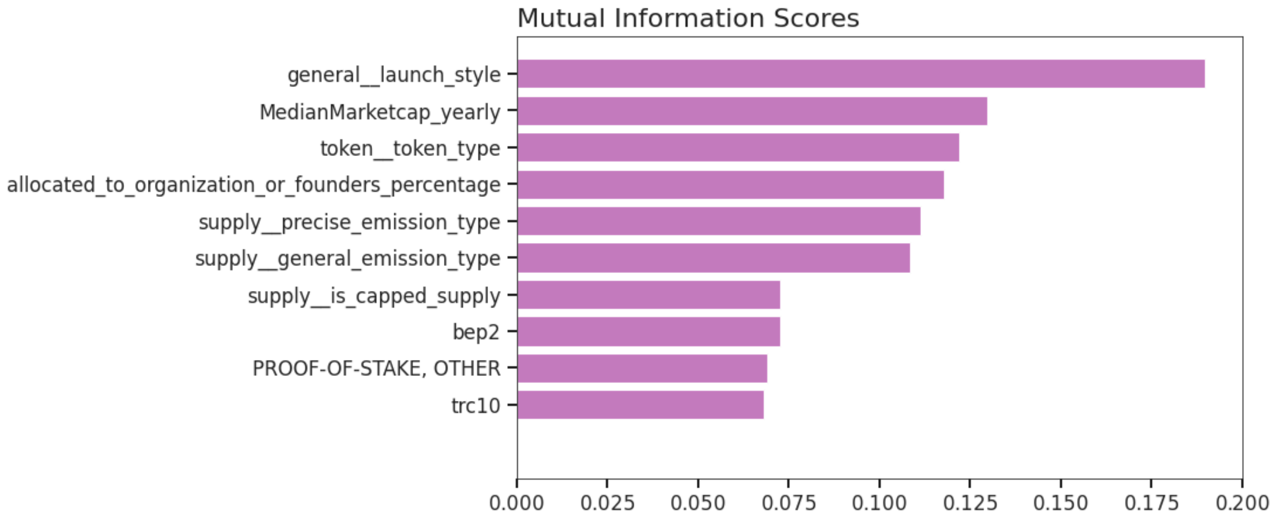
Feature Engineering

One feature that we create is the Median market cap of coins in that year. The feature was created as we suspect that launching a coin in the bull or bear market can make a significant impact on a coin's future market cap. Variables that were categorical, such as launch style or token type, were transformed to dummy variables. We also do some descriptive statistics to check the data we are dealing with. In this particular case, we will be using it to categorize the market cap of tokens into separate categories. Using percentiles, we can bin the market cap of tokens into low, medium, high and very high. Coins with a market cap less than the 25th percentile was low, greater than 25 - 50 and below 50 was medium, greater than 50 and below 75 was high and greater than that was very high.

	count	mean	std	min	25%	50%	75%	max
current_marketcap_usd	209.0	6.101972e+09	5.044152e+10	3118475.5	27338239.1	120201552.5	5.507598e+08	6.596436e+11

Mutual Information

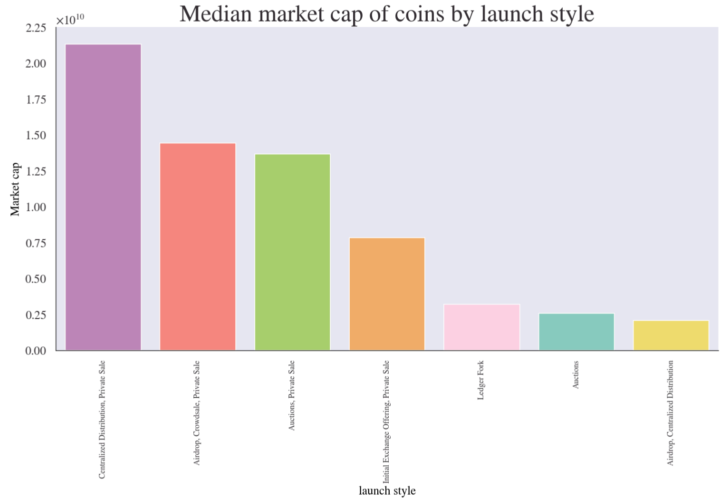
When dealing with data with a large amount of features it is beneficial to identify which are the potentially key features. To do so, we calculate the mutual information scores. What is measured is how much information is lost to predict/classify our target variable when a feature is removed. There are advantages of using mutual information over Pearson correlation as it is not limited to identifying just linear relationships.



When we calculate Mutual information(MI), we see that the model has narrowed down these few key features as the most important to determining a coin's future market cap. One of the limitations of MI is that it can only identify univariate relationships. Meaning, certain variables that might not be helpful on their own but are useful when combined with other variables are not captured. We will do further EDA on the features that our model has identified.

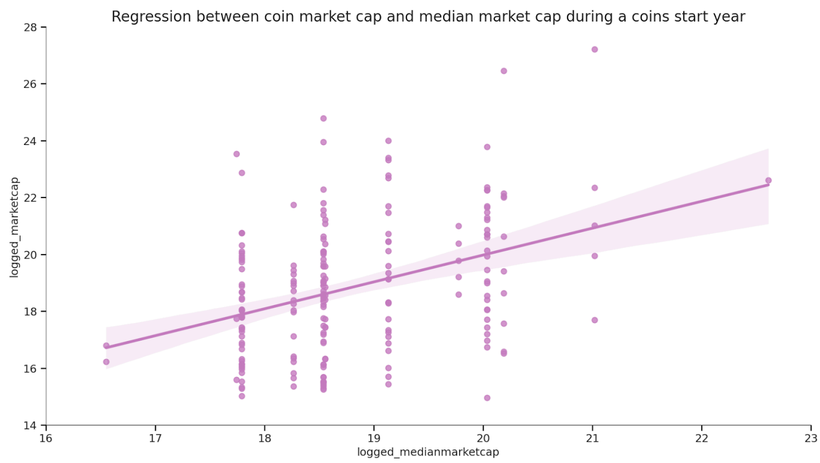
Model EDA

General Launch Style



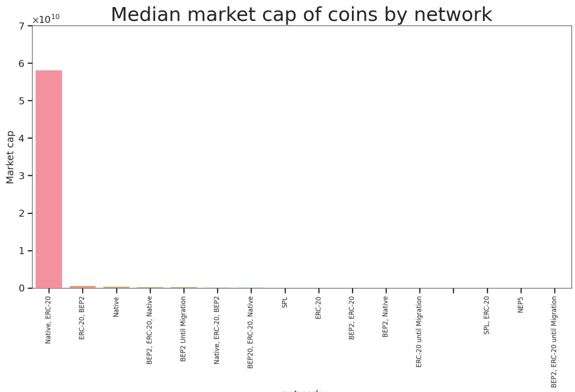
A coin's launch style also seems to be correlated to its future market cap. Those that have a Centralised distribution and Private sale launch style tend to be worth more as opposed to other launch styles. This could be that coins that were most centralised (initially) were easier to patch and update compared to those that were already decentralised from the beginning.

Median Market Cap During a Coin's Start Year (Logged)



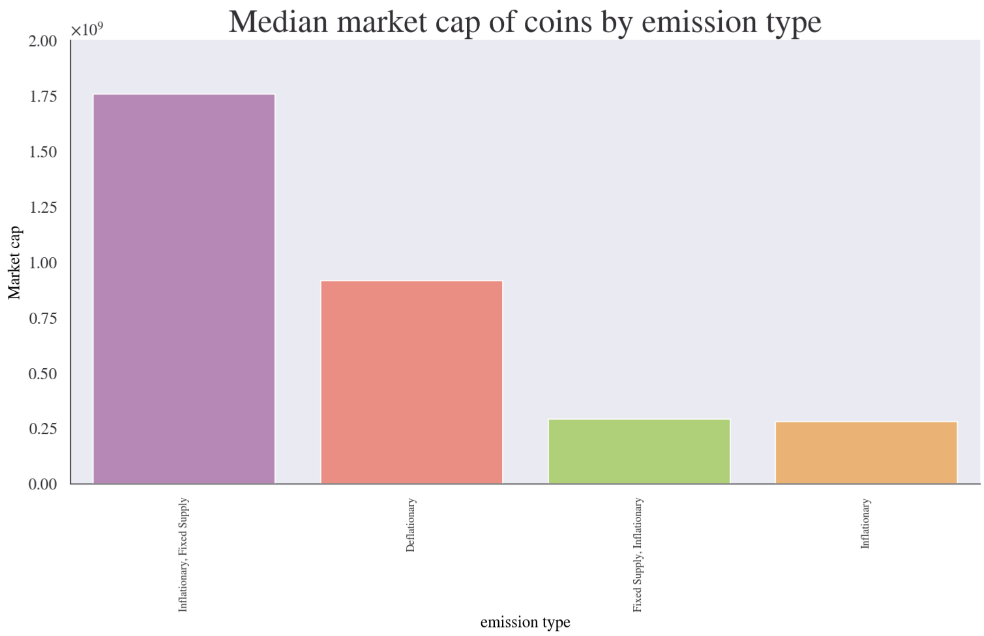
We can see that there is a relationship between the median market cap of coins of a start year to its current market cap value. This could be due to coins that are launched during the bull market benefit from the hype and tend to be valued more than those that launched during a bear market.

Median Market Cap by Token Standard



From the data provided, we can see that coins that have both ERC-20 and Native token type have the highest median market cap. There are a few potential reasons for this. The ERC-20 standard is a widely used one due to a large support base and wide variety of flexibility offered by the token standard. It can also be used by most wallets and DEXes.

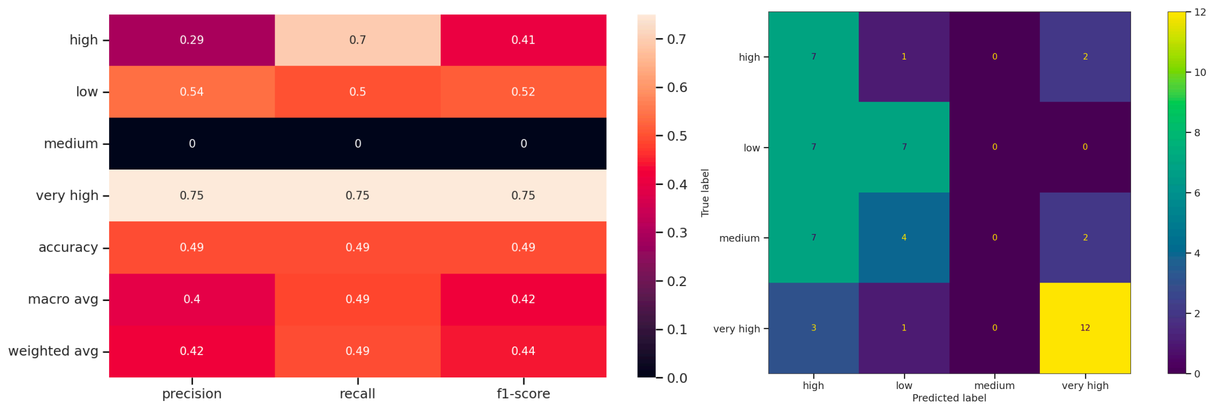
Median Market Cap by monetary policy



We can see that coins that are Inflationary with fixed supply (which will eventually be deflationary) and deflationary coins tend to have a higher market cap. This could be that the deflationary aspect of these currencies create scarcity which no longer exists in fiat currency where the central bank can print currencies as it likes. Both, anti-capitalist and hyper-capitalist movements have had this criticism of the central bank model and so this may be one of the factors making it appealing to potential investors.

Model Creation

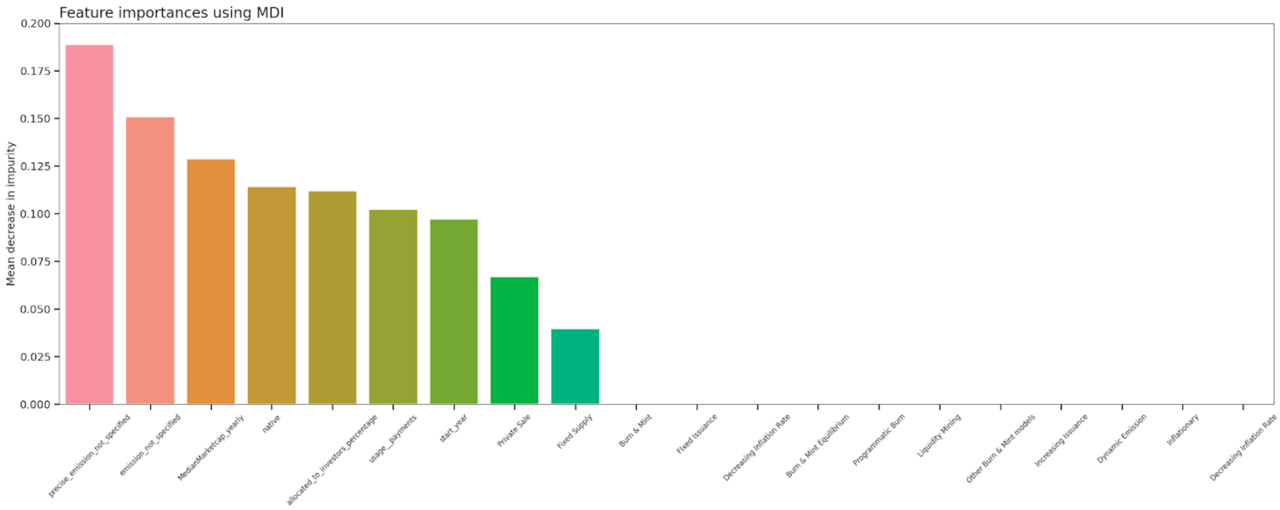
We use both Random Forest and XGBoost to predict a coin's future market cap. We utilized grid search to find the optimal hyper parameters for our XGBoost model.



CLASSIFICATION REPORT

The figure on the is the confusion matrix produced by our model.We then see that while our model only had a score of 49%, its F1 score for classifying coins that have a very high market cap was 12/16 or 75%. Our model particularly struggled with correctly classifying coins that have a medium market cap.

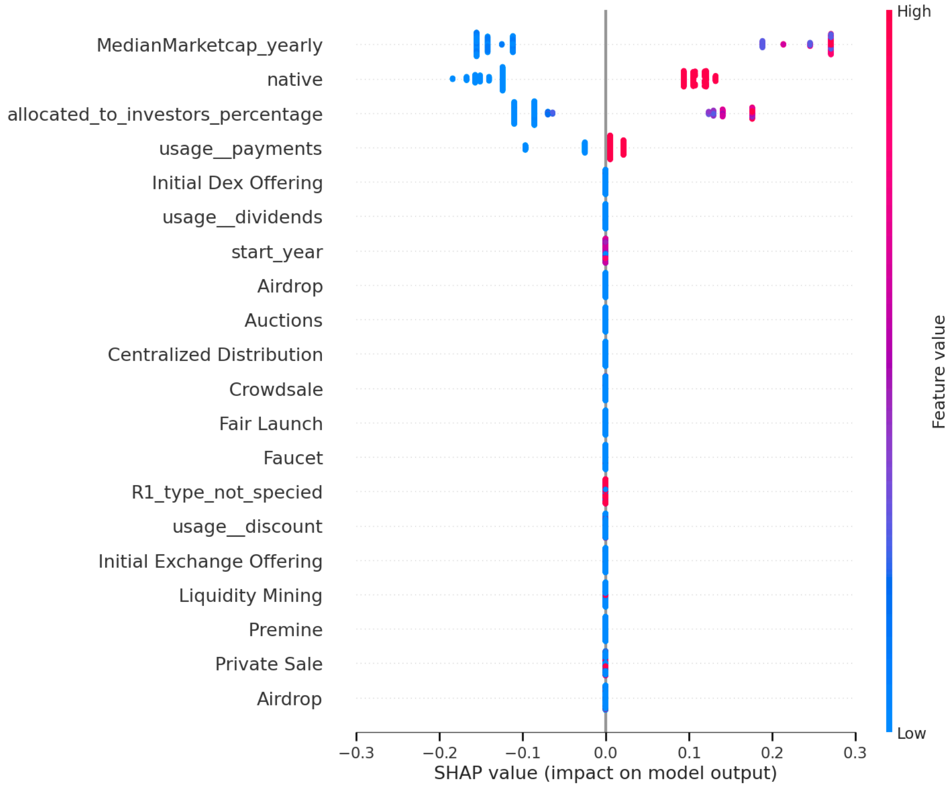
Feature Importance



FEATURE IMPORTANCES USING MDI

Looking at the feature importance according to our XGboost model. We can see that emission type, token type, median market cap and usage are key features to determining a coin's market cap. We can explore how each feature affects the output that our model produced by utilizing SHAP.

How our model decides to classify a token with very high market cap



SHAP is a library used to explain ML models. It uses a game theory approach to see which features are most important and how each feature impacts our model.
The y-axis is the features names and are arranged based on feature importance. Our x-axis shows how a value of that feature impacts the output of our model. The color of the data point indicates the range value of a particular sample's feature. In this plot we see how a model classifies a coin to be very high. We see that if the median market cap of the start year of the coin is high it is usually going to have a higher market cap. If it has its own native protocol it also increases the probability. The higher the percentage that was allocated to investors also increased its probability.

Conclusions

With regards to the first question, it is becoming evident that while overall funding seems to be back to pre-COVID values, the destination of these funds are going into tokens that follow a certain standard (i.e. ERC-20 or BEP-2) and are being used for an interesting potential use case for cryptocurrency tokens — voting. In broader terms, it could mean that the industry has decided to move onto the democracy part of cryptocurrencies' promises.

In regards to our findings produced by our models, we can see that the timing of launching the ICO is important. During bear markets, when the median market cap of that year is low. Investors are more cautious when investing. There is higher due diligence and overall less capital to go around. This may lead to a coins launching ICO's that year to garner less hype and less awareness at least in the short terms (5 years <). Whether or not coins have a native protocol also matters. This could be due to investors interpreting this as the team actively trying to address issues present with cryptocurrencies protocols by creating their own, which can also mean they have a high level of expertise and are proactively trying to solve such issues instead of solely using existing protocols. Coins that have a higher percentage that are allocated to investors tend to have a high valuation. This could be that these tokens are backed by VC's and tend to be more credible. Tokens that are ***initially*** more centralized may also benefit from faster software updates as it's easier to achieve consensus. Finally, it's utility such as for payments would also increase an ICO's values as opposed to more community based tokens, though this impact is relatively low.

Limitations and Suggestions for Future Research

One of the advantages of our dataset was that it was balanced in the sense that we had a relatively equal amount of samples that were had a low market cap all the way to very high. To be very cautious with the capabilities of our model, we also included a confusion to show how it performed for classifying each class. As shown in the classification report and accuracy, our model has has an overall accuracy of 49% but has an F1-score of 75% for classifying coins with a "very high market cap". In the mutual importance and feature importance section, we listed what variables was most important, however, within our dataset for quite a number of coins the details such as consensus mechanism, emission type etc were not specified. This might have caused our model to undervalue its true importance. Several potentially useful variables such as team experience, whether a white paper was included, lock up period for founding teams etc were not included. The exclusions of these variables may have also undervalued the importance of certain features as certain features might not be helpful on their own but useful when combined with others.

References

[1] "New research claims 21 accounts pumped the $4.4B EOS ICO with wash trades," *Cointelegraph*. https://cointelegraph.com/news/new-research-claims-21-accounts-pumped-the-4-4b-eos-ico-with-wash-trades

[2] "Top Cryptocurrency Exchanges Ranked By Volume," *CoinMarketCap*. https://coinmarketcap.com/rankings/exchanges/

[3] R. Browne, "Cryptocurrency firms Tether and Bitfinex agree to pay $18.5 million fine to end New York probe," *CNBC*, Feb. 23, 2021. https://www.cnbc.com/2021/02/23/tether-bitfinex-reach-settlement-with-new-york-attorney-general.html