Final_Report

May 24, 2022

1 Ape Minders: Trading Cryptocurrency on NFT Signals

By Andreas Michael, Michael Wander, Daria Vasileva, and Amit Kumar

1.1 Introduction

The market of non-fungible tokens, or NFT's for short, has boomed over the past two years, with interest and trading at an all-time high. Although the hype cannot last forever, the effects of NFT's on cryptocurrency as a whole should not be dismissed. The most popular and most expensive NFT collections are generally based on the Ethereum blockchain, mainly through Polygon on OpenSea, but marketplaces on other blockchains now exist for the trading and minting of NFT's, including Solana, Cardano, Avalanche, Decentraland, and others. With this in mind, we aim to trade various cryptocurrencies, from large to small-cap, which are directly or indirectly related to the trading of NFT's, on signals produced from NFT data, such as number of sales, active wallets, buyers, and sales in USD, as well as sentiment and polarity data of text posts related to NFT's. We presume that NFT related data correlates well with our chosen universe of cryptocurrencies, which should allow for an effective trading strategy.

1.2 Market Universe

1.2.1 Instruments

For our universe, we choose seven cryptocurrencies which are directly or indirectly related to NFT's and their marketplaces. All but two of these coins are based on their own blockchains. The two which are not, Polygon and Decentraland, are based on the Ethereum blockchain. The cryptocurrencies we choose range from large to small-cap in terms of market capitalization, which allows for a good spread of the volatility of altcoins. The instruments, along with their approximate market caps as of April 2022 are as follows:

- Solana (SOL), $44B
- Cardano (ADA), $39B
- Avalanche (AVAX), $26B
- Polygon (MATIC), $11.5B
- Decentraland (MANA), $5B
- Flow (FLOW), $2.5B
- Theta Fuel (TFUEL), $1.1B

The OHLCV and market cap data for these instruments is sourced from https://coincodex.com/. The date range that we use for zipline ingestion is from 2017-10-02 to 2022-04-17, but backtesting, model training, and strategy evaluation is done on the range from 2021-09-19 to 2022-03-20.

1.2.2 Exchange

For trading and transaction cost purposes, we will suppose that our trading is done on the Binance cryptocurrency exchange, despite being banned in New York. We choose Binance because we can trade all of our seven cryptocurrencies in our universe while incurring a transaction cost of 0.1% on each trade. This assumes that we convert all of our capital into Tether (USDT) before trading on Binance. The spread for each transaction in terms of slippage is 0.2%.

1.2.3 Benchmark: AMI index

A capitalization-weighted index was created to compare our algorithmic strategy against. The index consists of the 7 instruments that we are trading. We deploy a buy and hold strategy with it. This index is being rebalanced every quarter.

1.3 NFT Features: Quantitative Data

Quantitative and numerical features relating to NFT sales and collections are downloaded from the market tracker of https://nonfungible.com/. The data comes as a .csv file which contains 10 features, which are number of sales, sales USD, average in USD, active market wallets, primary sales, secondary sales, primary sales USD, secondary sales USD, unique buyers, and unique sellers. The collections and markets that the data pulls from are mainly ETH-based, though they still show correlation with the instruments of our universe.
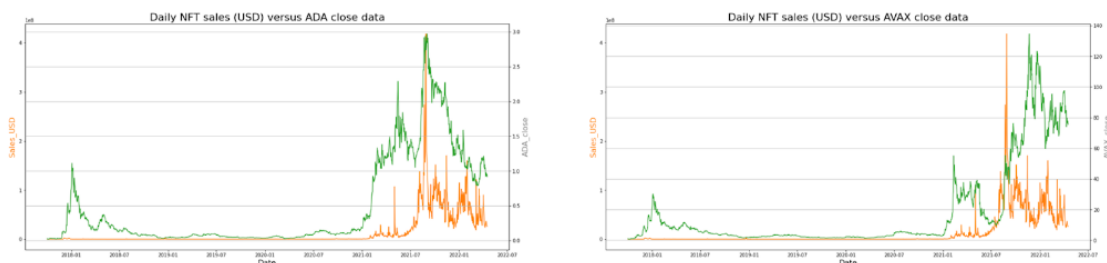
1.4 NFT Features: Alternative Data

Online sentiment about NFT's, and cryptocurrency in general, may contribute to a lot of the hype around the trading of these assets. Some notable places for discussions include Twitter, Reddit, Telegram, and Discord. For our collection of posts related to NFT's, we choose Reddit, specifically the r/NFT subreddit where many different NFT posts are made per day. This subreddit was created in 2016, but has historical posts dating back only to 2019, which does not affect our analysis.
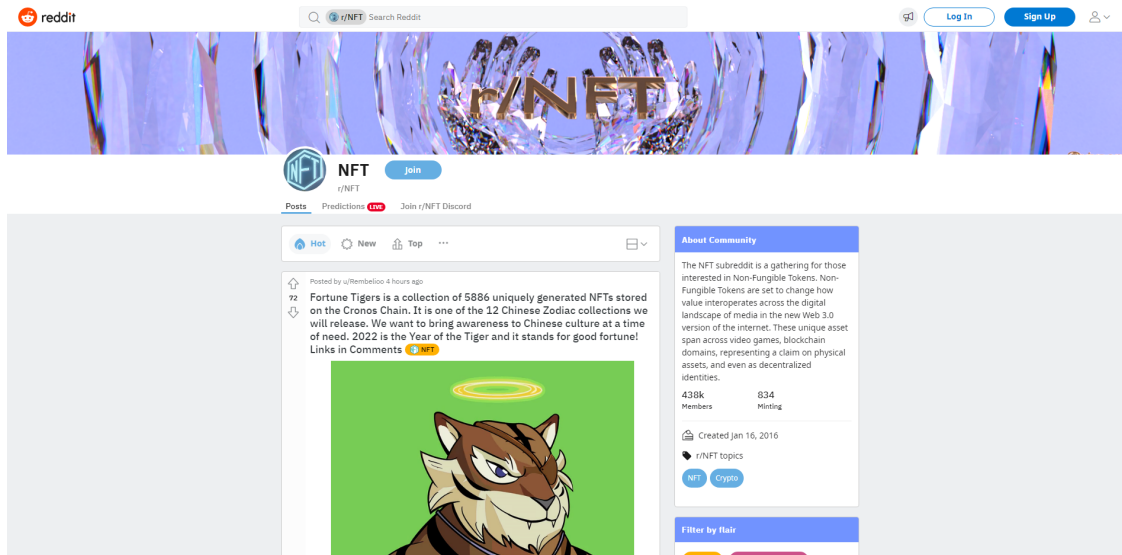
To collect the posts from Reddit, the Python Reddit API Wrapper (PRAW) and the Pushshift API Wrapper (PSAW) libraries are used. PSAW is only used due to the 1000 post limit per API call, and the inability to use date ranges to request posts, of the base Reddit API. Using these two libraries, around 300,000 posts from the r/NFT subreddit, from 2019 to 2022, are collected. For analysis, we only collect main posts and not comments or reply trees. These main posts usually do not have any text apart from their title, since most of them are memes, pictures, or other media related to NFT's. Due to this, the title of each post, as well as any post text that exists, are combined for each post and used for seniment analysis.

Date	Title	Selftext
2019-02-28	A NonFungible Token Stampede Is Coming – Coinm...	NaN
2019-02-28	Decentraland Creator Contest - Building on NFT...	NaN
2019-02-28	NFTY News	NaN
2019-02-28	'No Wallet Needed': Mobile Cryptopunk Game	NaN
2019-07-06	Yat Siu - NFT is why content will be king again	NaN

The sentiment analysis is done using SpaCy and the spacytextblob pipeline component. From each post title and text, the polarity is produced and collected as a column in the same dataframe.

Date	Comb	Polarity
2019-10-16	BLVD Collectible Map Styles : Season 2 is here!.	-0.625000
2019-12-13	Formula0x : One of a kind F1 inspired custom l...	0.600000
2020-03-24	NFTs site for info and promotion. [http://www....	0.600000
2020-04-28	Crazy Defense Heroes NFT collectibles.	-0.600000
2020-04-30	Are you ready for the TrailBlazer? Got your St...	0.604167
...	...	...
2022-04-17	Nansen Report: The non fungible tokens' market...	0.625000
2022-04-17	PROUD LIONS CLUB 20% DISCOUNT ON WHITELIST SAL...	0.800000
2022-04-17	GIVEAWAY FACE X TRY TO WIN 100 WL.	0.800000
2022-04-17	giving my welcome with this donkey spirit.	0.800000
2022-04-17	"What? You are collecting jpeg son? Absurd!".	-0.625000

To aggregate the polarity for each day, an aggregate sum and mean are taken of the polarity of each post associated with their respective day. Then, a threshold based on the polarity sum column is created, which produces either a label of -1, 0, or 1, for overall negative, neutral, or positive sentiment for the day. The threshold for the negative label is less than 0, and the threshold for the positive label is greater than 100.

| Date	Polarity_Sum	Polarity_Mean	Sentiment
2021-01-01	0.89	0.13	0
2021-01-02	1.37	0.11	0
2021-01-03	1.36	0.12	0
2021-01-04	1.42	0.08	0
2021-01-05	1.96	0.18	0
...	...	...	...
2022-04-13	69.98	0.15	0
2022-04-14	66.05	0.13	0
2022-04-15	61.89	0.14	0
2022-04-16	58.31	0.13	0
2022-04-17	66.96	0.15	0

While analyzing and collecting the sentiment data, we noticed that most of the posts, and therefore days, have neutral sentiment. This may be because we did not do any cleaning or preprocessing of the posts gathered from the subreddit. Many posts are just people writing about various NFT giveaways they claim to be doing, or other things that would not indicate positive or negative sentiment necessarily. Nevertheless, the sentiment data is used as is.

1.5 Regression Models for Predicted Returns

The trading algorithm for our universe is based on predicted forward returns of one day. To predict these returns, we will use a regression model with the previously described NFT features, as well as some general technical indicators for the close prices of our instruments. These indicators are the

RSI, Bollinger Bands (high and low), Average True Rate, and MACD (technical factor methods referenced from *Machine Learning for Algorithmic Trading*, Chapter 7). Three simple models are chosen for comparison and evaluation of the predicted returns: OLS, random forest, and XGBoost. All three models are trained on the same input features, which include the technical indicators and the NFT features. No hyperparameters are changed, and the input features are not scaled or normalized. In total, there are 17 input features: 5 technical features, 10 quantitative NFT features, and 2 sentiment NFT features. The target variable are the one-day forward returns of the close price of each coin.

The three models are all trained separately on each coin rather than all at once on the multi-indexed dataframe, due to issues with accurately setting the date range of the training, testing, and backtesting data. The feature importance and coefficients of the models are also calculated and visualized in the figures below. It seems that all three models use some or all of the quantitative NFT features, but the sentiment is only used slightly in the OLS model. The technical indicators do make up a bulk of the feature importance.

Linear Regression Coefficients

Random Forest Feature Importance


Random Forest Feature Importance
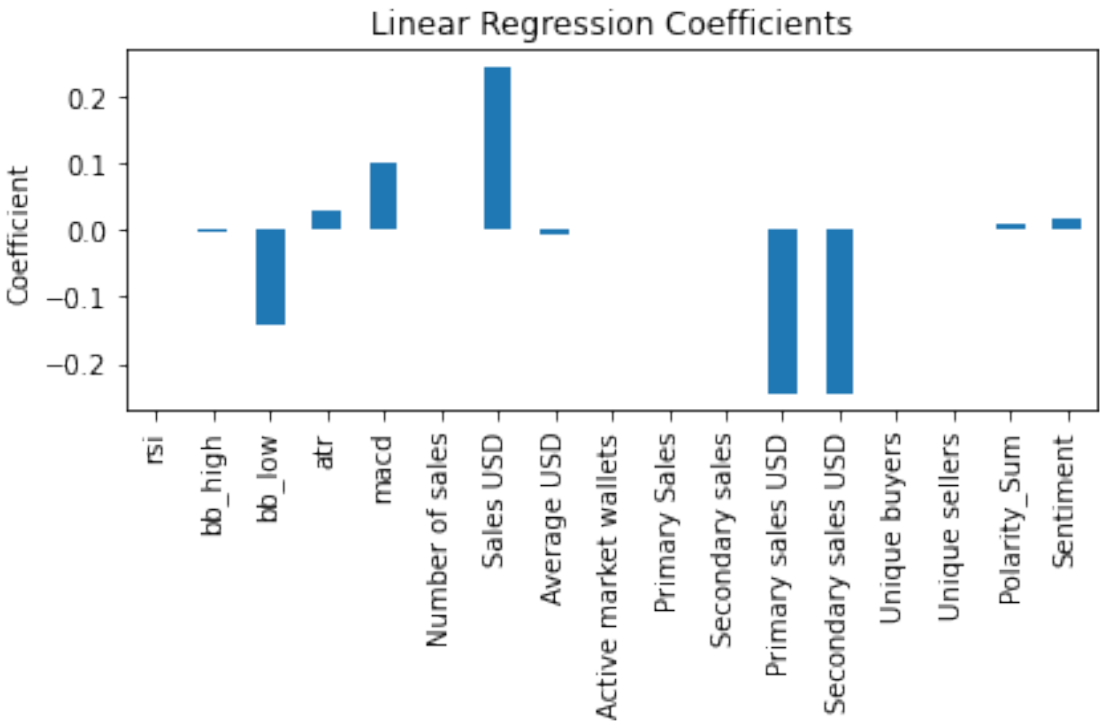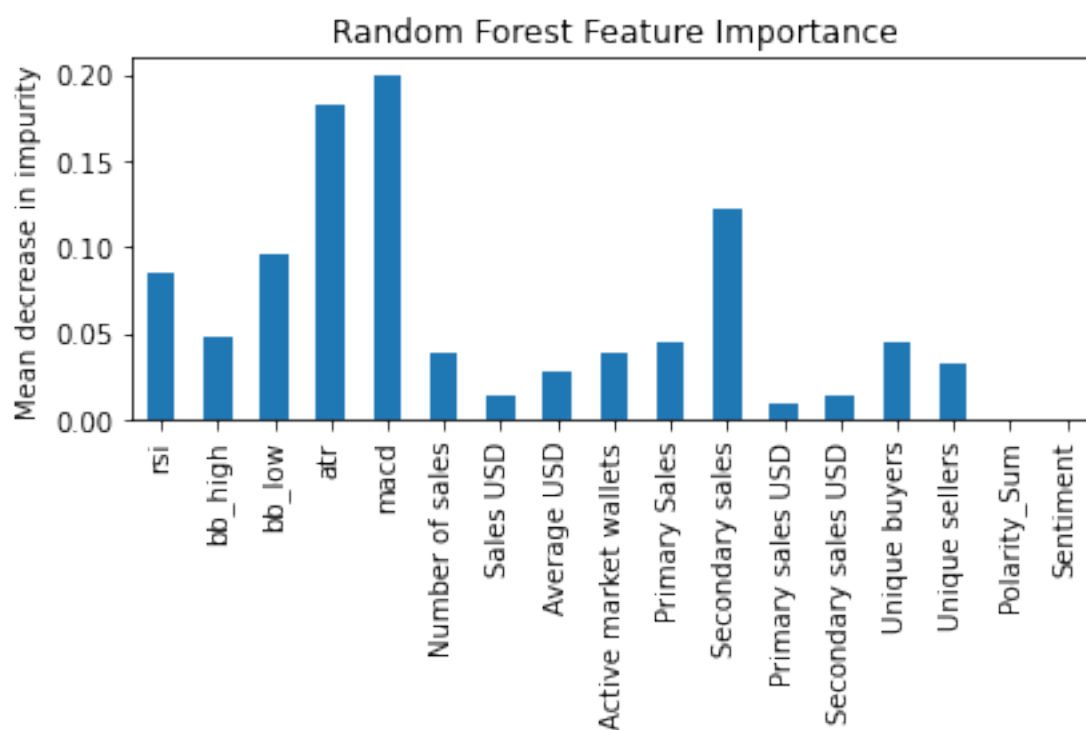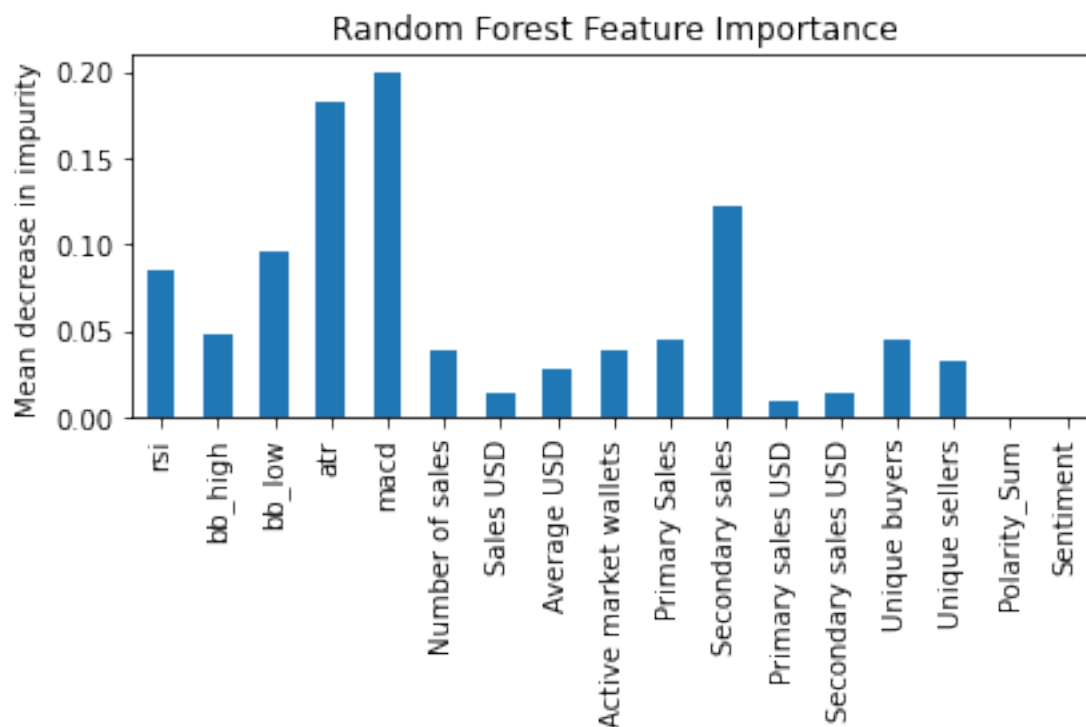
For evaluation, the MSE and MAE are calculated for each coin, then averaged together and compared. Overall, the random forest model appears to have the lowest error, though for Cardano the OLS model has a lower error.
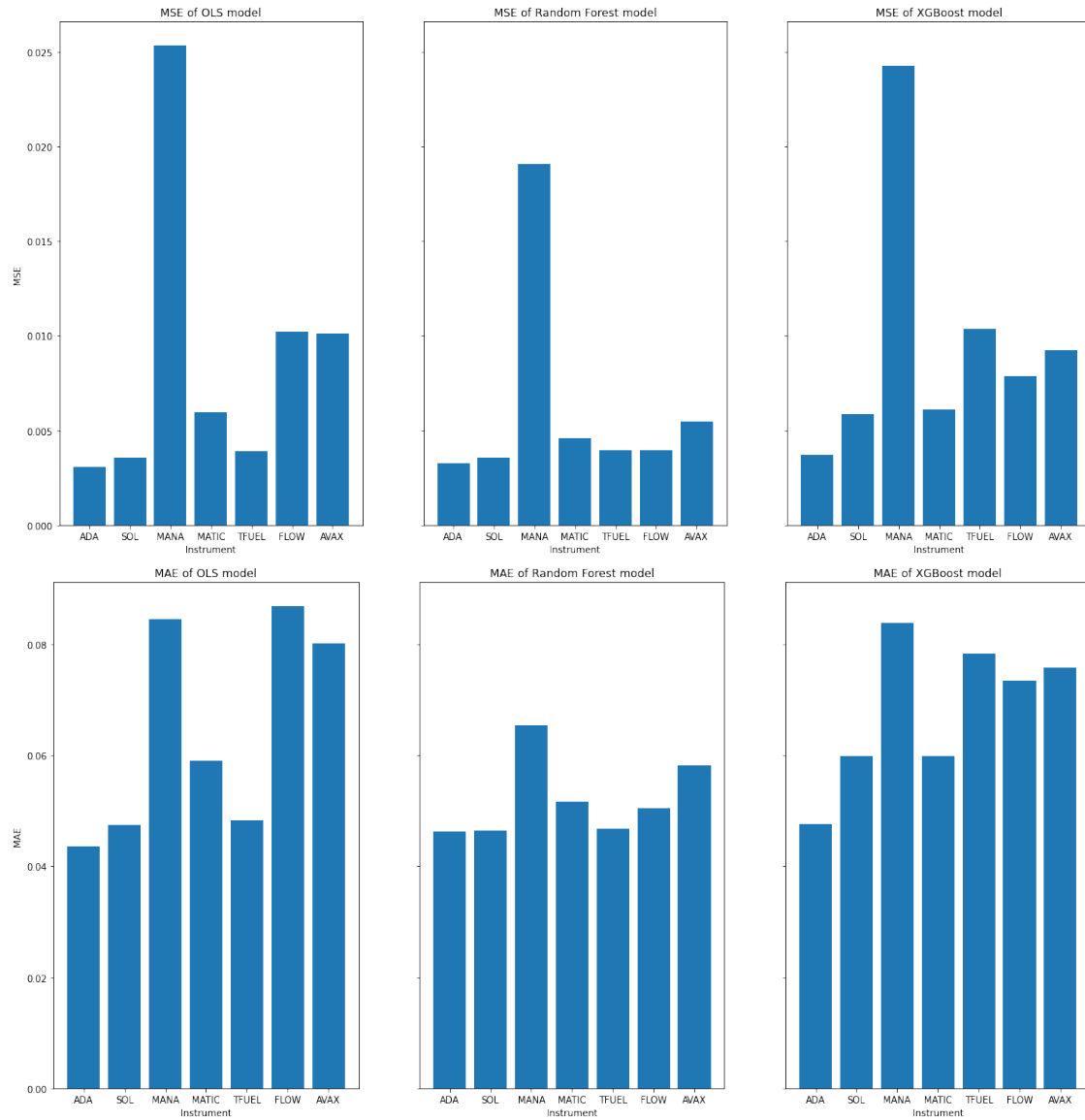
On average, the random forest has the lowest error, followed by the OLS model and the XGBoost model.

Average MSE of Models

Average MAE of Models
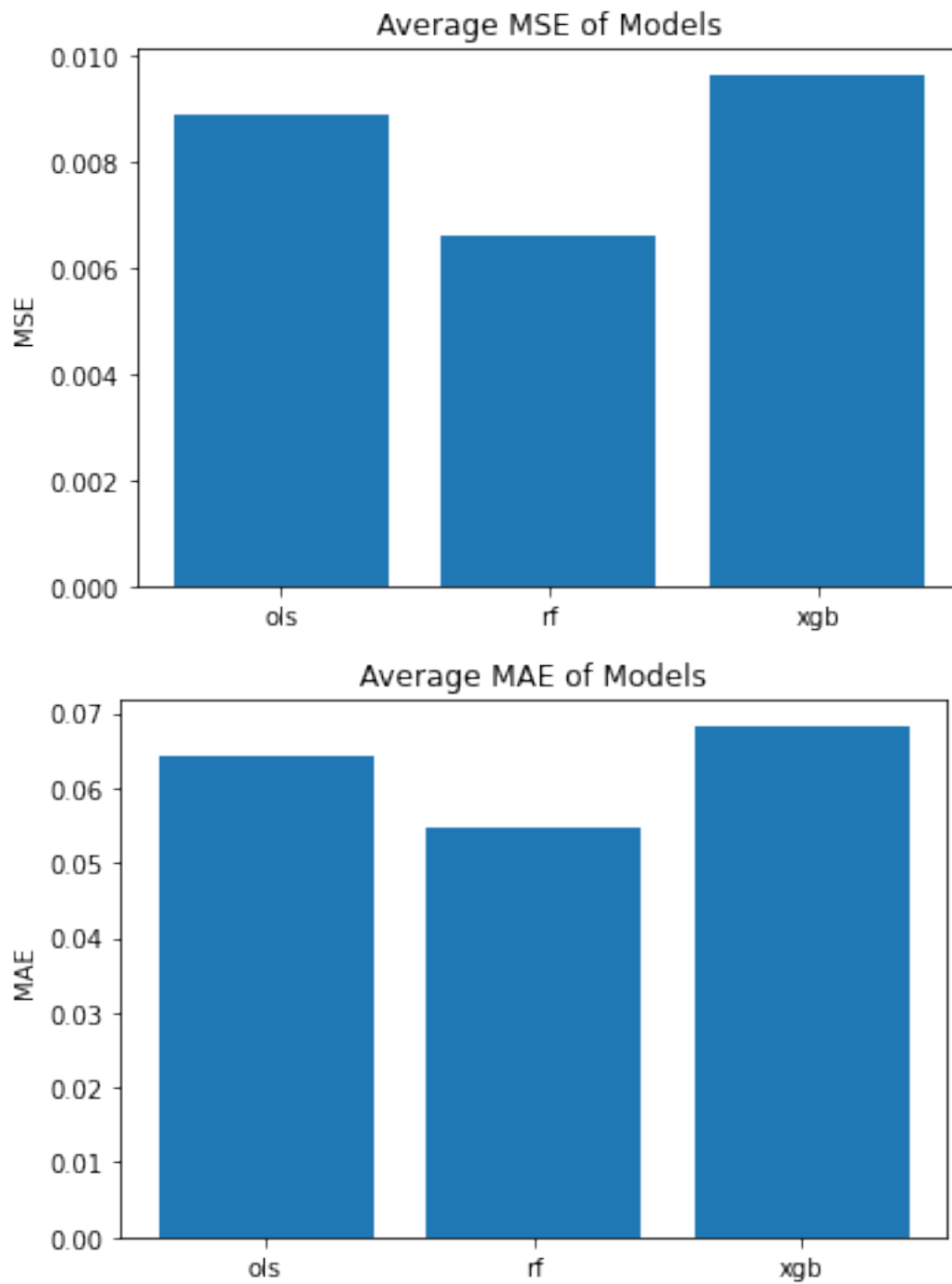
However, the error in Cardano may affect the returns produced in the backtesting, as we find that the OLS model's predicted returns produces much better results in our backtesting range than both the random forest and the XGBoost models. Thus, we will use the OLS model's predicted returns to create our trading signals.

1.6 Trading Algorithm

After collecting all the necessary features, future returns of each coin are predicted on a daily basis. Long signal in this strategy is defined as a more than 1% increase in predicted future returns. Short signal in this strategy is defined as a less than 0% decrease in predicted future returns or the holding days for that coin reaches 18 days. A coin can be bought when our current position of that coin is equal to 0. A coin can be sold only when our current position of that coin is more than 0.

1.7 Portfolio Evaluation

There is a wealth of performance statistics which most professional fund managers would use to analyze the performance of the algorithm. These metrics range from the algorithm's annual and monthly returns, return quantiles, rolling beta and sharpe ratios to the turnover of the portfolio. The most critical metrics are discussed as follows. ### Sharpe ratio The risk-adjusted return is an essential metric of any strategy. Risk-adjusted returns allow us to judge returns streams that have different individual volatilities by providing an avenue for meaningful comparison. There are different measures of risk-adjusted returns but one of the most popular is the Sharpe ratio. Sharpe ratios above 1.0 are generally considered "good", as this would suggest that the portfolio is offering excess returns relative to its volatility. Having said that, investors will often compare the Sharpe ratio of a portfolio relative to its peers. Therefore, a portfolio with a Sharpe ratio of 1.0 might be considered inadequate if the competitors in its peer group have an average Sharpe ratio above 1.0.

The Sharpe Ratio of the backtest is 1.296.
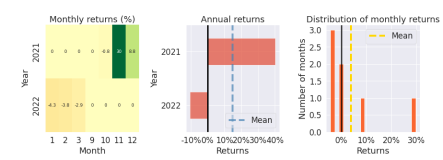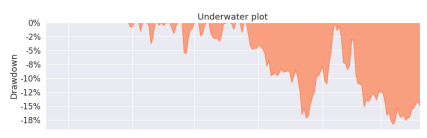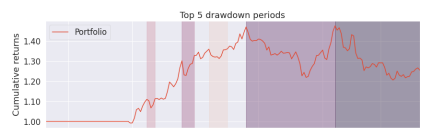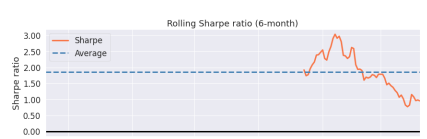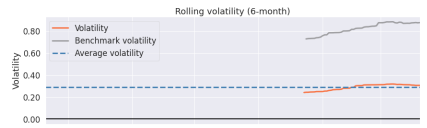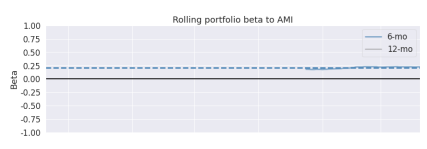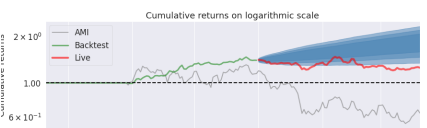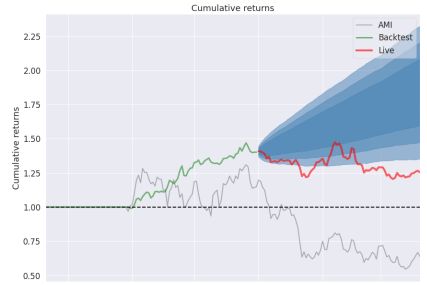
1.7.1 Market beta

The market beta of an algorithm is the exposure of that stategy to the broader market. For instance, a market beta of 1 would mean that you're buying the the market, while a beta of -1 means that you are shorting the market. Any beta within this range signifies reduced market influence, while any beta outside this range signifies increased market influence. A coin that swings more than the market over time has a beta above 1.0. If a coin moves less than the market, the stock's beta is less than 1.0. High-beta stocks are supposed to be riskier but provide higher return potential; low-beta coins pose less risk but also lower returns.

The market beta of the backtest is 0.014.

1.7.2 Drawdown

Drawdowns are important for measuring the historical risk of different investments, comparing fund performance, or monitoring personal trading performance.A drawdown refers to how much an investment or trading account is down from the peak before it recovers back to the peak. Drawdowns are typically quoted as a percentage, but dollar terms may also be used if applicable for a specific trader.They are a measure of downside volatility. The time it takes to recover a drawdown should also be considered when assessing drawdowns. A drawdown and loss aren't necessarily the same thing. Most traders view a drawdown as a peak-to-trough metric, while losses typically refer to the purchase price relative to the current or exit price. This is another measure of the financial risk of an algorithm. If the net drawdown of a strategy is very significant, this generally means that the volatility of the algorithm is more significant. Good strategies try to limit drawdowns. A good benchmark is to have a maximum drawdown of less than 20%.

The maxmimum drawdown of the backtest is: -0.183.

Cumulative returns

Cumulative returns volatility matched to benchmark

Cumulative returns on logarithmic scale

Returns

Rolling portfolio beta to AMI

Rolling volatility (6-month)

Rolling Sharpe ratio (6-month)

Top 5 drawdown periods

Underwater plot

Monthly returns (%)

Annual returns

Distribution of monthly returns

Return quantiles

10

1.8 Return Quantiles

These box and whisker plots provide an overview of the return quantiles broken down by the return timeframe (daily / weekly / monthly) across the entire backtest time period.

Return quantiles
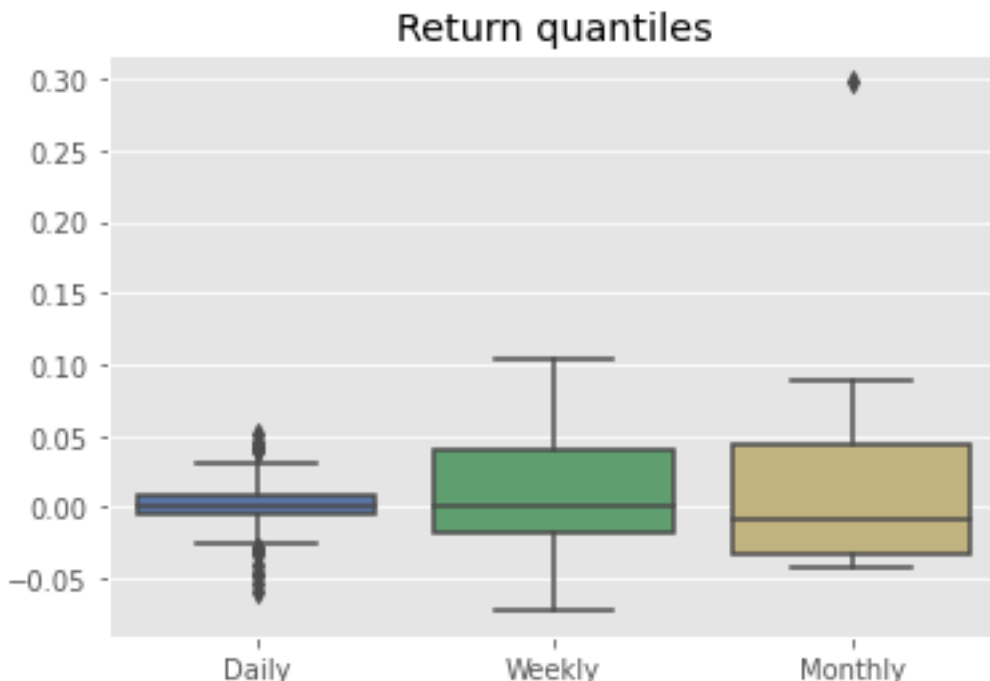
The center line in the middle of each box shows the median return, and the box shows the first quartile (25th percentile) as well as the 3rd quartile (75th percentile). While a high median return is always helpful, it is also important to understand the returns distribution. A tight box means that the bulk of the returns (25th - 75th percentile) fall within a tight bound - i.e. the returns are consistent and not volatile. A larger box means that the returns are more spread out. It is important, however, to take note of the scale to the left to put the quartiles in perspective. In addition, returns over longer periods of time will have a wider distribution as increasing the length of time increases the variability in returns.

The 'whiskers' at the end indicate the returns which fall outside the 25th and 75th percentile. A tight box with long whiskers indicate that there may be outliers in the returns - which may not be ideal if the outliers are negative. This may indicate that your strategy may be susceptible to certain market conditions / time periods.

Below, we have several rolling plots which show how an estimate changes throughout backtest period. In the case of the rolling beta and the rolling Sharpe ratio, the rolling estimate gives us more information than single point estimate for the entire period. A rolling estimate allows the user to see if the risk-adjusted return of the algorithm (Sharpe ratio) is consistent over time or if it fluctuates significantly. A volatile Sharpe ratio may indicate that the strategy may be riskier at certain time points or that it does not perform as well at these time points. Likewise, a volatile rolling beta indicates that it is exposed to the market during certain time points - if the strategy is meant to be market neutral, this could be a red flag. ### Rolling Beta Plot The plot below shows the rolling beta of the strategy against benchmark returns over the entire period of the backtest. In this instance, the benchmark return of the AMI was used. Thus, the lower the rolling portfolio beta to the AMI, the more market neutral an

Rolling portfolio beta to AMI
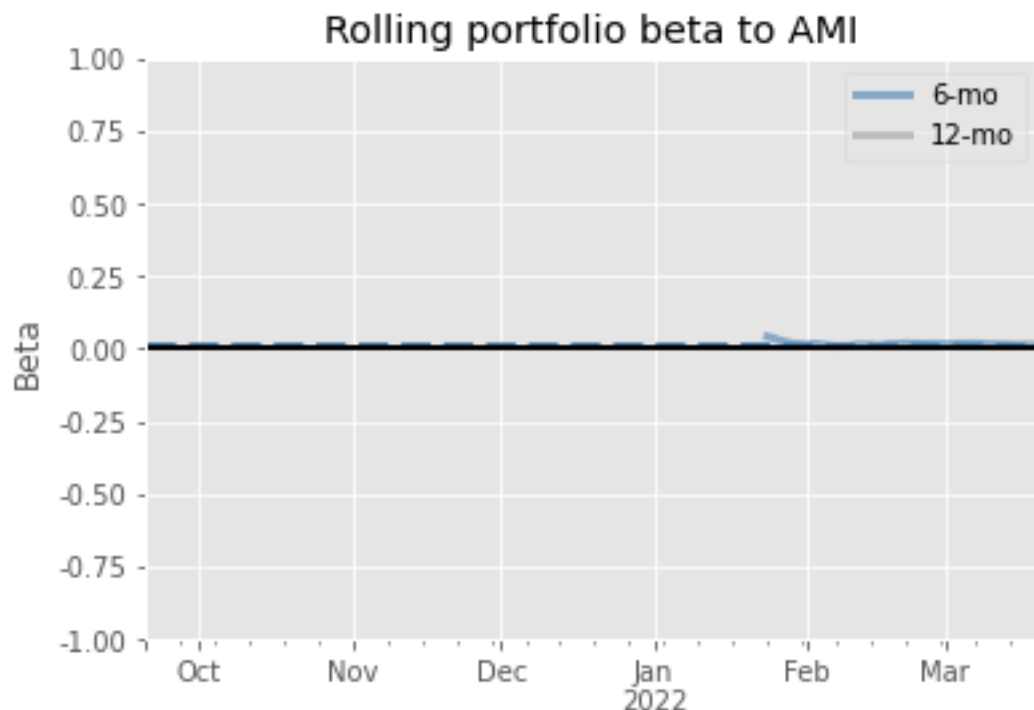
algorithm is.

Rolling Sharpe Ratio Plot The plot below shows the rolling Sharpe ratio over the period of the backtest. This allows you to understand the performance of the algorithm at different time

Rolling Sharpe ratio (6-month)
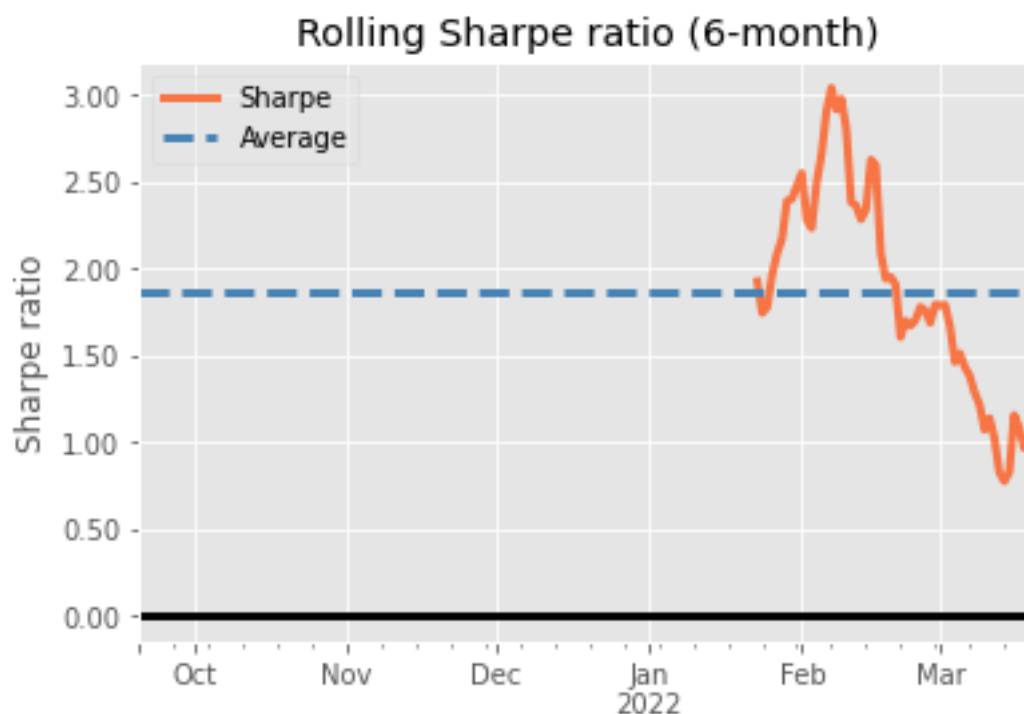
points.   ###
Drawdown & Underwater Plot In this plot, we visualize the drawdown chart described above. This chart provides an overview of the worst drawdown periods in the backtest. These periods show the time windows in the backtest in which the top 10 drawdowns occurred.

Rolling Sharpe ratio (6-month)


Underwater plot
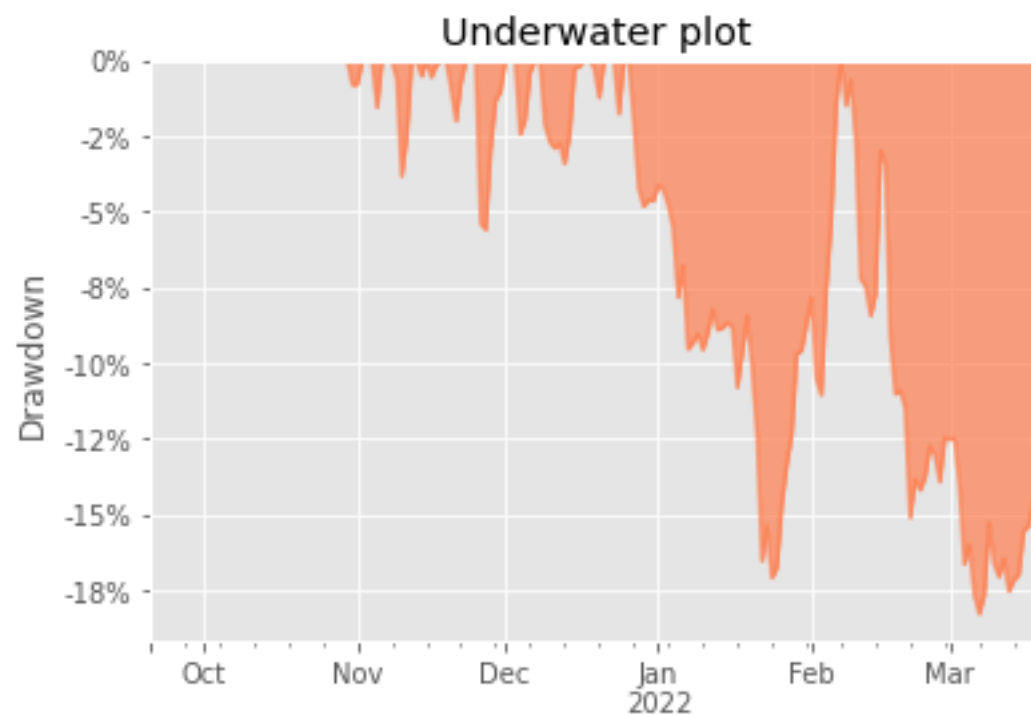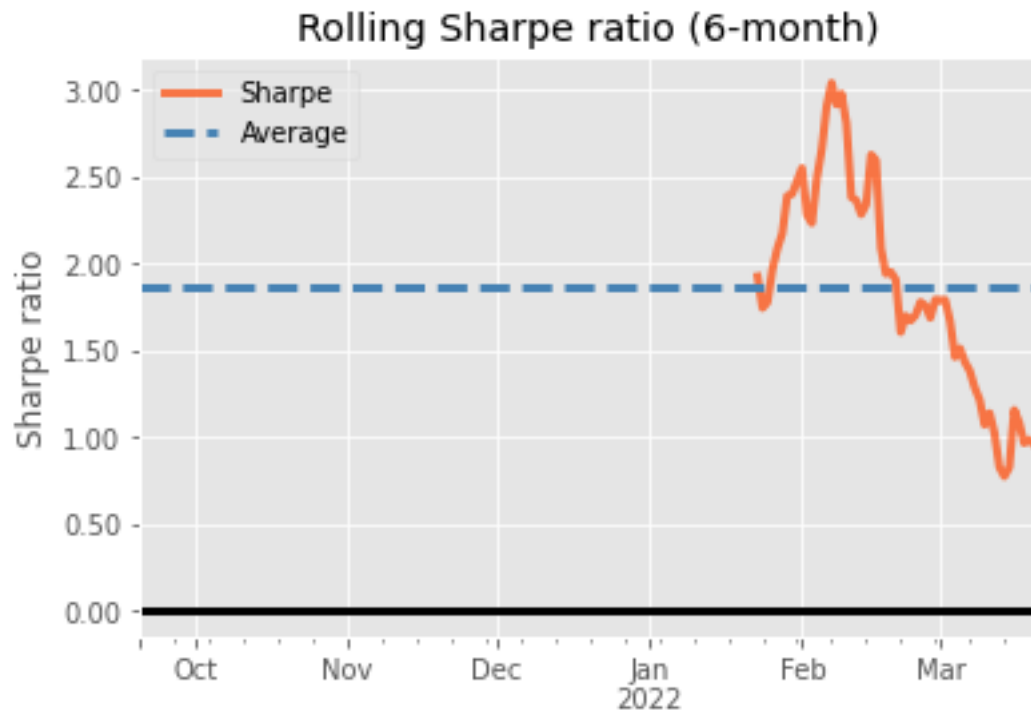
Both plots coupled together, allows for a quick check into the time periods during which the algorithm struggles. Generally speaking, the less volatile an algorithm is, the more minimal the drawdowns.

1.8.1 Top Positions

The tables below list the top 10 long and short positions of all time. The goal of each algorithm is to minimize the proportion of the portfolio invested in each security at any time point. This prevents the movement of any individual security from having a significant impact on the portfolio as a whole. The bigger the exposure a strategy has to any security, the greater the risk.

Generally, the biggest failure point for many strategies is high portfolio concentration in a few securities. While this may produce significant positive returns over a given time period, the converse can easily occur. Huge swings in a small number of equities would result in significant drawdowns. Good strategies tend to be those in which no security comprises more than 10% of the

Portfolio allocation over time, only top 10 holdings
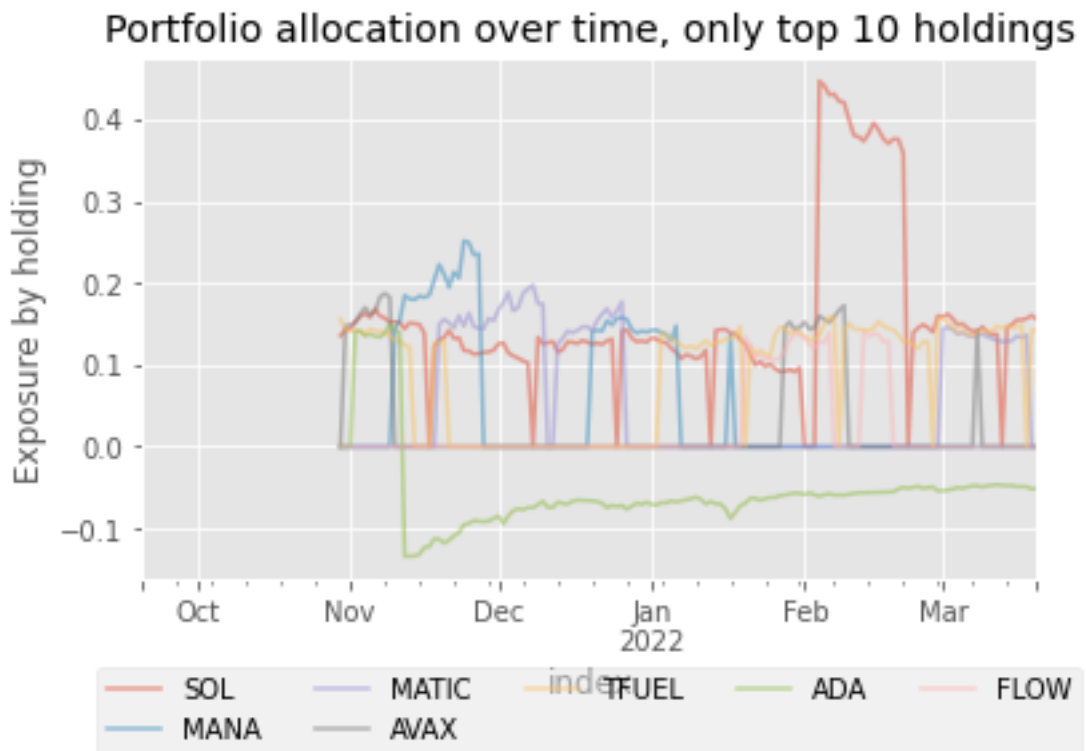
portfolio.

Top 10 positions of all time are: - SOL 44.67% - MANA 25.16% - MATIC 19.75% - AVAX 18.71% - TFUEL 15.85% - ADA 14.95% - FLOW 14.01% ### Holdings Per Day The holdings per day allows us to gain an insight into whether the total portfolio holdings fluctuate from day to day. This plot provides a good sanity check as to whether the algorithm is performing as it should, or if there were any bugs which should be fixed. For instance, we can use to holdings plot to check if the trading behavior is expected, i.e. if there are extended periods in which the number of holdings is exceptionally low or if that the algorithm is not trad-

Total holdings
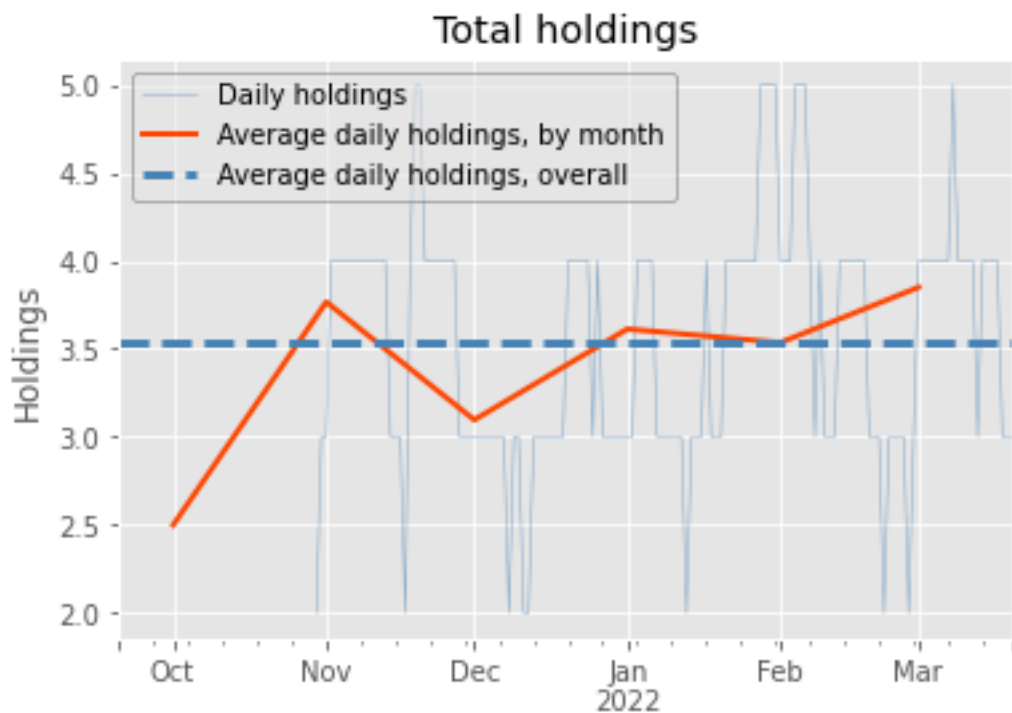
ing. ## Daily Turnover This plot reflects how many coins are traded as a fraction of total coins. The higher the daily turnover, the higher the transaction costs associated with the algorithm. However, this also means that the returns and risk metrics are better able to capture the underlying performance of the algorithm as the higher quantity of trades provides more samples (of returns, risk, etc.) to draw from. This would in turn give a better estimation on Out of Sample periods as well. Likewise, the Daily Turnover Histogram gives you an overview of the distribution of the turnover of your portfolio. This shows you both the average daily turnover of your portfolio and any outlier trading days.

Daily turnover

Distribution of daily turnover rates
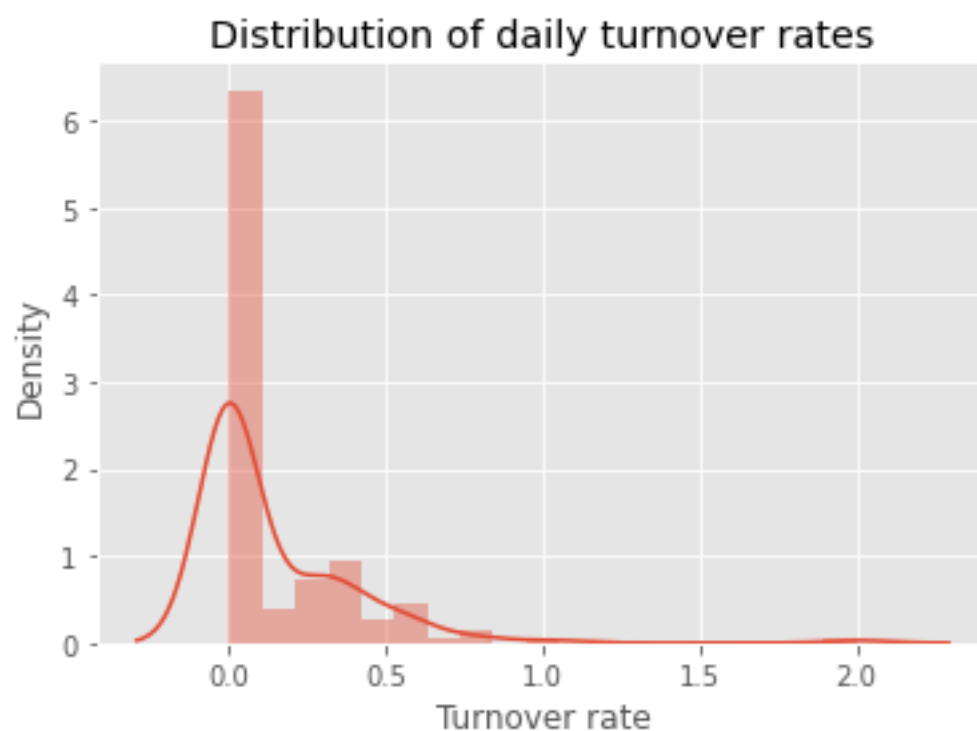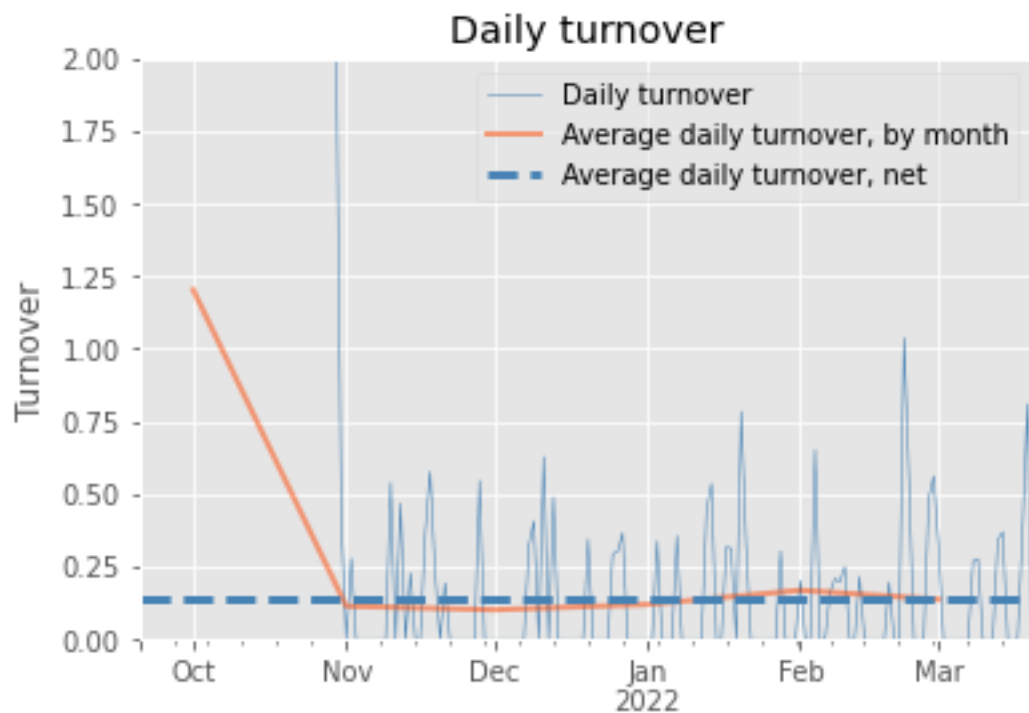
1.9 Conclusion

NFTs is a relatively new concept in the finance world. We utilized information regarding NFTs to inform our decisions in trading. Both quantitive and alternative data have been used to render features for or Machine Learning models. The Machine learning models predicted the next day's

future returns, which have been used to create long and short signals. The period which the algorithmic strategy was backtested on, is a period which the market is going down. This can be observed by observing the cumulative returns of buying and holding the AMI index. This strategy was used as a benchmark to compare the algorithmic strategy against. For that in-sample and out of sample period, our algorithm yielded us significant returns, while the AMI index was predominantly depreciating in value.