

— DBA4761

# Bitcoin Group Presentation



Completed By

Choo Jun Heng, Alvin: A0183030W

Dhameem Ansar S/O Bathusha: A0184366W

Lin Qiwei: A0183470E

Mohamad Rafi B Kamsani Chia: A0182838R

Nai Jing Wen, Bernessa: A0185739L

# 01

# Introduction



# What is Bitcoin?



- Decentralized digital currency that you can buy, sell and exchange directly, without an intermediary like a bank or financial institution



- Every transaction exists on a public ledger accessible to everyone, hard to reverse and difficult to fake



- Distributed digital record/currency build on blockchain

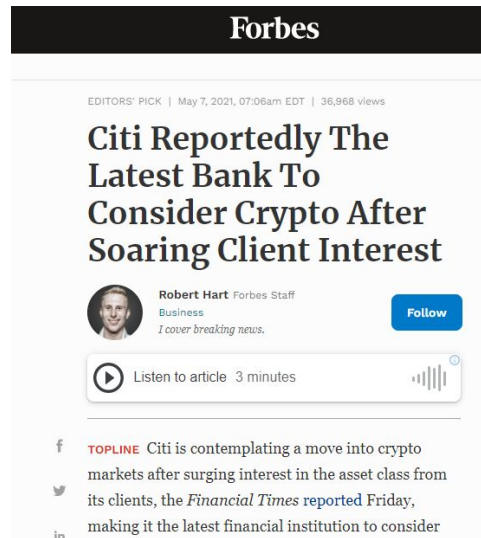


- Fixed Supply & Divisible

**Reduces risk of fraudulent transactions as unique codes are used to recognise user's wallets and transactions must conform to right encryption patterns**

# Relevance of Bitcoin Today

- **Expected to be more relevant in the future**, especially as more large, institutional investors begin treating it as a sort of digital gold to hedge against market volatility and inflation
- Citi is contemplating a move into crypto markets **after surging interest in the asset class** from its clients
- **Other major financial institutions have also moved into crypto markets this year**—including JPMorgan, BNY Mellon, Morgan Stanley and BlackRock







**Especially in the context of high volatility of bitcoin prices, will be advantageous to be able to predict future returns/trends in bitcoin prices so that investors can best maximise their returns**

## Objective of Project – Predictive Model to help us decide to buy, hold or sell

- To determine the best predictive model - which is the model that is able to **predict with maximum possible accuracy** on whether the future returns from **bitcoin is going to be positive** (i.e. bitcoin price increases the next day) **or negative** (i.e. bitcoin price decreases the next day) using **various predictive modelling/ML techniques**

By determining the best predictive model, we will then be able to do bitcoin momentum feature-based forecasting and formulate bitcoin trading strategies

# Data Sources

Data Source	Data Extracted
Bitstamp 	USD/Bitcoin exchange data (data on bitcoin prices)
Yahoo 	Data on VIX values, alternative asset class data - S&P500, DXY, USD/Yuan
NASDAQ  Nasdaq	Data on Gold Prices
CryptoCompare 	Sentiments based on daily news articles from CryptoCompare - website that tracks cryptocurrency markets with live prices, charts, free portfolio, news
“101 Formulaic Alphas” Research Paper by Zura Kakushadze (CEO Quantigic Solutions LLC)	Contains code on real-life quantitative trading alphas

# 02

## Feature Engineering



# Target Variable



**Future Direction of  
Bitcoin Returns**

- **Aim:** Predict whether future returns for Bitcoin are positive or negative, given all the information we have today.
- Informs our business decision today whether to buy Bitcoin today, hold it, or to sell it to reap maximum returns.



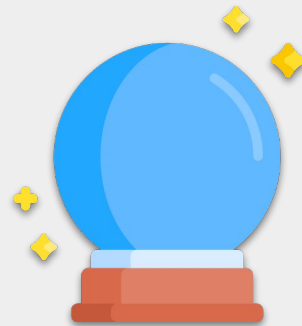
# Features - Market Volatility



## Historical Volatility

**Measures:** Rolling Daily Return Volatility

Captures effect of past volatility/risk perception on Bitcoin returns



## Future Expectation of Volatility

**Measures:** CBOE Volatility Index (VIX)

Captures effect of fear/expectations on Bitcoin returns

# Features - Relative Measure of Bitcoin's Returns



**Measures Comparing Bitcoin  
Returns to Other Assets**

## **Alternative Asset Class Data**

Captures co-movement of Bitcoin returns with alternative assets that fulfil similar purposes

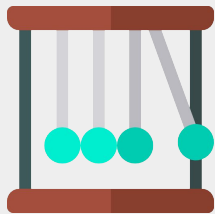
## **Alphas (Techfactors)**

Captures the effect of its market performance on its return

# Features - Relative Measure of Bitcoin's Returns



**Measures  
Comparing  
Bitcoin Returns  
to its Historical  
Returns**



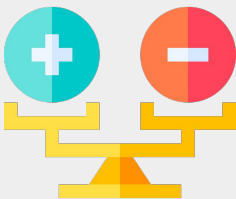
**Momentum of Price Movement/ Bitcoin  
Price Change**

Captures the strength, direction & rate of change of Bitcoin return trends



**Bitcoin Drawdown**

Captures the relative position of Bitcoin prices compared to its peak



**Number of Consecutive Positive/Negative Days**

Captures the effect consecutive runs of positive/negative days indicative of a bullish or bearish runs for Bitcoin prices

# Features - Relative Measure of Bitcoin's Returns



**Measures  
Comparing  
Bitcoin Returns  
to its Historical  
Returns**

## Technical Analysis Indicators



Captures signals investors use to either buy/sell/hold an asset due to an upcoming predicted bullish/bearish run

### Measures:

- Exponential Moving Average
- Golden Cross
- Moving Average Convergence Divergence (MACD)
- Stochastic Movement Index (SMI)

# Features - Time Series Features

## Extent of Shifts

**Measures:** Max level, variance & Kulback-Leibler divergence shifts

## Reversion to Mean

**Measures:** Crossing points

## Predictability of Time Series

**Measures:** Entropy

## Variance of Time Series

**Measures:** Stability & Lumpiness



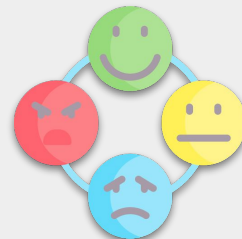
# Features - Sentiment Analysis

Cryptocurrencies

**Bitcoin Slips Below \$60,000 as ETF-Related Bliss Evaporates**

**Islamic Organization in Indonesia Issues Fatwa Against Cryptocurrency**

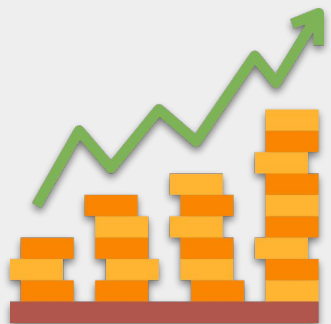
**Elon Musk backtracks, says Tesla won't accept bitcoin**



## Sentiment Analysis

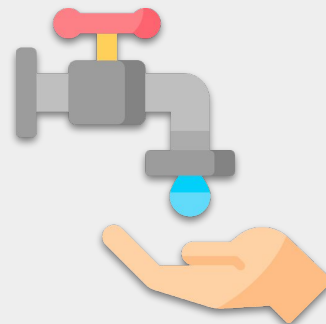
Captures bearish/bullish instincts & speculation from media releases

# Features - Other Relevant Features



## Daily Miner Revenue

Captures the effect of total mining power & the reward per mined Bitcoin



## Stock-to-Flow Ratio

Captures the effect of scarcity of Bitcoin on returns

# 03

## Predictive Modelling





# Overview

---

1

**Data Split**

2

**Apply Model**

- Background
- Hyperparameter Search
- Performance

3

**Evaluation**

4

**Further Analysis (Cost)**

# Data Split



## Train-Test Split

80% of the data is split into training and 20% into testing

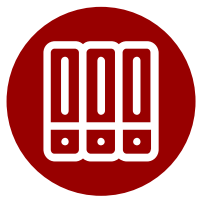


1 Jan 2018

1 Oct 2020

21 Jun 2021

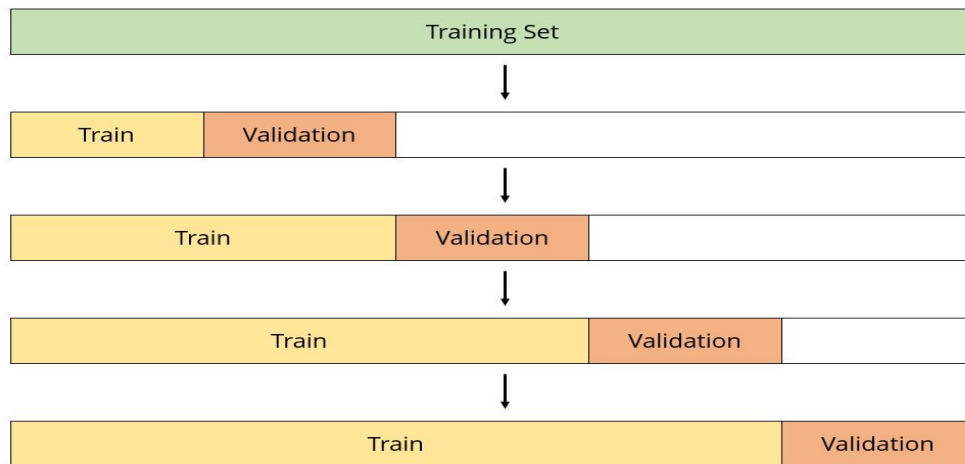
# Data Split



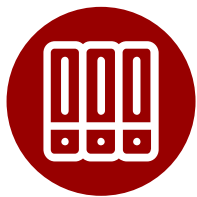
## Cross Validation Split

3 months rolling basis

10 Splits



# Data Split



## Cross Validation Split

3 months rolling basis

10 Splits



# Model Overview

1

**XGBoost**

2

**Catboost**

3

**Logistics  
Regression**

# Model Overview - XGBoost & Catboost

## Background

### XGBoost

Part of gradient boosting, which produces prediction models in the form of an ensemble of weak prediction models.

XGBoost further improve by **using more regularised model formalisation to control over-fitting**, a common problem associated with decision tree models.

### CatBoost

Developed by Yandex.

A variant of the gradient boosting method, and is claimed to be one of the models that is revolutionising the machine learning game.

# Model Overview - XGBoost & Catboost

## Background



### Advantages of Catboost over XGBoost

- CatBoost implements symmetric trees.
- Trains only *log* number of data points in a procedure known as ordered boosting.
- Has a very good vector representation of categorical data unlike XGBoost which cannot handle categorical features by itself
- CatBoost is comparatively 8x faster than XGBoost while predicting

# Model - XGBoost & Catboost

## Hyperparameter Search

## Loss Function

- Mean logarithmic loss used as estimator.
- Allows us to evaluate the absolute probabilistic difference of each prediction:
  - The more certain the model is of an observation that is true, the lower the error.
  - Conversely, the more certain the model is of an observation that is untrue, the heavier the penalty.



# Model - XGBoost & Catboost

## Hyperparameter Search

### Cross-Validation

Determine the **optimal number of trees to build the model**.

This is the optimal number of weak trees to build to reduce variance of the model and prevent overfitting

### XGBoost

Trees	Metric	Estimator	Mean	Standard Error
400	nm_log_loss	binary	0.708	0.00904
800	nm_log_loss	binary	0.723	0.0153
1200	nm_log_loss	binary	0.740	0.0218
1600	nm_log_loss	binary	0.756	0.0276
2000	nm_log_loss	binary	0.772	0.0355

### CatBoost

Trees	Metric	Estimator	Mean	Standard Error
400	nm_log_loss	binary	0.698	0.00366
800	nm_log_loss	binary	0.711	0.0105
1200	nm_log_loss	binary	0.735	0.0201
1600	nm_log_loss	binary	0.764	0.0321
2000	nm_log_loss	binary	0.791	0.0415

# Model - XGBoost & Catboost

## Performance

### XGBoost

	Metric	Estimator	Estimate
Train	nm_log_loss	binary	0.654
Test	nm_log_loss	binary	<b>0.694</b>

### CatBoost

	Metric	Estimator	Estimate
Train	nm_log_loss	binary	0.664
Test	nm_log_loss	binary	<b>0.682</b>

# Model Overview - Logistics Regression

## Background

### What

A statistical model that uses a logit function to model a binary dependent variable. Essentially, it is like linear regression of a classification model.

### How

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

# Model - Logistics Regression

## Hyperparameter Search

## Loss Function

- Mean logarithmic loss used as estimator.
- Allows us to evaluate the absolute probabilistic difference of each prediction:
  - The more certain the model is of an observation that is true, the lower the error.
  - Conversely, the more certain the model is of an observation that is untrue, the heavier the penalty.

# Model - Logistics Regression

## Hyperparameter Search

### Cross-Validation

Determine the penalty to be imposed on the model for having too many variables. This shrinks the coefficients of the less contributive variables towards 0 to reduce variance of the model and prevent overfit.

Penalty	Metric	Estimator	Mean	Standard Error
0.0000000001	nm_log_loss	binary	5.89	1.11
0.00000001	nm_log_loss	binary	5.89	1.11
0.000001	nm_log_loss	binary	5.89	1.11
0.0001	nm_log_loss	binary	5.89	1.11
0.01	nm_log_loss	binary	1.46	0.227

# Model - Logistics Regression

## Performance

	Metric	Estimator	Estimate
Train	nm_log_loss	binary	0.588
Test	nm_log_loss	binary	1.425

# Evaluation

## Summary

Model	Data	Metric	Estimate
Logistics Regression	Train	nm_log_loss	<b>0.588</b>
XGBoost	Train	nm_log_loss	<b>0.654</b>
Catboost	Train	nm_log_loss	<b>0.664</b>
Catboost	Test	nm_log_loss	<b>0.682</b>
XGBoost	Test	nm_log_loss	<b>0.693</b>
Logistics Regression	Test	nm_log_loss	<b>1.425</b>

**Best Performance on Testing**  
Catboost

# Evaluation

## Model Performance

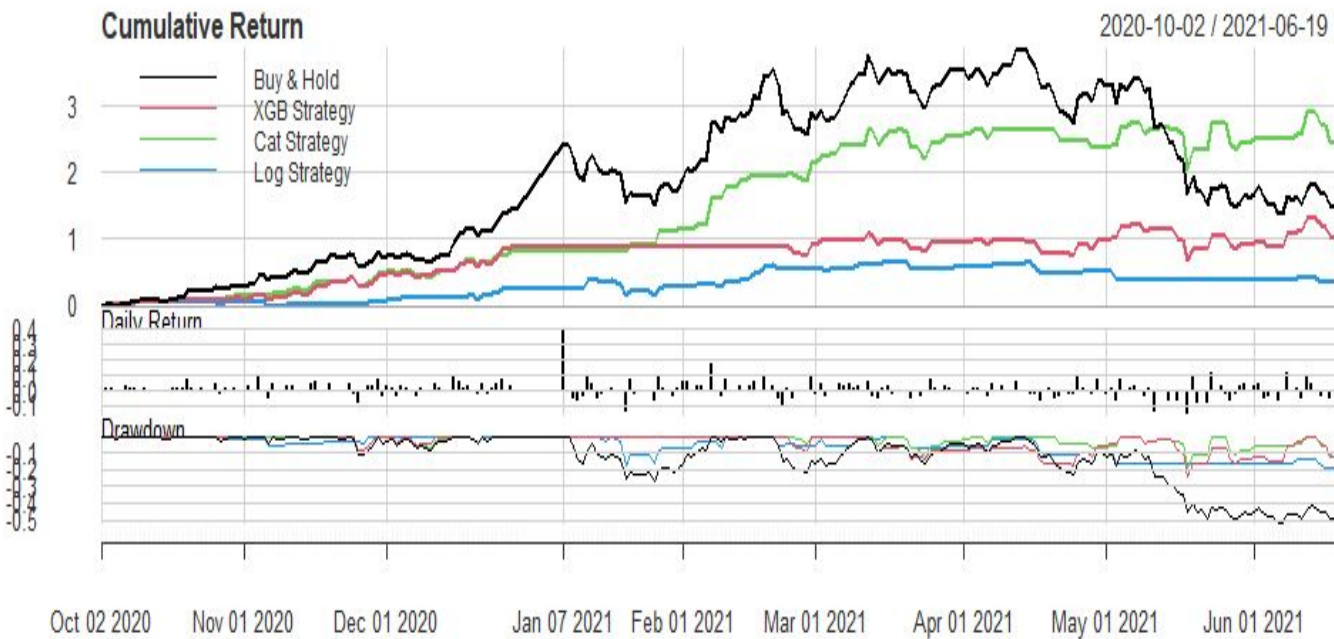
	Buy & Hold	XGBoost	Catboost	Logistics Regression
Annualized Return	1.492	1.038	2.451	0.348
Annualized Std Dev	0.799	0.460	0.464	0.339
Annualized Sharpe (RF=0%)	1.869	2.256	5.285	1.025

**Catboost**  
**Highest Annualized Return**  
**Only strategy that perform better than 'Buy & Hold'**



# Evaluation

## Strategy Performance



# Further Analysis

## Cost

- Transaction cost should also be considered when determining the best strategy since too many buying and selling would incur high transaction cost and lead to overall lower return.
- In Singapore, the trading fees can range from 0.20% to 2.5% depending on trading platforms, and this is not inclusive of other miscellaneous fees.

	Number of Crypto Listed	Trading Fees
Huobi Global	219	0.20%
Crypto.com	80+	0.20%
Gemini	40+	0.25% to 1.49%
Binance.sg	8	0.60%
Coinbase	37	0.5% to 2.5%
Independent Reserve	10	0.5%

# Further Analysis

## Cost

Transaction Cost = 0.003 (0.3%)

	Buy & Hold	XGBoost	Catboost	Logistics Regression
Annualized Return	1.492	0.559	1.535	0.0789
Annualized Std Dev	0.799	0.461	0.463	0.339
Annualized Sharpe (RF=0%)	1.869	1.212	3.317	0.233

### Catboost

Remains the **best performance strategy**  
**No longer as effective** with cost factored in  
**Only marginally better performance** than 'Buy & Hold'

# 04

## Feature Selection



# Overview



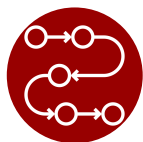
## What is feature selection?

Process of reducing the number of input variables when developing a predictive model



## Why is feature selection necessary?

- Mitigate overfitting
- Reduce the variance of the model
- Improve the model's efficiency



## Process of Applying Feature Selection

- Prepare the dataframe
- Split the data into train and test
- Apply the feature selection technique
- Evaluate the model

## Dimensionality Reduction Techniques

### Feature Transformation

Principal Component Analysis (PCA)

### Feature Selection

Information gain

Maximum Relevancy Minimum Redundancy (MRMR)

Boruta

# Principal Component Analysis

## What

Projection method whereby data is reduced into its principal components that maximise the amount of variance explained

## Result

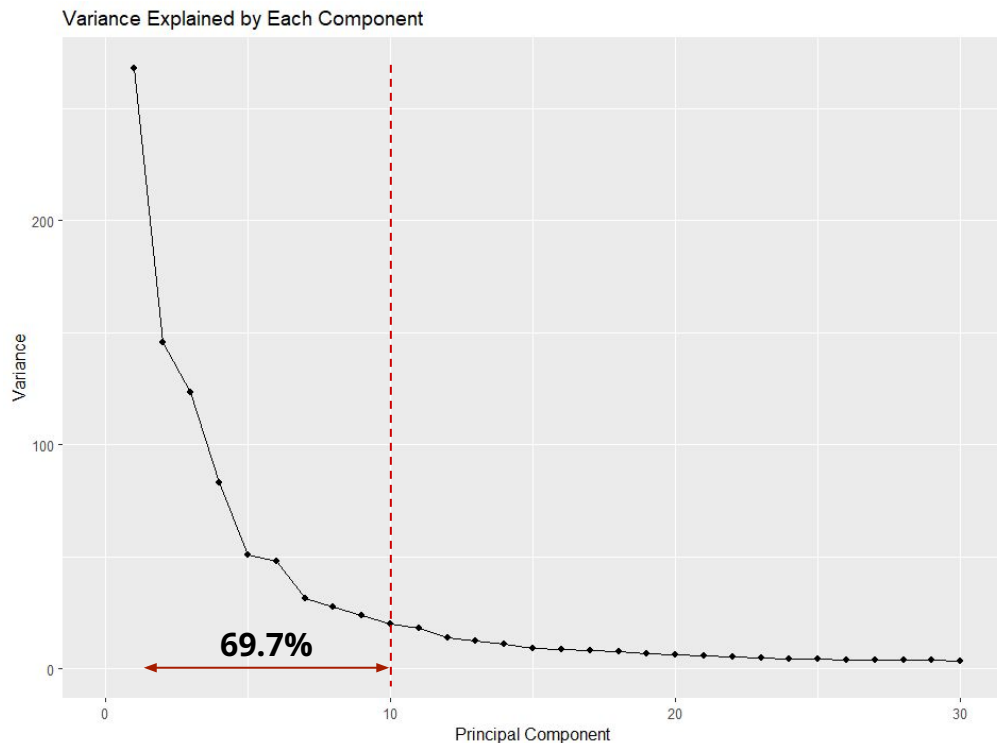
Resulting features are not correlated with each other

## Limitation

Ignores categorical variables



# Principal Component Analysis



- The impact on variance reduces significantly after the 10th component
- 69.7% of the model explained

**Performance on Testing**  
Error estimate: 0.697

# Information Gain

**Goal** Maximise information gain

**How** Information Entropy: How much variance a dataset has

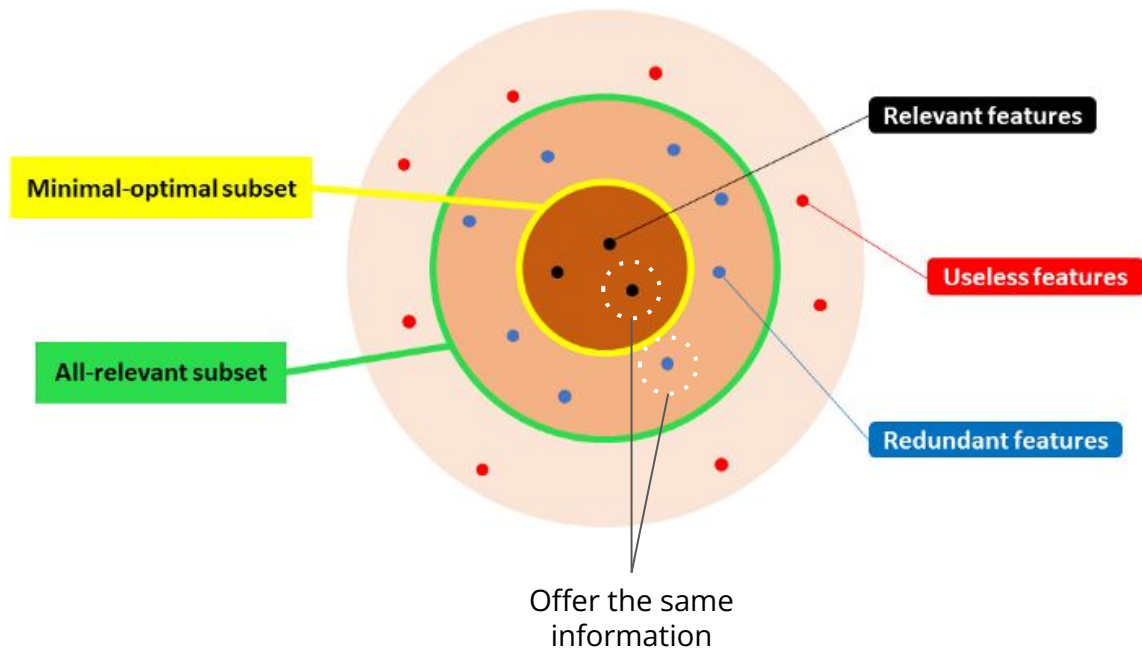
- The data is subsetting
- Each subset's information entropy is calculated and summed
- The higher the amount of information entropy removed, the greater the information gain

**Performance on Testing**

Error estimate: 0.682



# Types of Subsetting



## Minimal-Optimal

Identify a small set of features that have maximum possible predictive power in combination

## All-Relevant

Select all features that individually have any predictive power

# Maximum Relevancy Minimum Redundancy (MRMR)

**Method** Minimal-Optimal

**How** After feature selection, redundant features of data are removed

- At every iteration, the algorithm selects the feature that has **maximum relevance** to the target variable and **minimum redundancy** to the variables selected in the previous iterations

**Performance on Testing**

Error estimate: 0.694

# Boruta

**Method** All-Relevant

**How** After feature selection, useless features of data are removed

- Creation of **shadow features** - another dataframe with randomly shuffled features
- A random forest is fitted on the shadow features dataframe and the target variable
- Compare the importance of the original features against the highest feature importance of the shadow features
- Only features that do better than the shadow features are kept
- Above steps are repeated and it is stopped when either all features are either confirmed/rejected or the specified limit of runs is reached

**Limitation** Takes all features even if they give the same information

**Performance on Testing**  
Error estimate: 0.710

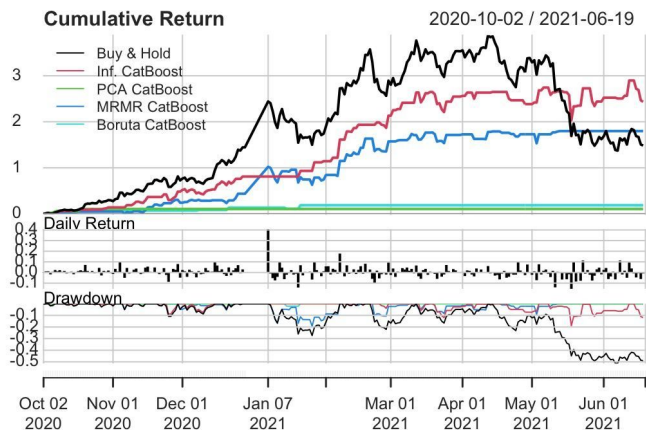
# Comparison of Results

## Error Estimates

	PCA	Info Gain	MRMR	Boruta
Train	0.681	<b>0.664</b>	0.674	0.672
Test	0.697	<b>0.682</b>	0.694	0.710

## Performance

### Strategy Performance



	PCA	Info Gain	MRMR	Boruta	Buy & Hold
Annualized Return	0.105	<b>2.451</b>	1.809	0.187	1.492
Annualized Std Dev	0.047	<b>0.464</b>	0.599	0.095	0.799
Annualized Sharpe (RF=0%)	2.22	<b>5.285</b>	3.022	1.959	1.869

Information Gain performed the best

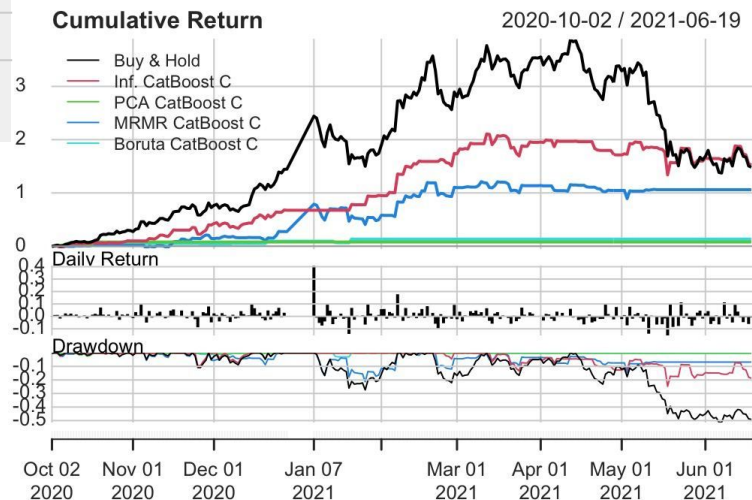
# Accounting for Cost

Transaction Cost = 0.003

	PCA	Info Gain	MRMR	Boruta	Buy & Hold
Annualized Return	0.079	1.535	1.067	0.131	1.492
Annualized Std Dev	0.048	0.463	0.600	0.093	0.799
Annualized Sharpe (RF=0%)	1.632	3.317	1.778	1.4187	1.869

Information Gain performed the best

## Strategy (with Cost) Performance



# 05

## Possible Improvements

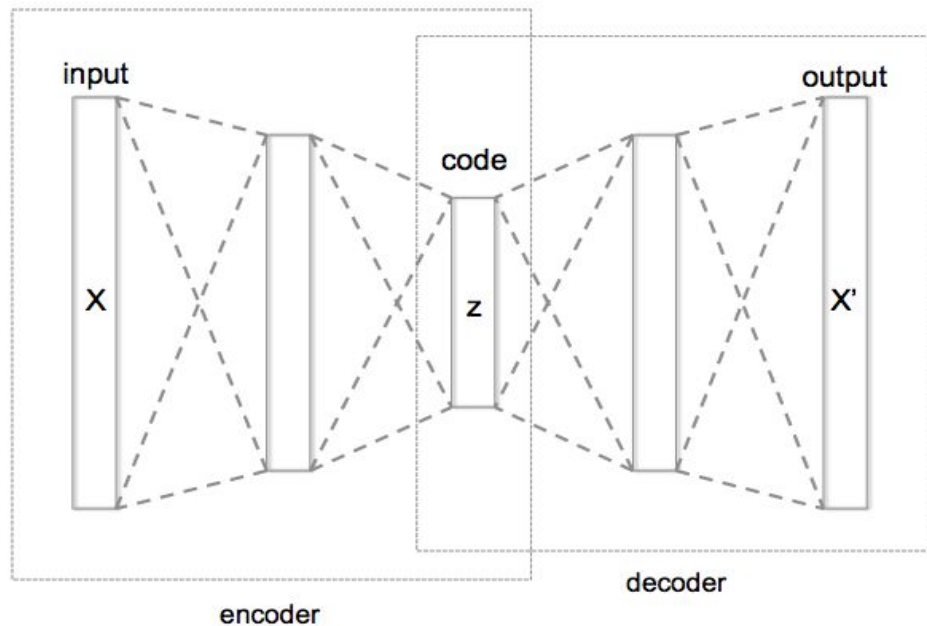


# Autoencoder

---

- An unsupervised artificial neural network
- Based on the Encoder-Decoder architecture
- Encoder encodes the high-dimensional data to a lower-dimension
- Decoder tries to reconstruct the original high-dimensional data
- Remove noise and redundancy

# Autoencoder





# Autoencoder

PCA	Autoencoder
Only linear transformation	Works for both linear and non-linearities
Deterministic, Faster	Gradient Descent, Slower
Results in orthogonal features	Only to minimize reconstruction loss

# Reinforcement Learning

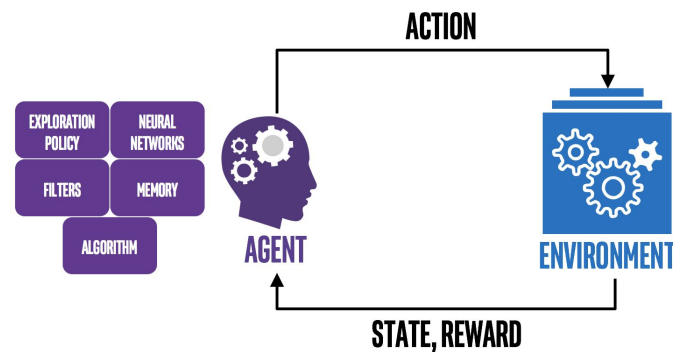
---

- Current modelling approach presents a relatively inefficient trading strategy
- Only predicting the trading signal of the next day
- Does not provide a recommendation on how much quantity to buy or sell
- Reinforcement Learning can address this

# Reinforcement Learning

- Trains an agent to interact with the environment
- Sequentially receiving states and rewards from the environment
- Taking actions to reach better rewards
- Model the crypto trading process as a RL

Problem



# Reinforcement Learning

- DRL uses a reward function to optimize future rewards
- Profit maximization as trading goal
- Reward function as the change of the portfolio value
- DRL maximizes the portfolio value over time



— Thank  
You —



# 06

## Appendix A: Feature Engineering



# Features Explaining Target Variable

- **Market volatility**
  - To account for how risk affects buying decisions
- **Relative measures of Bitcoin's current return**
  - Compared to other assets' returns
    - To proxy for different purposes of Bitcoin
  - Compared to Bitcoin's previous returns
    - To understand trends to identify optimal points to buy/sell/hold an asset
- **Time series features**
- **Sentiment analysis**
  - To account for how speculation/instinct drives erratic buying decisions
- **Other relevant features**
  - Daily Bitcoin miner revenue
  - Stock-to-Flow ratio

# Market Volatility

- **Rolling Daily Return Volatility (Past Market Volatility)**

- The standard deviation of previous bitcoin prices could indicate the degree of volatility, which may be a valid indicator for the future prices of Bitcoin.
- Here, we include the standard deviation of prices from 1 to 90 days worth of previous Bitcoin prices. (although  $n = 1$  doesn't really make sense since  $SD = 0$ )

- **VIX (Future Market Volatility)**

- VIX, also known as the CBOE volatility index, is a real-time market index representing the market's expectations for volatility over the coming 30 days. Investors use the VIX to measure the level of risk, fear, or stress in the market when making investment decisions.
- Given the volatile nature of bitcoin, which is also considered to be “digital gold”, it is good to investigate the effect of fear/expectations (as represented by VIX) on bitcoin prices
- Data consisting of all the historical VIX values is loaded from Yahoo and saved, and the VIX values for the corresponding dates in the bitcoin data are added to the bitcoin predictive model.





# Relative Measures of Bitcoin's current return (Performance compared to other stocks)

- **Alternative asset class data (Relative Measure of Performance to Other Stocks)**
  - Alternative asset class data refers to data on alternative stocks/assets that possess at least one of the potential attributes/characteristics of bitcoin. The attributes are as the following:
    - **Safe haven asset** → Alternative asset/stock: Gold
    - **Risk on investment** → Alternative asset/stock: S&P500
    - **Hedge for the dollar** → Alternative asset/stock: DXY (USD) and USDCNY(USD/Yuan) currency pair
  - Data on Gold, S&P 500 stocks, DXY and USD/Yuan currencies are extracted from NASDAQ and Yahoo. The data is then used to calculate the lag values and daily returns before being added to the dataset of the predictive model

# Relative Measures of Bitcoin's current return (Performance compared to other stocks)

---

- **Techfactors (alpha)**

- Techfactors are the alphas that are used to evaluate whether a stock is outperforming or underperforming its competitors in the market.
- Here, we calculate all 191 alphas from this research paper for Bitcoin, using the baseline stock of SP500.

# Relative Measures of Bitcoin's current return (Performance compared to bitcoin past performance)

- **Bitcoin price change**

- Bitcoin price change refers to the rate of change of bitcoin close prices over a range of days. In our model, the bitcoin price change variable is the rate of change in bitcoin prices over a period of 90 days. We use both continuous and discrete values for the rate of change of bitcoin close prices

- **Momentum of Price Movement**

- Momentum of price movement shows the rate of change in price movement over a period of time to help investors determine the strength of a trend. This can help to determine whether a particular stock/asset is moving in a particular direction with the strength of momentum.
- Using the bitcoin closing price change rates and 1-day lag price values, we also can calculate the momentum of price movement, which in this case is the derivative/the rate at which the rate of change of bitcoin prices is changing. We use both continuous and discrete values for the momentum of price



# Relative Measures of Bitcoin's current return (Performance compared to bitcoin past performance)

- **Bitcoin Drawdown**

- Drawdown is the difference between the current price of Bitcoin and when it is at its peak.

- **Number of positive days**

- The number of positive days in the past n days could indicate whether Bitcoin prices the next day could be positive or not.
- Firstly, we include 2 indicator variables to see whether the change in closing prices are positive or negative.
- Then, we include 90 more variables for the number of positive closing prices for the past 1 to 90 days.

- **Number of consecutive positive/negative days**

- The consecutive runs of positive or negative days could indicate whether there is a bullish or bearish run for Bitcoin prices due to speculation etc. This could help predict future Bitcoin prices too.
- We include two new variables which indicate a positive number if there are n number of days with consecutive positive/negative changes in Bitcoin prices.



# Relative Measures of Bitcoin's current return (Performance compared to bitcoin past performance)

- **Technical Indicators (EMA, Golden Cross, MACD, SMI)**
  - **20, 50 & 200 Day Exponential Moving Average (EMA):**
    - An indicator for entry/exit when yesterday's and today's candle are above/below EMA.
    - Equals 1 when yesterday's/today's prices are above EMA but two days ago was below, 0 if yesterday's/today's prices are below EMA but two days ago was above.
  - **Golden Cross:**
    - A golden cross is a chart pattern in which a short-term moving average crosses above a long-term moving average.
    - Hence, when the 20 day EMA crosses above the 50 day EMA, we indicate it with a 1 and when the 50 day EMA crosses over the 20 day EMA, we indicate it with a 0.
  - **Moving Average Convergence Divergence (MACD):**
    - The MACD is calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA
    - A nine-day EMA of the MACD called the "signal line," is then plotted on top of the MACD line, which can function as a trigger for buy and sell signals.
    - Traders may buy the security when the MACD crosses above its signal line and sell—or short—the security when the MACD crosses below the signal line.
    - When the MACD line crosses over its signal line, we denote it as 1, and 0 if the MACD line dips below its signal line.

# Relative Measures of Bitcoin's current return (Performance compared to bitcoin past performance)

- **Technical Indicators (EMA, Golden Cross, MACD, SMI)**
  - **Stochastic Momentum Index (SMI):**
    - The SMI relates the closing prices to the midpoint of the high/low range.
    - The SMI has a normal range of values between +100 and -100. When the present closing price is higher than the median, or midpoint value of the high/low range, the resulting value is positive. When the current closing price is lower than that of the midpoint of the high/low range, the SMI has a negative value.
    - Traders also use the SMI as a general trend indicator, interpreting values above 40 as indicative of a bullish trend and negative values greater than -40 as showing a bearish trend.
    - Here, we denote a value 1 if closing prices surpass the EMA-20 and SMI is below -40 and a value 0 if closing prices drop below the EMA-20 and SMI is above 40.

# Time Series Features

Using the “tsfeatures” function from the “tsfeatures” package, we include time series features such as:

- **Entropy**
  - Measures the "forecastability" of a series. Low values indicate high signal-to-noise while large values indicate a difficulty to forecast
- **Stability**
  - Stability is the variance of the means
- **Lumpiness**
  - Lumpiness is the variance of the variances
- **Mean level, variance and KL shifts**
  - max\_level\_shift: Finds the largest mean shift between two consecutive windows (returns two values, size of shift and time index of shift)
  - max\_var\_shift: Finds the largest variance shift between two consecutive windows (returns two values, size of shift and time index of shift)
  - max\_kl\_shift: Finds the largest shift in the Kulback-Leibler divergence between two consecutive windows (returns two values, size of shift and time index of shift)
  - crossing\_points: Number of times a series crosses the mean line



# Sentiment Analysis

---

- Using news articles from CryptoCompare, we analyze the sentiments of words in their titles.
- Namely, we use the imported “sentiments” dataset from ‘tidytext’ which links words with a sentiment to classify whether a particular day’s news articles are largely positive or negative.
- Each positive/negative word is given a score of 1 or -1 and this is summed up over the day to give a net sentiment.
- Such sentiments could be a valid predictor for bitcoin prices since there is significant association with bearish/bullish instincts from such news sources.





# Other Relevant Features

- **Stock to flow (S2F) multiple (how scarcity drives prices)**
  - Ratio which measures the scarcity of bitcoin by taking **ratio of spot price (i.e. current market price) against the fair value of bitcoins (i.e. actual worth)**
  - If close to 1: current market price matches actual worth
  - If greater than 1: current market price overestimates actual worth
- **Bitcoin miner revenue (how much miners earn per bitcoin)**
  - Miners get a profit for each successful solution to algorithms that unlock a bitcoin i.e. those that solve the puzzles first get the rewards for a successful mining. It makes sense that bitcoin prices are correlated with how much these miners make.
  - We include 1-90 days lag of miner's revenue, the day-to-day percentage changes and 5 to 50-day moving averages of miner revenue as well as the current drawdown from the peak of bitcoin miner's revenue.

