



# G Research Crypto Forecasting Competition

Optimize winning model's performance in bear market

財金所碩二 葉力嘉  
胡祖望  
黃元裕



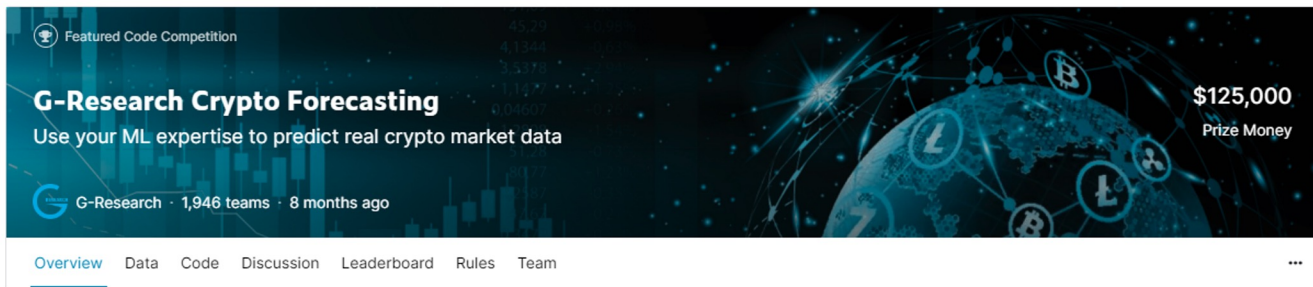
## Outline

- Background and Motive
- Data Preprocessing
- Model Architecture
- Performance Analysis
- Improvement

---

# Background

- G-Research is Europe's leading quantitative finance research firm.
- Held a crypto forecasting competition on Kaggle.
- Training Timeline: November 2, 2021 ~ February 1, 2022
- Forecasting Timeline: February 1, 2022 ~ May 4, 2022



# Background

- Training Timeline: November 2, 2021 ~ February 1, 2022
- Forecasting Timeline: February 1, 2022 ~ May 4, 2022
- But then crash happened and bear market began.

市場概況 > 比特幣

16,804.40 USD

+ 追蹤

-32,150.40 (65.67%) ↓ 過去 1 年

12月21日 下午5:44 [UTC] · 免責聲明




























1 天 | 5 天 | 1 個月 | 6 個月 | 本年迄今 | 1 年 | 5 年 | 最久



# Motive

- We wonder whether those gold medal models will continue to perform well in bear market.
- So we picked the third place models to test our ideas and tried to enhance performance comparing to the original model.

■ Prize Winners

#	Team	Members		Score	Entries	Last	Solution
1	Meme Lord Capital	 		0.0326	2	1y	
2	Nathaniel Maddux			0.0283	2	1y	
3	GABA			0.0232	2	1y	
4	cmanning			0.0229	2	1y	
5	Stamford009			0.0225	2	1y	
6	yuto			0.0203	2	1y	
7	Patrick Yam			0.0202	2	1y	
8	404NOTFOUND	    		0.0201	2	1y	
9	bturan19			0.0194	2	1y	



# Data Preprocessing



## Generate Training Data

在競賽中，除了競賽期間給的training data，2018-01-01~2021-03-12

及supplemental data，2021-03-12~2022-01-24

我們將兩者重疊的日期先處理掉，再合併兩個檔案為一個檔案，時間為

2018-01-01~2022-01-24

還另外去Binance抓了從2022-01-24~2022-12-5的data準備當作testing data



## Data Given

Asset ID: 0~13 e.g. Bitcoin, ETH, Binance coin

timestamp: 時間(年月日)-> 整數 (每1分鐘一筆數據)

Open: 開盤價

Close: 收盤價

High: 一分鐘內的最高交易價格

Low: 一分鐘內的最低交易價格

Volume: 一分鐘內的交易量

Target :  $\log\_return \text{ of future 15 mins } - \beta * \text{market\_log\_return}$  預測目標 $y$  (CAPM)





## Original Feature

lag\_time = [60,300,900] (minutes)

1.  $\log\_close\_mean = \log( close / \text{mean}(close) )$  : 近期的平均漲(跌)勢
2.  $mean\_close / mean$  : 前者的所有資產平均
3.  $\log\_close / mean - mean\_close / mean$  : 兩者相減
4.  $\log\_return\_lag = \log( close / close(lag) )$  : 近期的漲(跌)勢
5.  $mean\_log\_return$  : 前者的所有資產的平均
6.  $\log\_return\_lag - mean\_log\_returns$  : 兩者相減

# Generating New Features

新feature: lag\_time = [60,300,900] (minutes)

1. 漲跌幅超過2% in lag time  $t=[i-\text{lag}:i]$

return 1 if  $r > 2\%$

return -1 if  $r < -2\%$

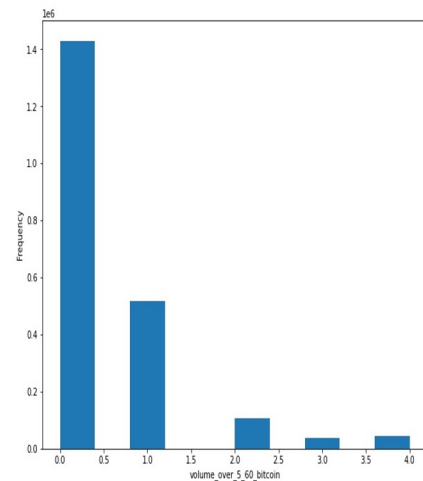
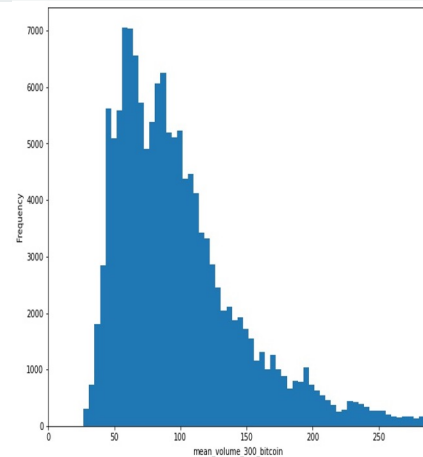
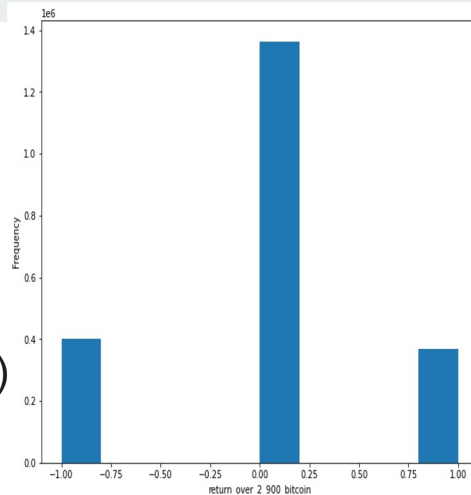
return 0 otherwise

2. mean\_volume = mean( volume(lag) ) 過去一段時間的平均交易量

3. volume\_over\_mean: 交易量超越平常的k倍  $k = 1,2,3,4$

return k if volume  $> k * \text{mean}(\text{volume}(\text{lag}))$

return 0 otherwise

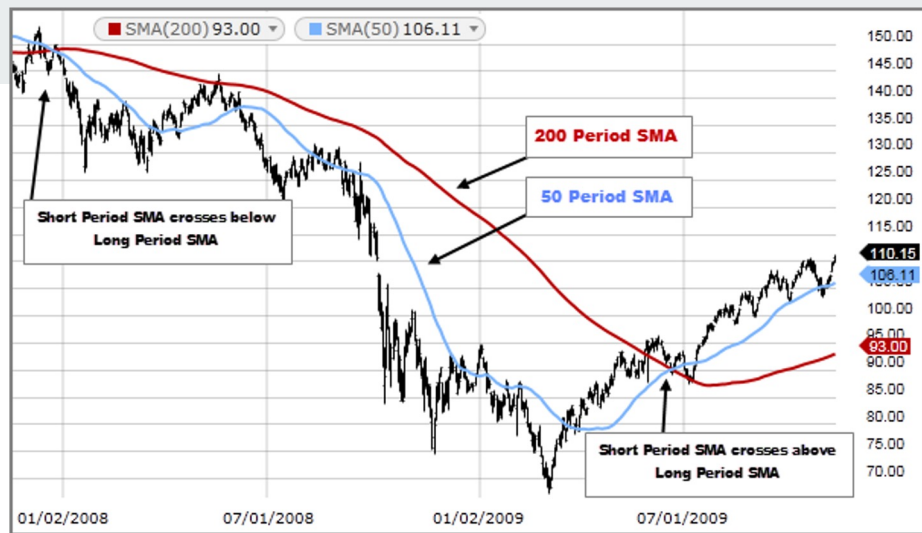


## Generating New Features

新feature：

股市技術分析常用的指標：

1. SMA\_diff 解釋短長期K線差距
2. RSI 解釋市場熱度(value = 0~100, 0:近lag分鐘都跌，100:近lag分鐘都漲)
3. bbands 解釋現在的價格與近lag分鐘價格相比是percentile多少
4. ATR 代表真實波動區間，與bbands很像，但多考慮了價格並非連續變動，因此做出調整
5. NATR = ATR/ 此分鐘收盤價

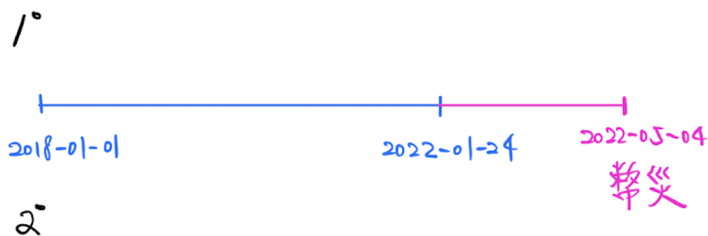


# Timeline

## 1. 預測幣災前

train & validation: 2018\_01\_01 ~ 2022-01-24

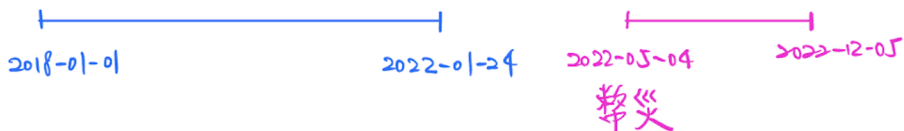
test: 2022-01-24 ~ 2022-05-24



## 2. 預測幣災

train & validation: 2018\_01\_01 ~ 2022-01-24

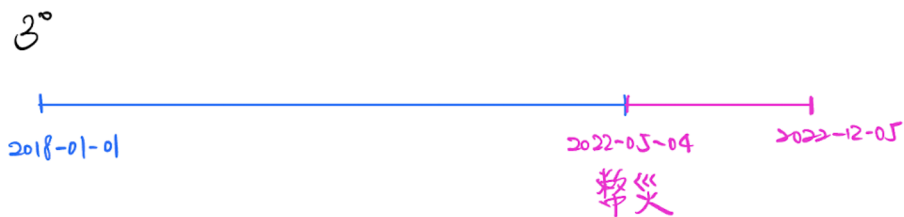
test: 2022-05-24 ~ 2022-12-04



## 3. 加新資料預測幣災

train & validation: 2018\_01\_01 ~ 2022-05-24

test: 2022-05-24 ~ 2022-12-04





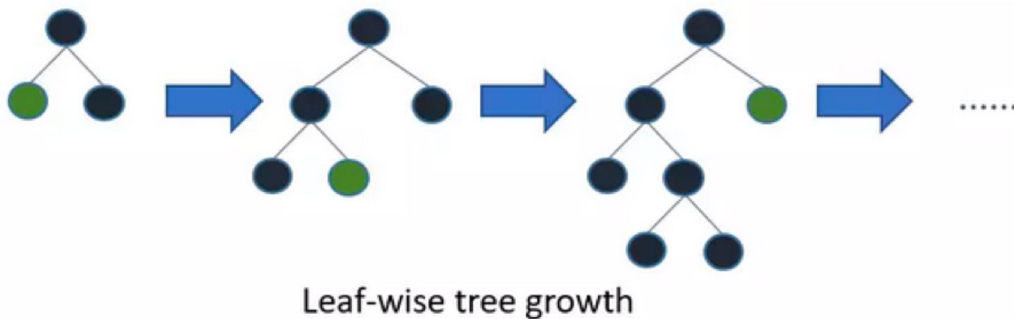
# Model

# Model: Single model of LightGBM (7-fold CV)

## LightGBM 優點

LightGBM 其主要想法是利用決策樹為基底的弱學習器，不斷地迭代訓練並取得最佳的模型。同時該演算法進行了優化使得訓練速度變快，並且有效降低被消耗的資源。

- 更快的訓練速度和更高的效率
- 低記憶體使用率
- 更好的準確度
- 支援 GPU 平行運算
- 能夠處理大規模數據



LightGBM 使用 leaf-wise tree 演算法，因此在迭代過程中能更快地收斂。  
<https://ithelp.ithome.com.tw/articles/10206014>  
<https://ithelp.ithome.com.tw/m/articles/10274577>



## Model types

1. model 1: 原作者(lgb)
2. model 2: 原作者(lgb) + 新feature
3. model 3: 原作者(lgb) + 新feature + 多model(lgb)



# Performance Analysis





# 1. 原作者(lgb) 在 三個不同時間下的預測

## 1. 預測幣災前

bit RMSE = 0.001403743

eth RMSE = 0.001525924

## 2. 預測幣災

bit RMSE = 0.00148702

eth RMSE = 0.001799466

## 3. 加新資料預測幣災

bit RMSE = 0.00139661

eth RMSE = 0.001745011

這邊可以看到 model 1 在預測幣災時的能力變差，RMSE皆變大，但在新增了額外的data後，預測能力有變好

至於貨幣方面的預測是：比特幣表現比以太幣好



## 2. 原作者(lgb) + 新feature 在三個不同時間下的預測

### 1. 預測幣災前

bit RMSE = 0.001402193

eth RMSE = 0.001525162

### 2. 預測幣災

bit RMSE = 0.00146021

eth RMSE = 0.00176843

### 3. 加新資料預測幣災

bit RMSE = 0.00139561

eth RMSE = 0.001743911

這邊可以看到 model 2 在預測幣災時的能力變差，RMSE皆變大，但在新增了額外的data後，預測能力有變好

至於貨幣方面的預測是：比特幣表現比以太幣好，結果與 model 1 相似



### 3. 原作者(lgb) + 新feature + 多model(lgb)

#### 1. 預測幣災前

bit RMSE = 0.001401184

eth RMSE = 0.001521367

#### 2. 預測幣災

bit RMSE = 0.00145767

eth RMSE = 0.001766942

#### 3. 加新資料預測幣災

bit RMSE = 0.00139176

eth RMSE = 0.001741172

一樣的可以看到 model 3 在預測幣災時的能力變差，RMSE皆變大，但在新增了額外的data後，能力有變好

至於貨幣方面的預測是：比特幣表現比以太幣好，結果與 model 1 和 model 2 相似



## 4. 預測幣災前 比較3種 model

1. model 1: 原作者(lgb)

bit RMSE = 0.001403743

eth RMSE = 0.001525924

2. model 2: 原作者(lgb) + 新feature

bit RMSE = 0.001402193

eth RMSE = 0.001525162

3. model 3: 原作者(lgb) + 新feature + 多model(lgb)

bit RMSE = 0.001401184

eth RMSE = 0.001521367

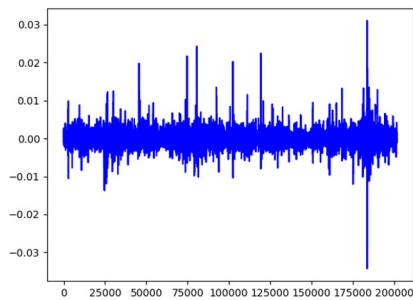
這邊可以看到3個model 在預測幣災的能力差不多，而在RMSE上model 3 < model 2 < model 1

然後貨幣的預測是比特幣表現比以太幣好

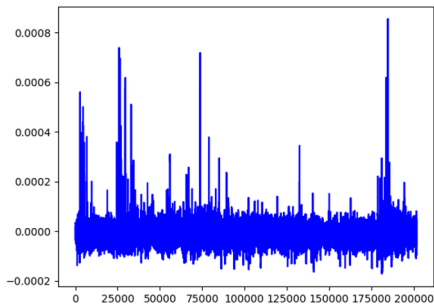
# 預測幣災前 Bitcoin (example)



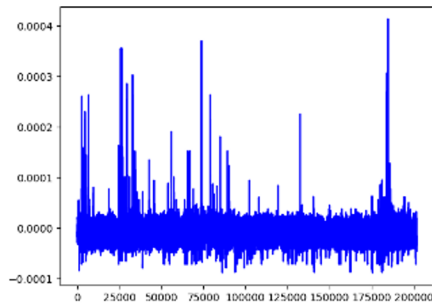
True Y



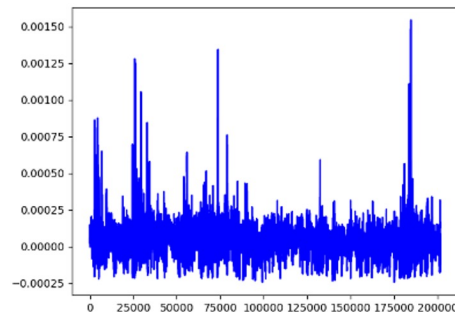
model 1 predict Y



model 2 predict Y



model 3 predict Y





## 5. 預測幣災 比較3種 model

1. model 1: 原作者(lgb)

bit RMSE = 0.00148702

eth RMSE = 0.001799466

2. model 2: 原作者(lgb) + 新feature

bit RMSE = 0.00146021

eth RMSE = 0.00176843

3. model 3: 原作者(lgb) + 新feature + 多model(lgb)

bit RMSE = 0.00145767

eth RMSE = 0.001766942

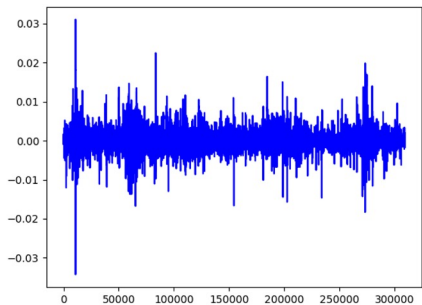
這邊可以看到3個model 在預測幣災時的能力差不多，而在 RMSE上model 3 < model 2 < model 1

至於貨幣的預測是：比特幣表現比以太幣好

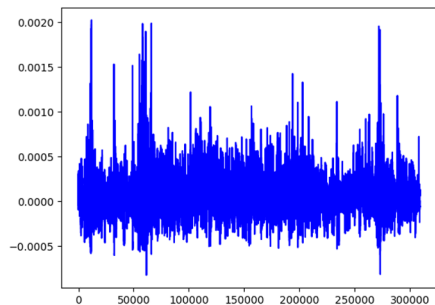
# 預測幣災 Bitcoin (example)



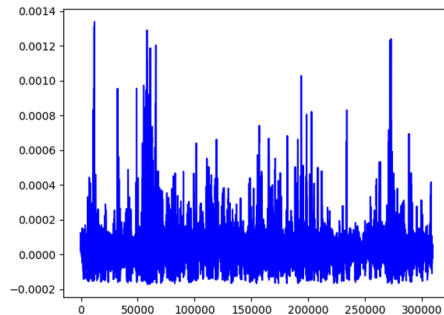
True Y



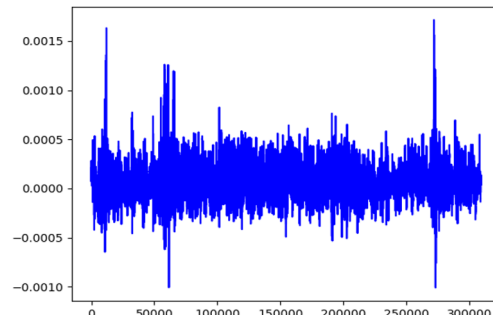
model 1 predict Y



model 2 predict Y



model 3 predict Y





## 6. 加新資料預測幣災 比較三種 model

1. model 1: 原作者(lgb)

bit RMSE = 0.00139661

eth RMSE = 0.001745011

2. model 2: 原作者(lgb) + 新feature

bit RMSE = 0.00139561

eth RMSE = 0.001743911

3. model 3: 原作者(lgb) + 新feature + 多model(lgb)

bit RMSE = 0.00139176

eth RMSE = 0.001741172

可以看到 三個model 在預測幣災的能力就有比較明顯一點的差別，RMSE上model 3 < model 2 < model 1

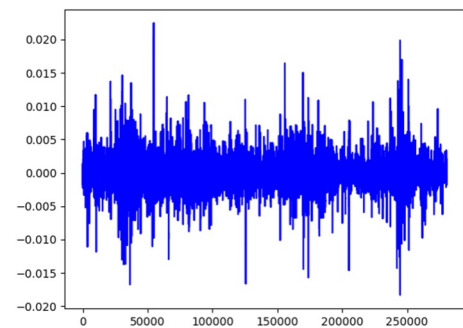
貨幣的預測是比特幣表現比以太幣好



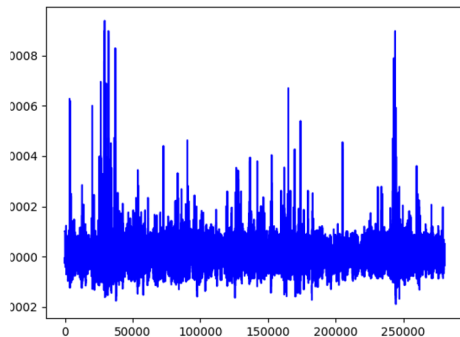
# 加新資料預測幣災 Bitcoin (example)



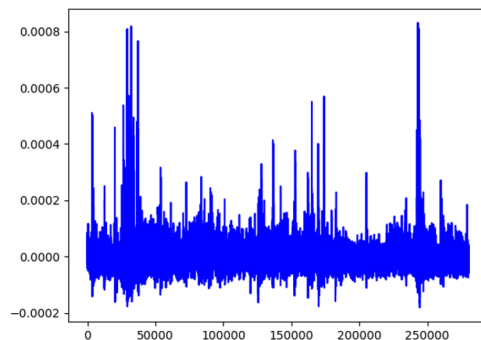
True Y



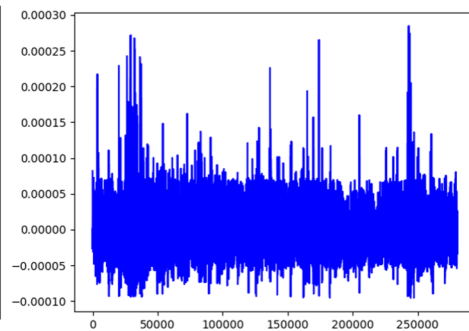
model 1 predict Y



model 2 predict Y



model 3 predict Y





# Bitcoin

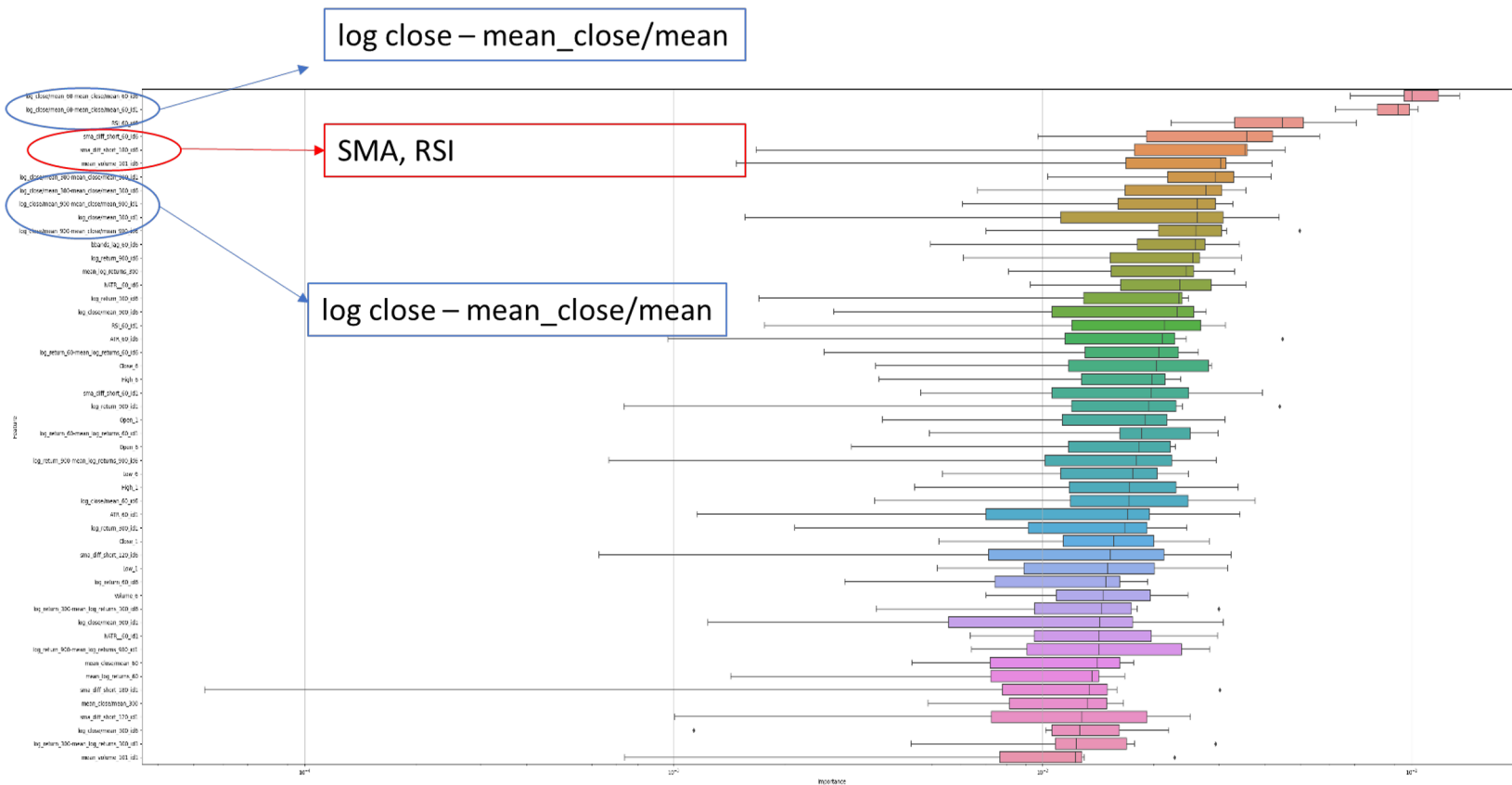
	幣災前	幣災	加新資料預測幣災
model 1 RMSE	0.001403743	0.00148702	0.00139661
model 2 RMSE	0.001402193	0.00146021	0.00139561
model 3 RMSE	0.001401184	0.00145767	0.00139176



# Ethereum

	幣災前	幣災	加新資料預測幣災
model 1 RMSE	0.001525924	0.001799466	0.001745011
model 2 RMSE	0.001525162	0.00176843	0.001743911
model 3 RMSE	0.001521367	0.001766942	0.001741172

# Feature Importance





## Some Top Feature Importance

`log_close/mean_{60}-mean_close/mean_{60}_id{6}`

`log_close/mean_{60}-mean_close/mean_{60}_id{1}`

發現在 train不同時期與不同model時，這兩個都是非常重要的指標



# Improvement



## 1. 用全部14種幣來訓練模型

由於資料量太過龐大，記憶體無法處理

我們只選出了其中兩種幣：比特幣 以及 以太幣，來train model，

但其實有可能其他貨幣也會影響比特幣和以太幣的走勢，但我們就沒有考慮到。

故之後的改進可能會想加入其他具代表性的Model進來train。



## 2. 模型的改善

聽完上次演講者所說再經過討論後，我們這邊想到了一些模型的改進方法

像是用CNN or RNN，

CNN用來做pattern recognition；

RNN可以加入time series, Autoregression跟LSTM，

多一些判斷趨勢的部分，來讓模型能夠學習得更完善。





THANK YOU !