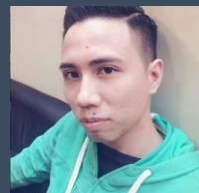


Stock In the Game

AI技術應用工程師班 第一組

開發主管

官定憶



開發人員

洪子惟



林恆成



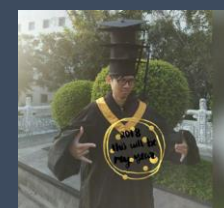
李家豪



陳柏安



黃紀凡

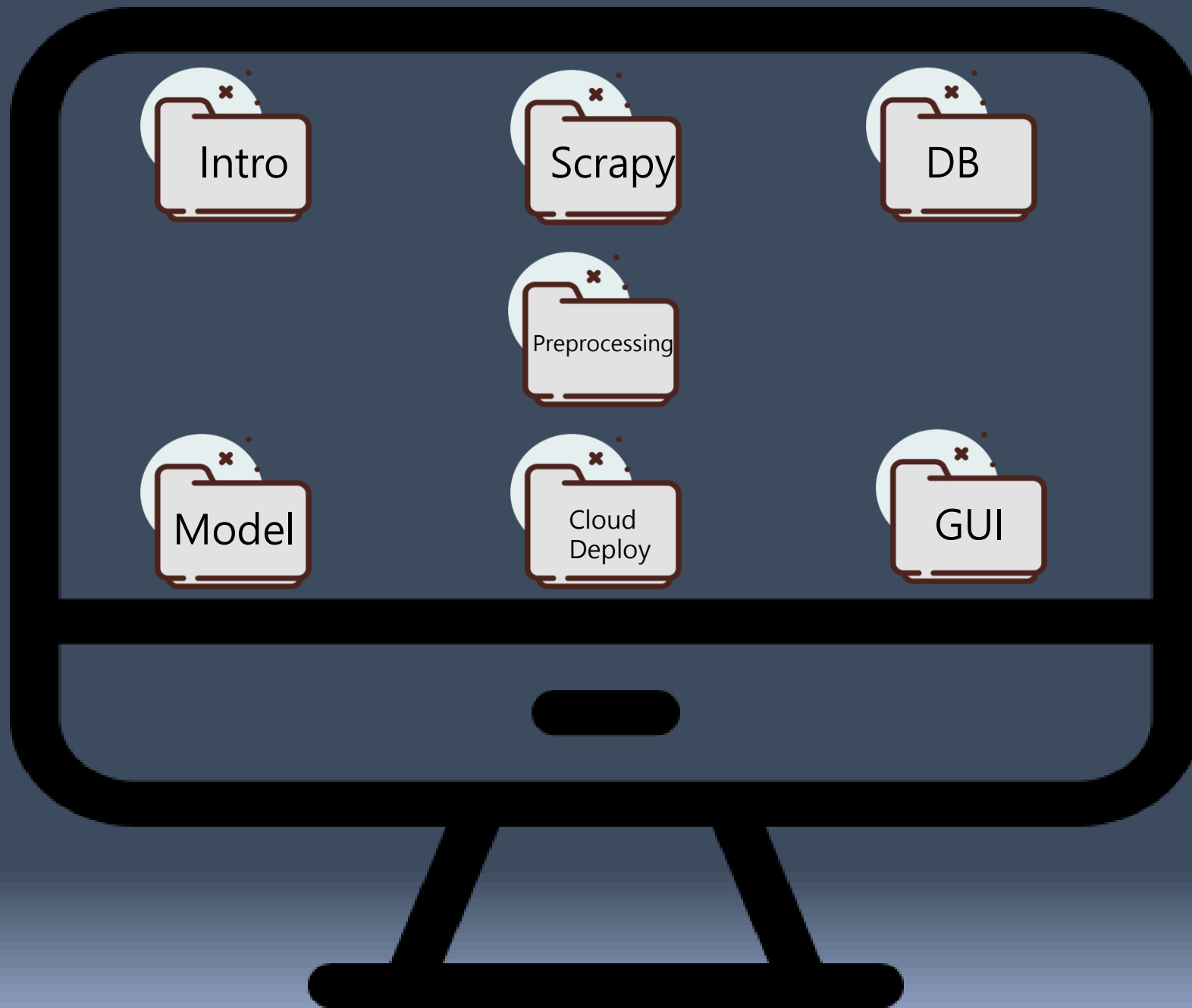


工作分配

分工 組員	爬蟲	Preprocessing	Model	Cloud	News	GUI	DataBase	Structure	PPT
官定憶	@@@		@@@					@@@	
洪子惟	@		@@@	@@@			@@@	@@	
李家豪	@@@					Backend @@@			@@@
林恆成	@@@	@@@	@@						@
陳柏安	@					Backend+d esign @@@			@@@
黃紀凡	@@				@@@	Design @@	@@		

報告流程





Intro

Scrapy

DB

Preprocessing

Model

Cloud
Deploy

GUI



動機 Why

選題

在這薪水成長緩慢，房價高升的年代
我們不能只以技術、勞力賺錢

必須以錢賺錢

最小的成本 最大的回報

發大財

自動交易程式差異

市面上:

- 一般是以價格界線判斷買賣
- 漲多少%停利，跌多少%停損
- 現在設定的規則未來不適用
- 股票瞬息萬變，需要人為操作

本系統:

- 結合AI判斷最佳買賣點
- 以每5秒、每天的資料會不斷更新

目標客群

Who

股市新手

股市散戶

金融巨鱷

波克夏·海瑟威

想要透過股市賺錢的人

成果展示

系統架構

Scrapy



Requests

Breaking

中美貿易戰
台灣挖到石油



Database



PostgreSQL



Data
Preprocessing



Finance Data

+



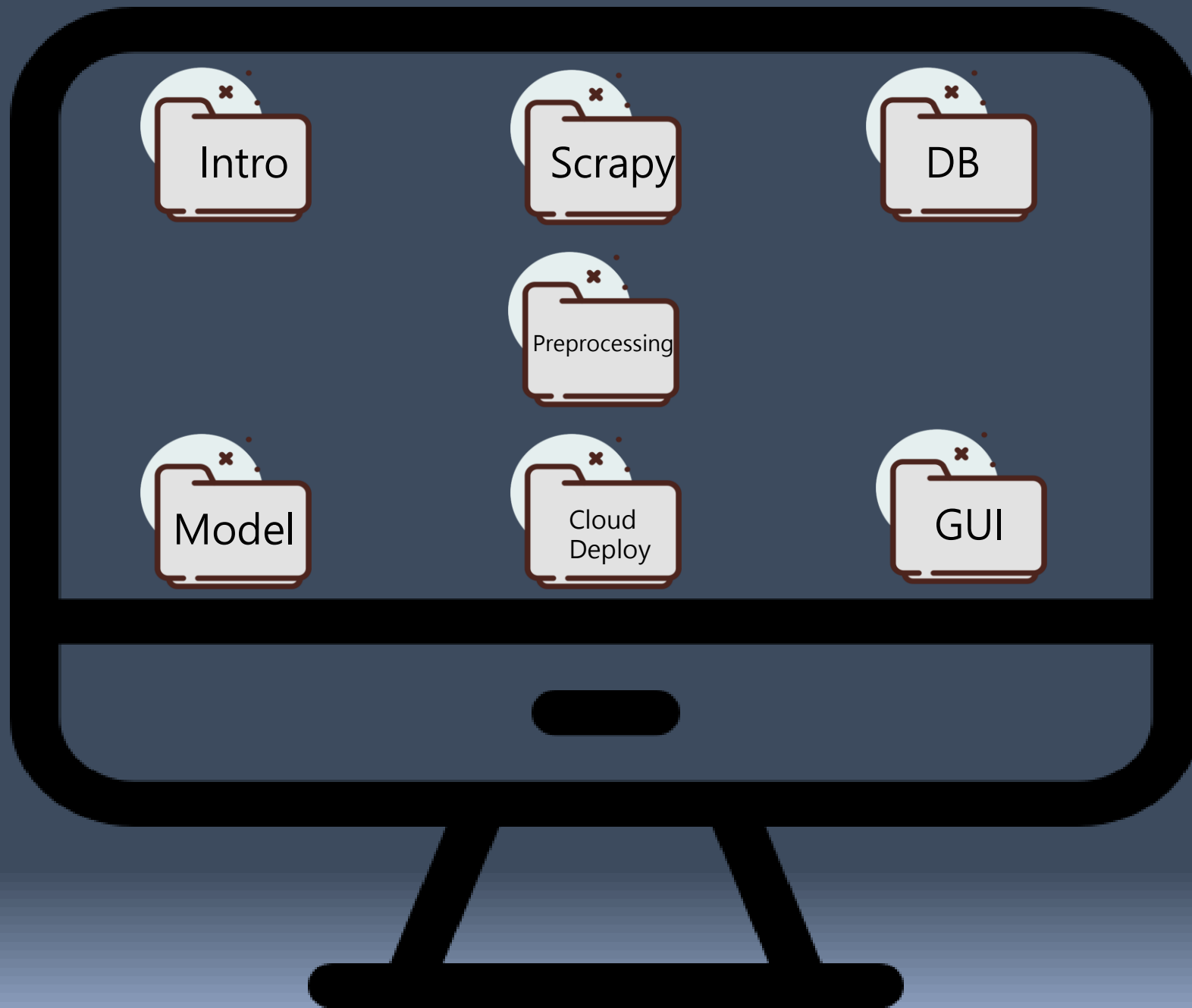
NLP

Model training



GUI





Intro

Scrapy

DB

Preprocessing

Model

Cloud
Deploy

GUI

黃紀凡

報告人:



作法 How

爬蟲資料

Yahoo

- 委買賣

證交所

- 委買賣
- 日成交
- 法人
- 融資融券

鉅亨網

- 新聞

Investing

- 外國股票
- 期貨
- 外匯

股票取前50筆高交易量為主

問題 解決方法

無法取得歷年五買五賣 取得資料要320萬!!!!!!

自去年12月開始爬取五買五賣資料

開盤10分鐘資料較難爬取

利用多線程多進程爬取多IP及Data

spy one

快代理
kuaidaili.com

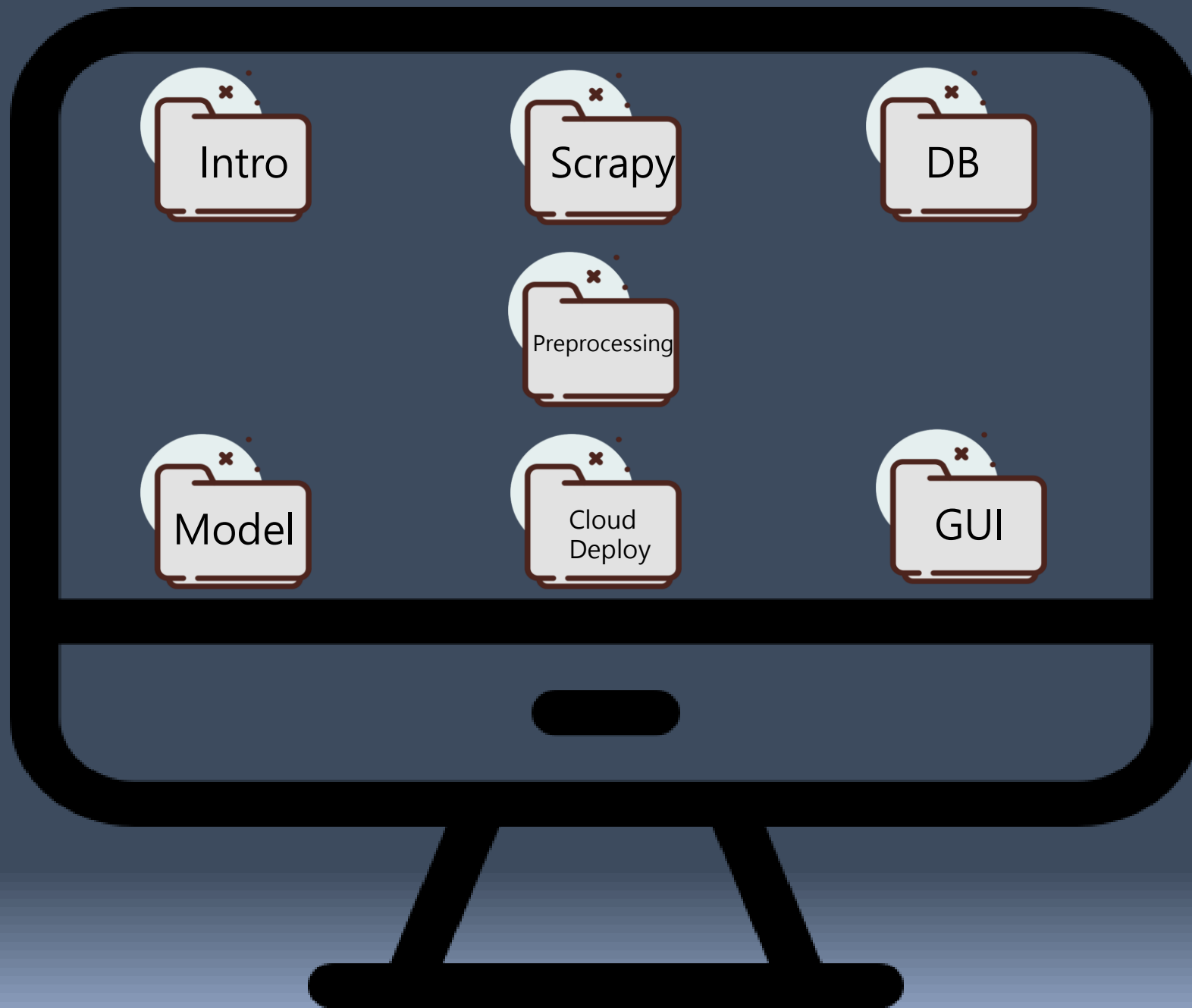
IP3366.NET
云代理

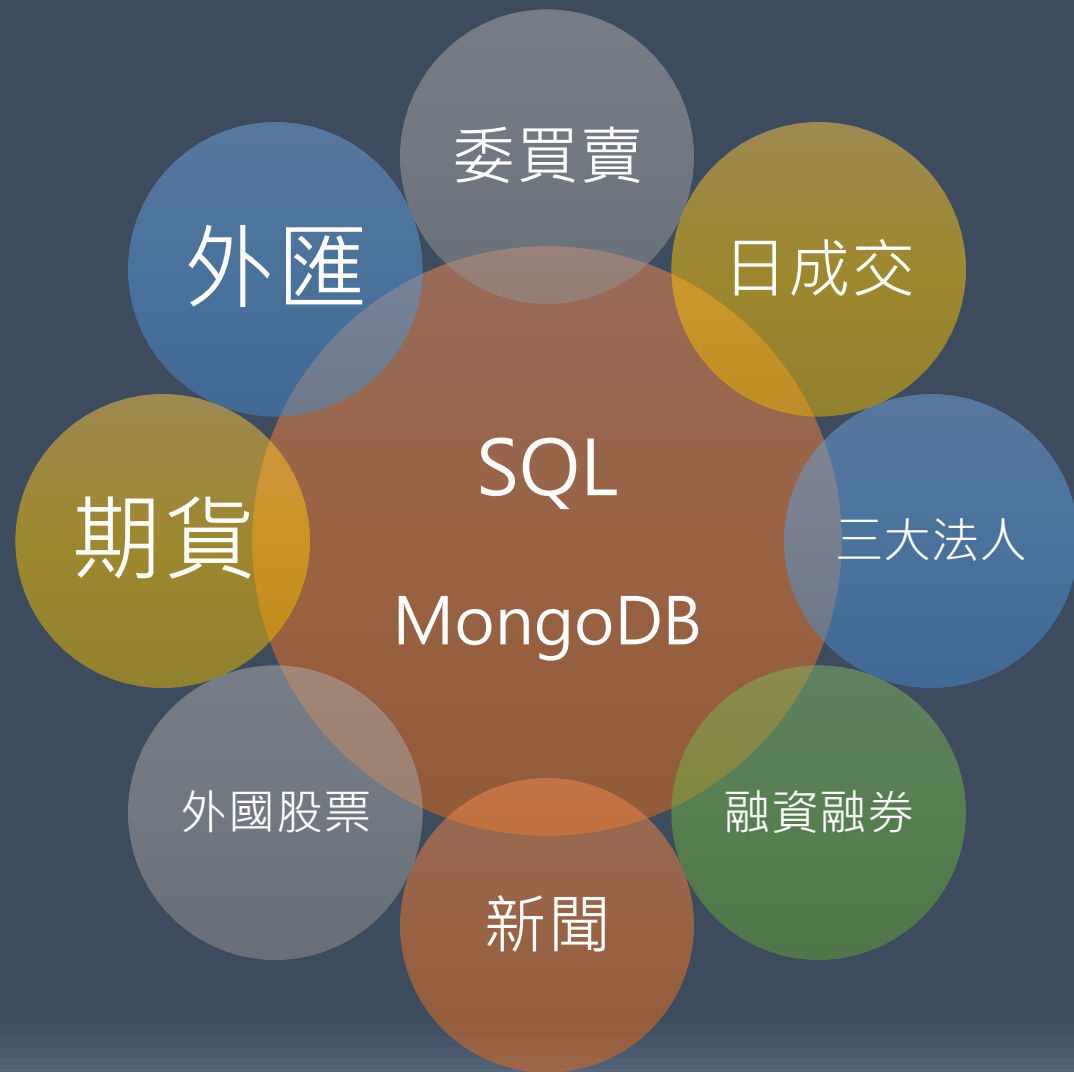
PROXY
RANKER



TAIWAN
STOCK EXCHANGE
臺灣證券交易所









	id ▾	timestamp ▾ 1	stock_id ▾	strike_price ▾	open_price ▾	low_price ▾	high_price ▾
1	5030410	2020-03-06 05:31:00.000000	2379	236	235	235	238.5
2	5030411	2020-03-06 05:31:00.000000	2338	28.95	29.2	28.8	29.5
3	5030412	2020-03-06 05:31:00.000000	2481	25.4	25.6	25.25	25.65
4	5030413	2020-03-06 05:31:00.000000	4968	170	167.5	166	171
5	5030414	2020-03-06 05:31:00.000000	1907	13.5	13.65	13.45	13.7
6	5030404	2020-03-06 05:30:59.000000	2409	9.34	9.53	9.3	9.55
7	5030405	2020-03-06 05:30:59.000000	3545	33.05	30.95	30.95	33.65
8	5030406	2020-03-06 05:30:59.000000	5203	127.5	125	125	131
9	5030407	2020-03-06 05:30:59.000000	2890	12.8	12.9	12.8	12.95
10	5030408	2020-03-06 05:30:59.000000	2417	12.85	13.1	12.8	13.35
11	5030409	2020-03-06 05:30:59.000000	2408	78.5	78.7	77.5	79.5
12	5030398	2020-03-06 05:30:58.000000	2456	108	110.5	108	110.5
13	5030399	2020-03-06 05:30:58.000000	3661	240	232.5	232.5	241.5
14	5030400	2020-03-06 05:30:58.000000	2344	16.25	16.45	16.2	16.6
15	5030401	2020-03-06 05:30:58.000000	3045	106	107.5	105	107.5
16	5030402	2020-03-06 05:30:58.000000	2812	12.25	12.3	12.2	12.3
17	5030403	2020-03-06 05:30:58.000000	2303	15.45	15.55	15.4	15.6
18	5030392	2020-03-06 05:30:57.000000	3530	114	112.5	111	114
19	5030394	2020-03-06 05:30:57.000000	3016	53.5	55.2	53.5	56.3
20	5030395	2020-03-06 05:30:57.000000	2363	7.98	7.86	7.77	7.98

s.public.best_5_test [postgres@35.2

-500 v of 4,203,548 > >| 5

420 多萬筆

id ↑ timestamp

News_db.newsdbDocuments

News_db.newsdb

DOCUMENTS 61.1kTOTAL SIZE 119.3MBAVG. SIZE 2.0KBINDEXES 4TOTAL SIZE 177.1MBAVG. SIZE 44.3MB

DocumentsAggregationsExplain PlanIndexes

FILTER

OPTIONS

FIND

RESET

...

ADD DATA

VIEW

Displaying documents 1 - 20 of 61120

REFRESH

_id: ObjectId("5e1570e07ec5f2e0212d8bb4")

title: "鴻海強化邊緣AI應用 鎖定智慧製造及零售市場"

newsId: 4431000

stock: Array

time: 2020-01-08T13:15:20.000+00:00

content: "鴻海 (2317-TW) 與日本系統晶片 (SoC) 業者索尼未來 (Socionext)、美系 IP 影像軟體商 Network Opti..."

hour: 13

_id: ObjectId("5e1570e07ec5f2e0212d8bb5")

title: "國旅補助發威 亞都、老爺知去年12月營收衝高"

newsId: 4431004

stock: Array

time: 2020-01-08T13:13:27.000+00:00

content: "秋冬旅遊補助政策下激勵飯店業者業績表現，業者公布的去年 12 月合併營收來看，陸續繳出亮眼成績，亞都 (5703-TW) 達 8442 萬元..."

hour: 13

_id: ObjectId("5e1570e07ec5f2e0212d8bb6")

title: "材料-KY今年添新產能 年營收逐季走揚估年增15%"

newsId: 4430987

stock: Array

time: 2020-01-08T12:54:54.000+00:00

content: "材料 - KY(4763-TW) 擴產效益顯現，去年第 4 季營收 6.07 億元創新高，2019 年全年營收 21.75 億元，也改寫年..."

hour: 12

_id: ObjectId("5e1570e07ec5f2e0212d8bb7")

title: "〈CES大展〉台科技廠新品亮相 面板廠搶車用商機 筆電廠續攻電競"

newsId: 4430993

stock: Array

time: 2020-01-08T12:52:03.000+00:00

DB

資料庫

► OPTIONS

FIND

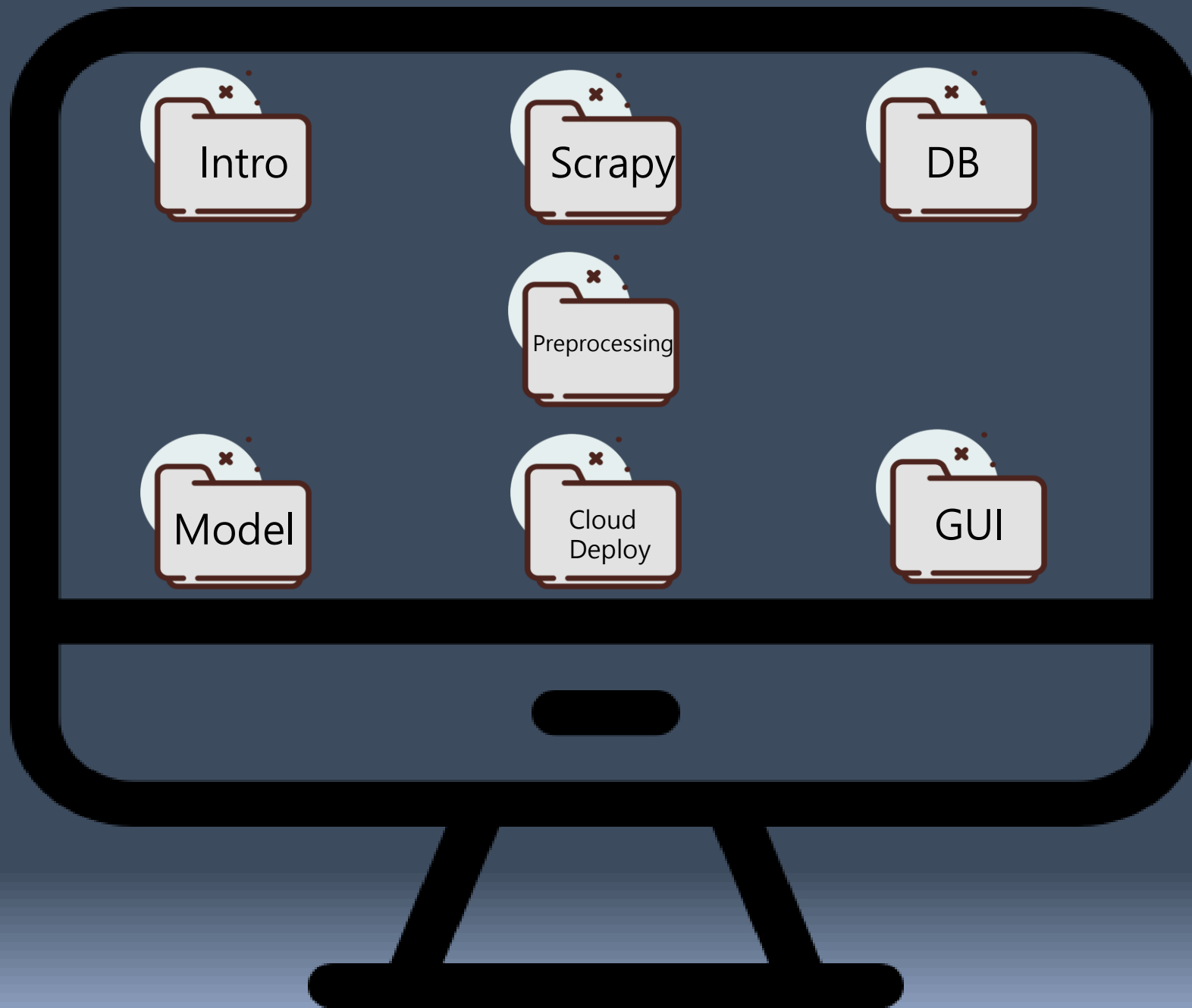
RESET

playing documents 1 - 20 of 61120



⌂ REFRESH

6萬多筆



Intro

Scrapy

DB

Preprocessing

Model

Cloud
Deploy

GUI

林恆成

報告人:



Preprocessing

前處理

將資料處理成日單位資料
資料合併及分類
將資料轉換為技術指標
新聞正負面指標

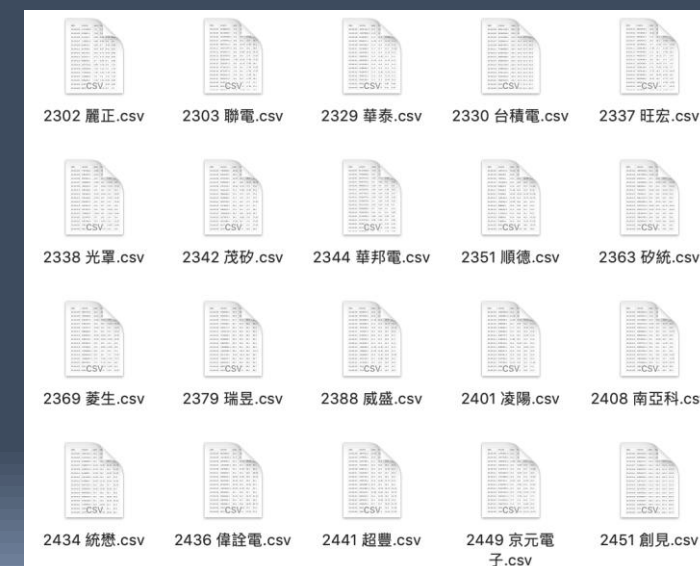
2010/01 Daily Trading Value/Volume of 2330			
Date	Trade Volume	Trade Value	Opening Price
2010/01/04	39,511,138	2,557,720,928	65.00
2010/01/05	38,394,084	2,464,115,096	65.00
2010/01/06	52,734,385	3,390,698,544	64.50
2010/01/07	53,294,614	3,437,221,996	64.90
2010/01/08	48,047,497	3,068,341,466	63.50

原始資料：
每筆(月)資料包含單股單月資料
Ex: 台積電2010年1月的股價

個股資料

將每月資料依個股合併

date	volume	open	high	low	close
20100127	53876237.0	60.1	60.4	59.5	59.6
20100128	44033633.0	60.5	61.5	59.8	60.0
20100129	98124608.0	60.1	61.5	59.4	61.5
20100201	82234872.0	60.7	60.7	59.5	59.8
20100202	72141345.0	60.2	60.4	59.0	59.0
20100203	42627861.0	60.2	60.3	58.8	59.8
20100204	46032165.0	60.0	60.0	59.0	59.1
20100205	66295018.0	57.6	57.9	57.2	57.2



三大法人 融資融券

證券代號	證券名稱	date	投信買進股數(總)
2330	台積電	20041217	36,000
2330	台積電	20041220	280,000
2330	台積電	20041221	200,000

將每日資料
依個股分裝

093年12月17日 投信買賣超彙總表				
證券代號	證券名稱	買進股數	賣出股數	買賣超股數
2303	聯電	500,000	2,600,000	-2,100,000
2317	鴻海	20,000	313,000	-293,000
2330	台積電	36,000	1,377,345	-1,341,345



證券代號	證券名稱	date	投信買進股數(總)
2317	鴻海	20041217	20,000
2317	鴻海	20041220	95,000
2317	鴻海	20041221	60,000

原始資料：

每筆(日)資料包含所有個股資料
Ex:2004年12月17日每股投信買賣超

證券代號	證券名稱	date	投信買進股數(總)
2303	聯電	20041217	500,000
2303	聯電	20041220	735,000
2303	聯電	20041221	0

依日期合併



股市關聯指標

~~基本面：公司基本營運~~

籌碼面：大戶資金流動

技術面：股價趨勢判定

消息面：新聞輿情影響

技術指標

股市操盤常看的技術指標：
移動平均、布林通道、KD指標、
相對強弱指數、動量指標...等共
計29項指標

```
def moving_average(dataframe, days = [5, 10, 20]):  
    for d in days:  
        dataframe['EMA{}'.format(d)] = abstract.EMA(dataframe, timeperiod = d)  
  
def bias(dataframe, days = [3, 6, 10, 25]):  
    series = dataframe['close']  
    for d in days:  
        dataframe['BIAS{}'.format(d)] = (series - sma(series, d)) * 100 / sma(series, d)
```


自定義分類

移動平均的概念是以當日相對前幾日價格的比較判斷弱點。

油門：當日數值與前幾日排名

方向盤：過去幾日上漲天數排名

分類變數包括：高低收價格、買賣超、融資融券餘額、外國指數

自定義分類

date	close
20100226	58.8
20100301	59.6
20100302	59.9
20100303	60.0
20100304	59.2
20100305	60.4
20100308	61.2
20100309	61.3
20100310	61.4
20100311	61.4
20100312	60.8

date	close_rank_5	close_increase_5
20100226	3	3
20100301	2	3
20100302	1	3
20100303	1	2
20100304	4	2
20100305	1	2
20100308	1	2
20100309	1	2
20100310	1	2
20100311	1	2
20100312	5	3

日期	2/26	3/1	3/2	3/3	3/4
價格	58.8	59.6	59.9	60	59.2
排序	5	3	2	1	4
上漲	--	✓	✓	✓	X

自定義分類

成交量(漲跌)

價格
(漲跌平)

融資融券餘額
(漲跌)

三大法人買賣
(買賣超)

消息面

從鉅亨網中爬取大量新聞資料及影響個股

手動標記200則新聞

藉由Model訓練生成正負面新聞指標

義總理簽封城令抗疫

危大利：水啦！



疫大利

喔：is it good to 疫？



消息面指標

網爆新iPhone規格！相機大升級

-100% 100%

NLP

	Predicted value	Size	Loss
Bert	98%	1.1GB	0.05
Albert(base)	91%	120MB	0.28
Albert(large)	93%	190MB	0.20

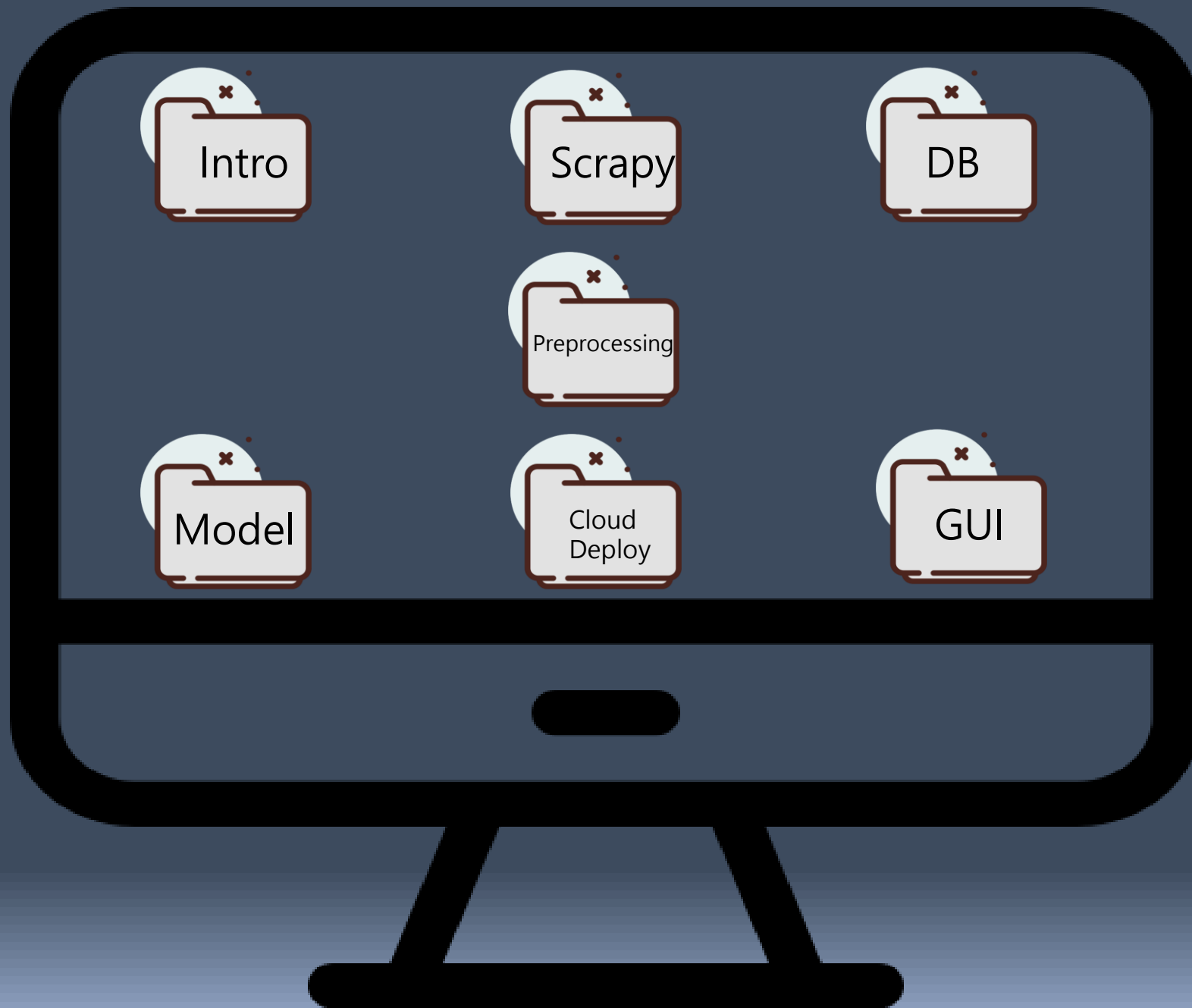
訓練資料

$X_{t-n,\dots,t}$

- 原始價量
- 技術指標
- 自定義分類指標
- 國外指數
- 期貨指數
- 外匯
- 新聞正負面

Y_{t+1}

- 隔日價格
- 隔日差額
- 隔日漲跌幅



Intro

Scrapy

DB

Preprocessing

Model

Cloud
Deploy

GUI

第三步

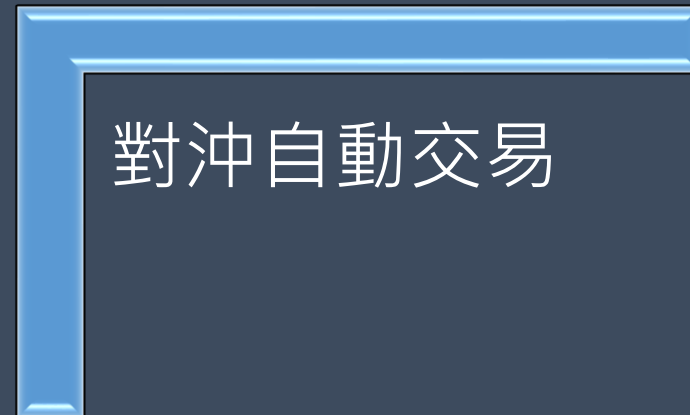
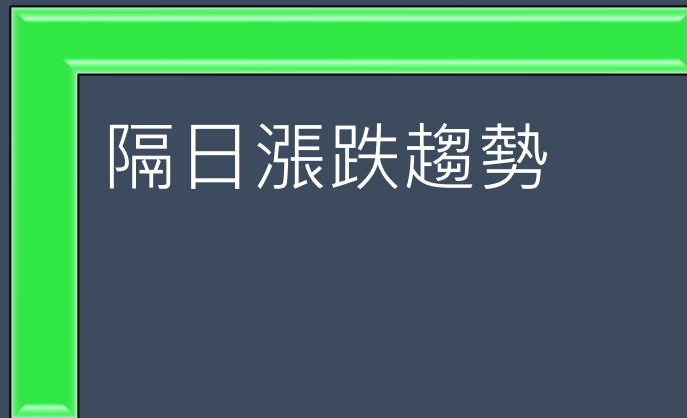
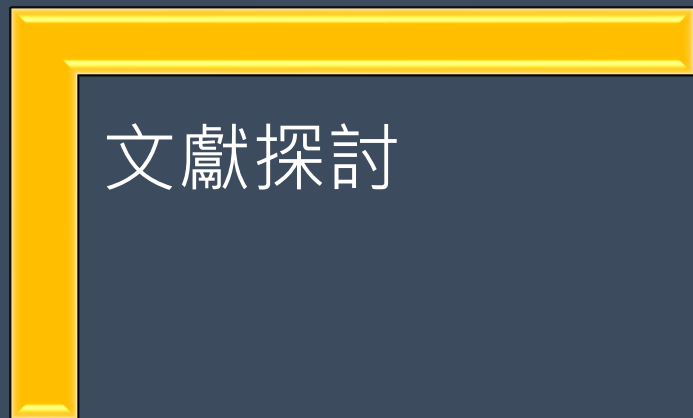
第二步

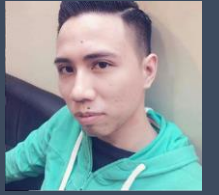
第一步

對沖自動交易

隔日漲跌趨勢

文獻探討





第一步 文獻探討

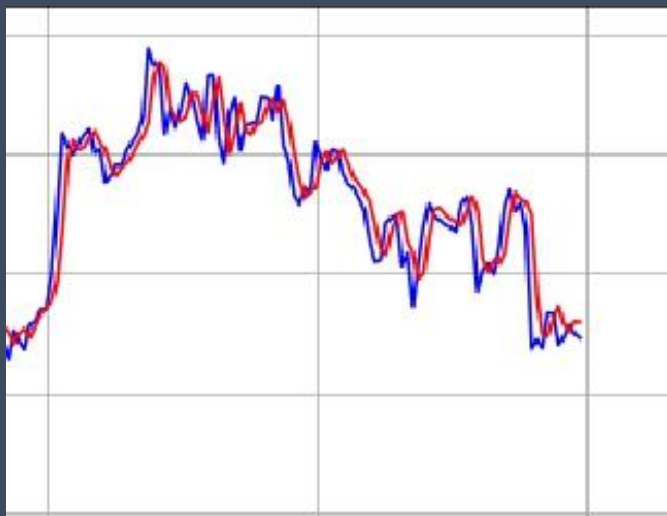
kaggle

arXiv.org

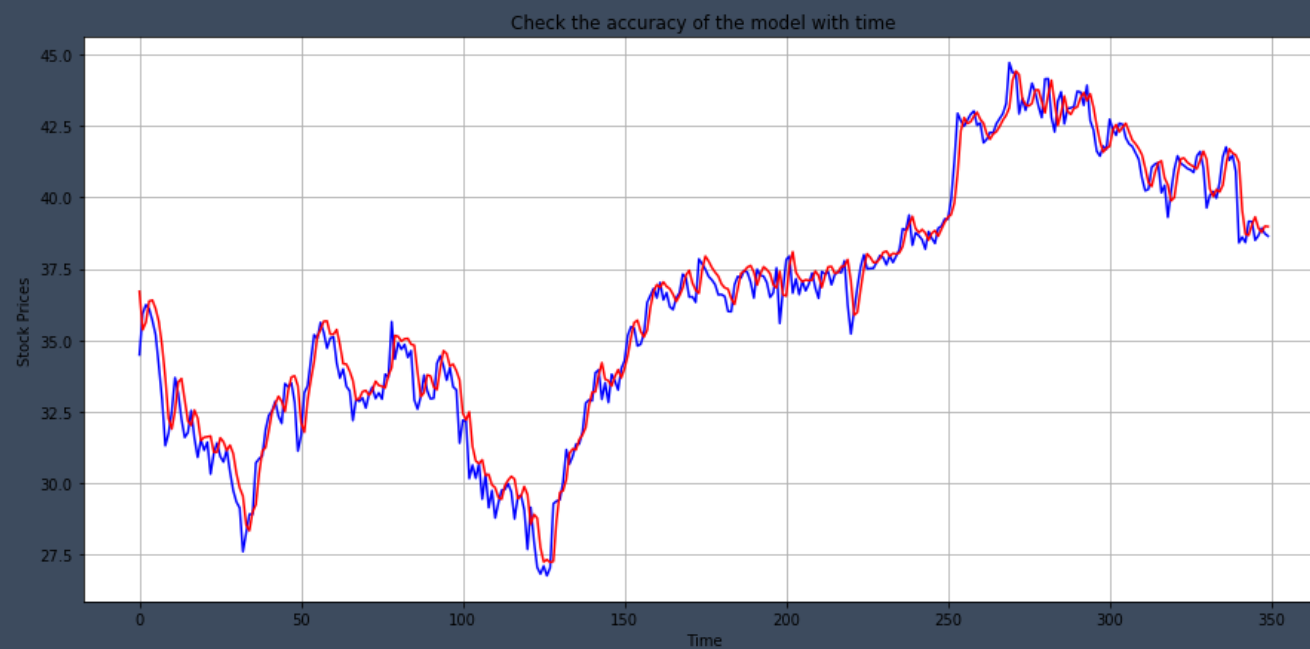
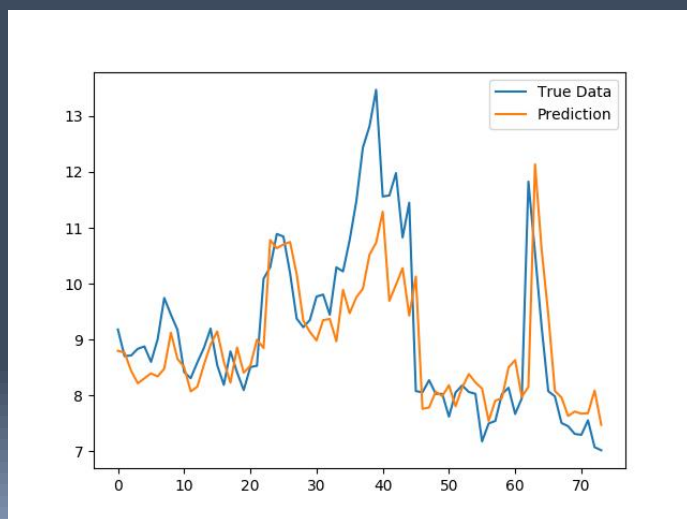


國家圖書館

National Central Library



Loss:0.0023
Acc:0.9831



Public
Your Work
Shared With You
Favorites
Sort by Hotness

Outputs
Languages
Tags
Search notebooks

194

Time-series plotting (Optional)
2y ago • tutorial

Py
18

135

NY Stock Price Prediction RNN LSTM GRU
2y ago • time series, lstm, rnn

Py
27

97

LSTM_Stock_prediction-20170507
10mo ago • finance, deep learning

Py
40

88

Predict stock prices with LSTM
3y ago • finance

Py
33

27

Stock Prediction
3y ago

Py
0

24

Best on Kaggle, Test MSE = 0.000328
2y ago

Py
4

17

predicting stock market using Linear Regression
2y ago • finance, forecasting, linear regression, future prediction

Py
1

5

Predict stock price using RNN with LSTM
24d ago • time series, lstm, rnn, stocks and bonds

Py
1

2

LSTM stock time series
1y ago

Py
0

10

AR, ARIMA, LSTM
9mo ago

Py
7

29

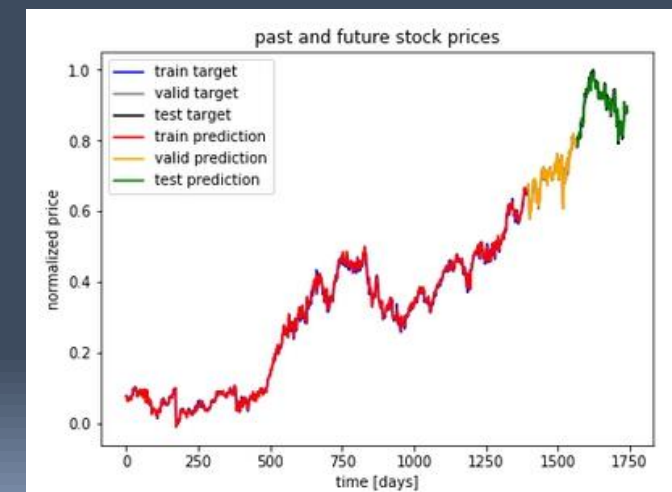
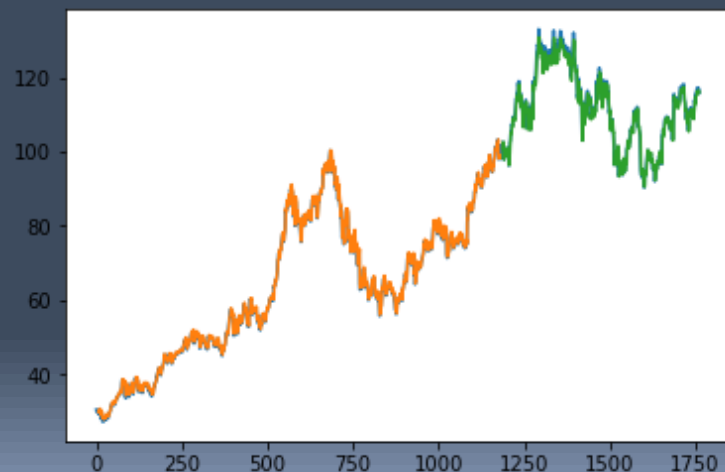
LSTM-GRU models for Stock Movement Analysis
2y ago • finance, data cleaning, classification, deep learning, lstm

Py
8

130

LSTM model of StockData
2y ago • time series, finance, regression analysis, deep learning

Py
27



問題 解決方法

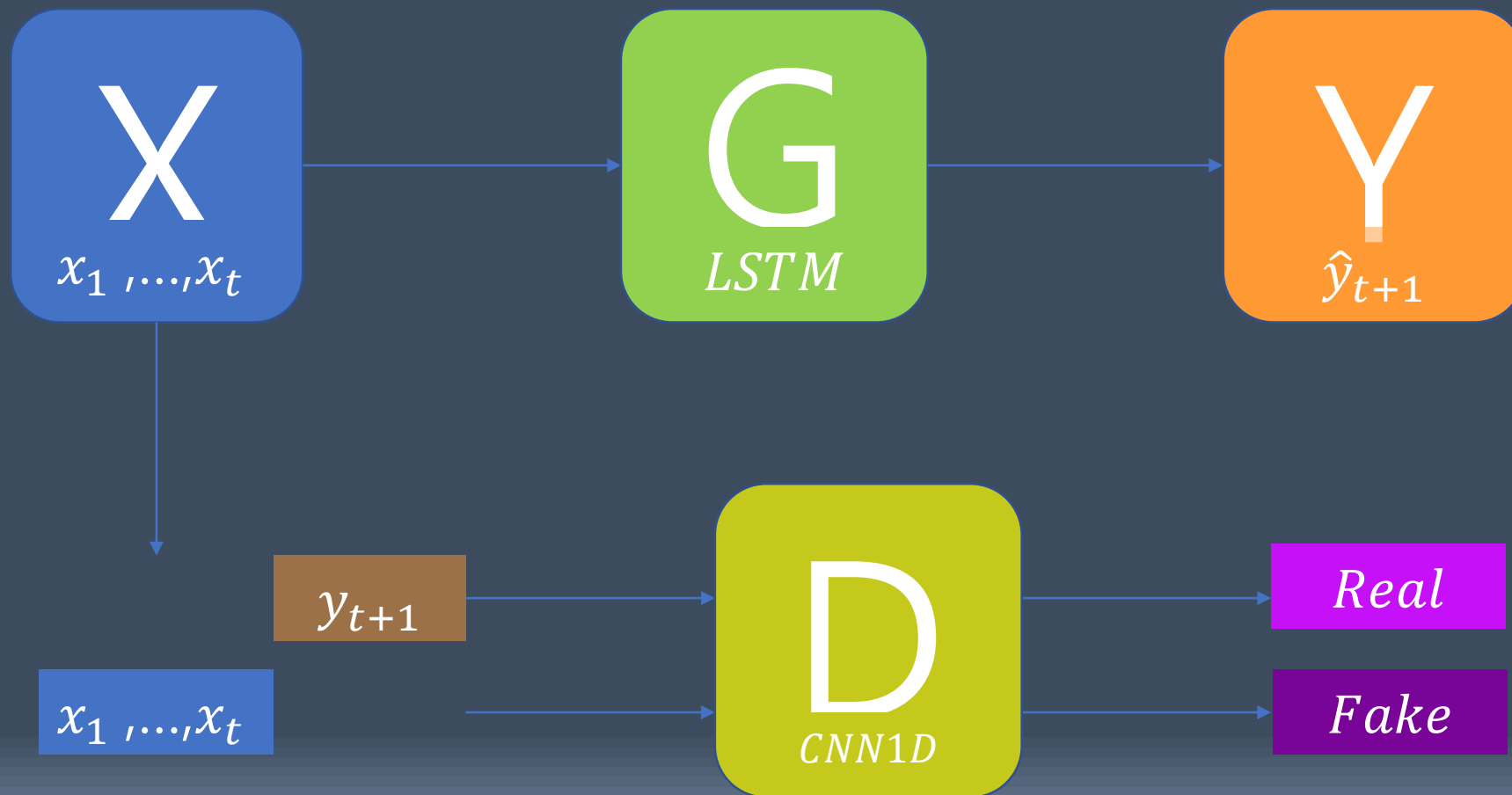
模型只是學會了下個點，會落在上個點差不多的位置

增加特徵及自己訓練模型

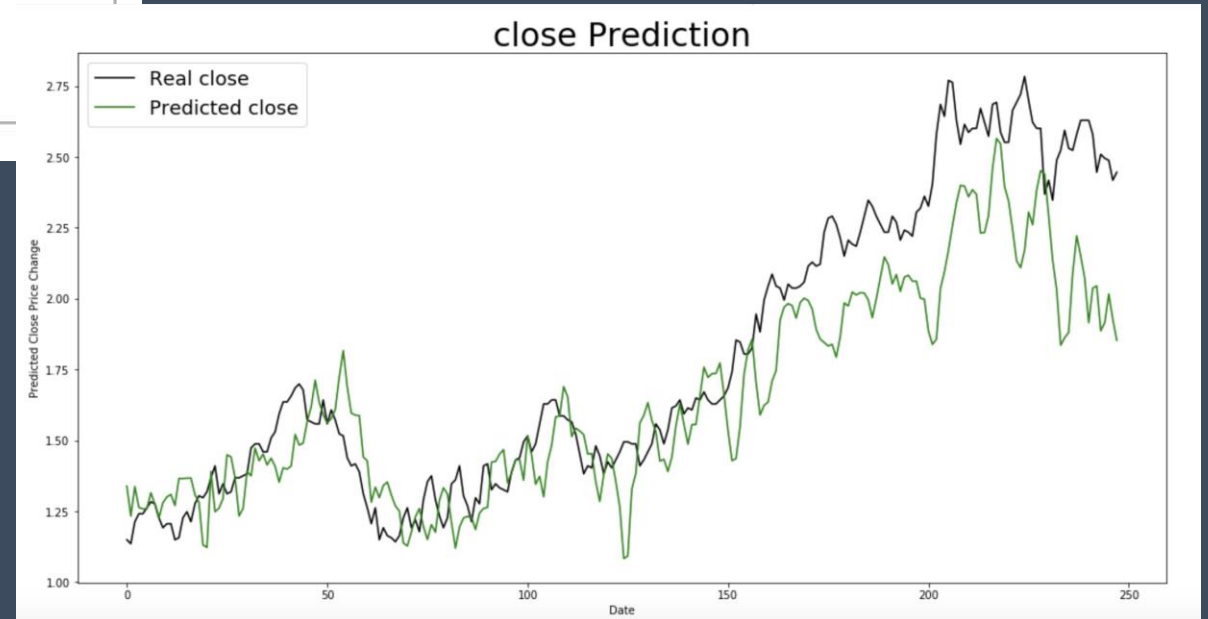
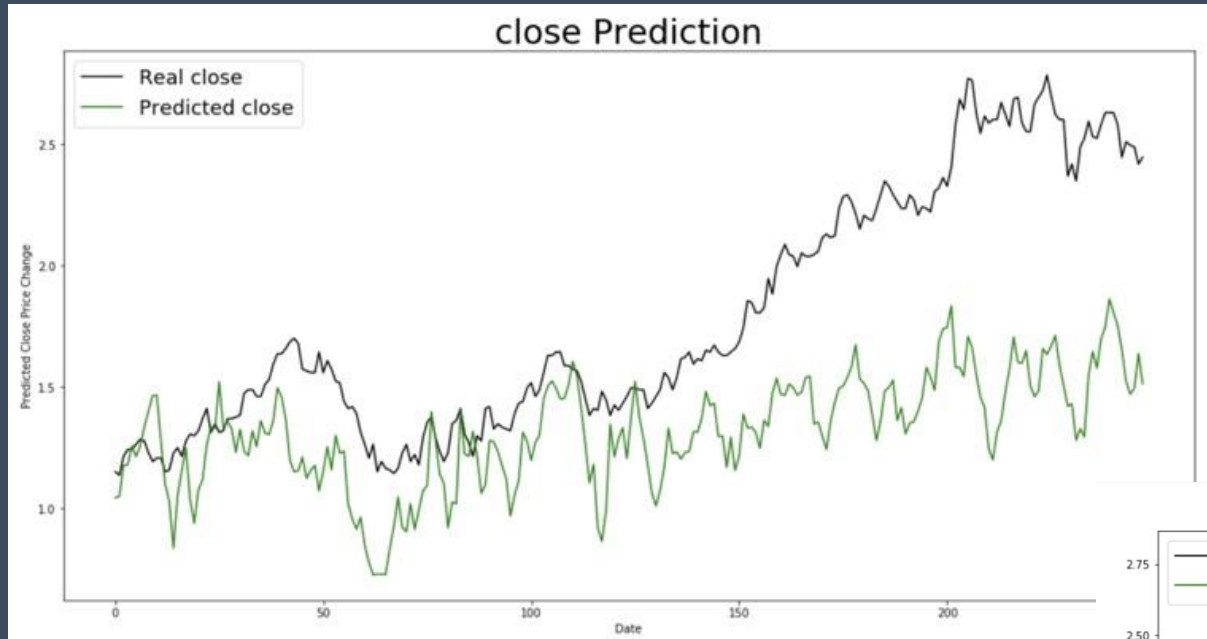
第二步

隔日落點及趨勢

GAN



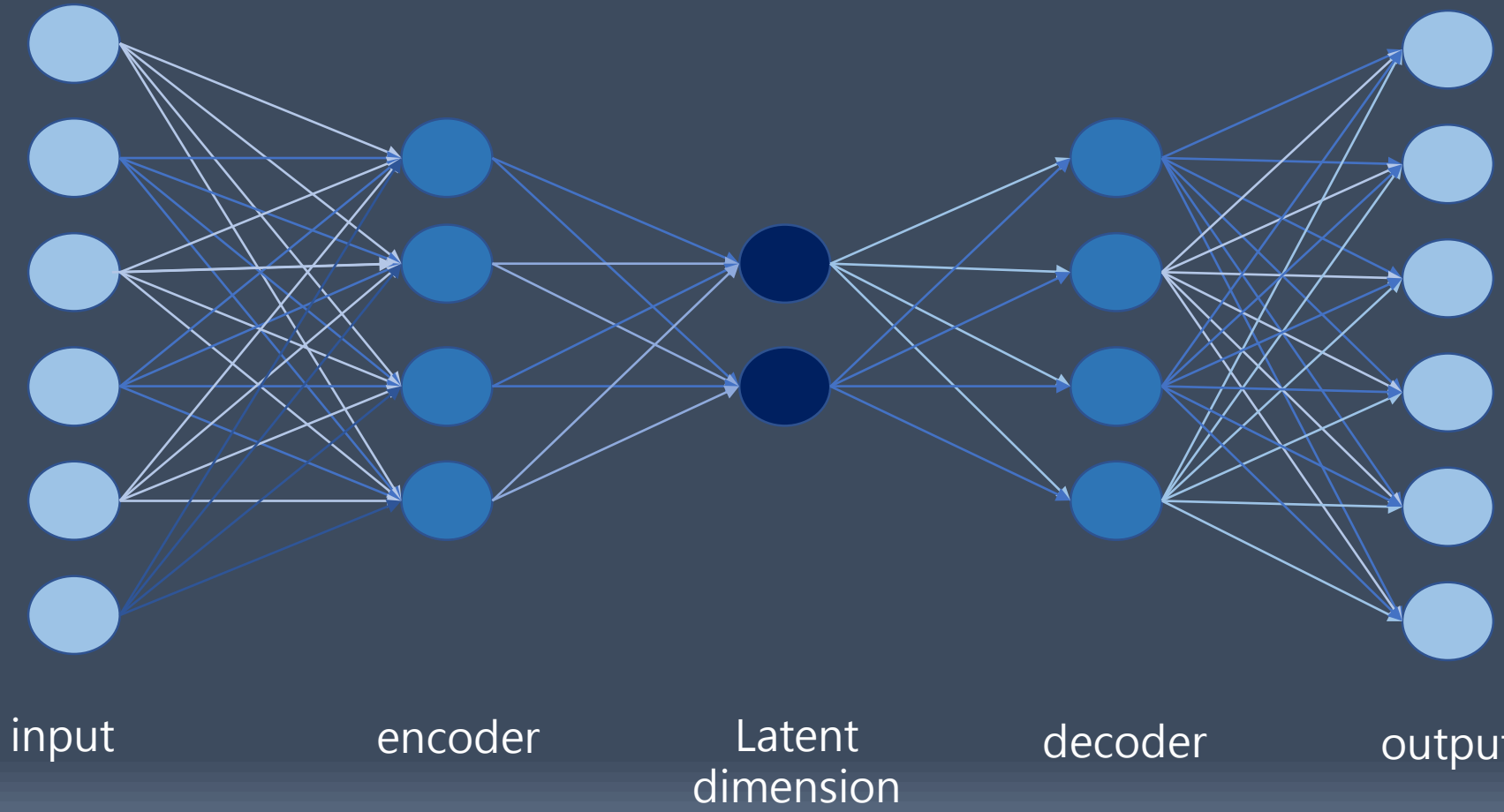
GAN

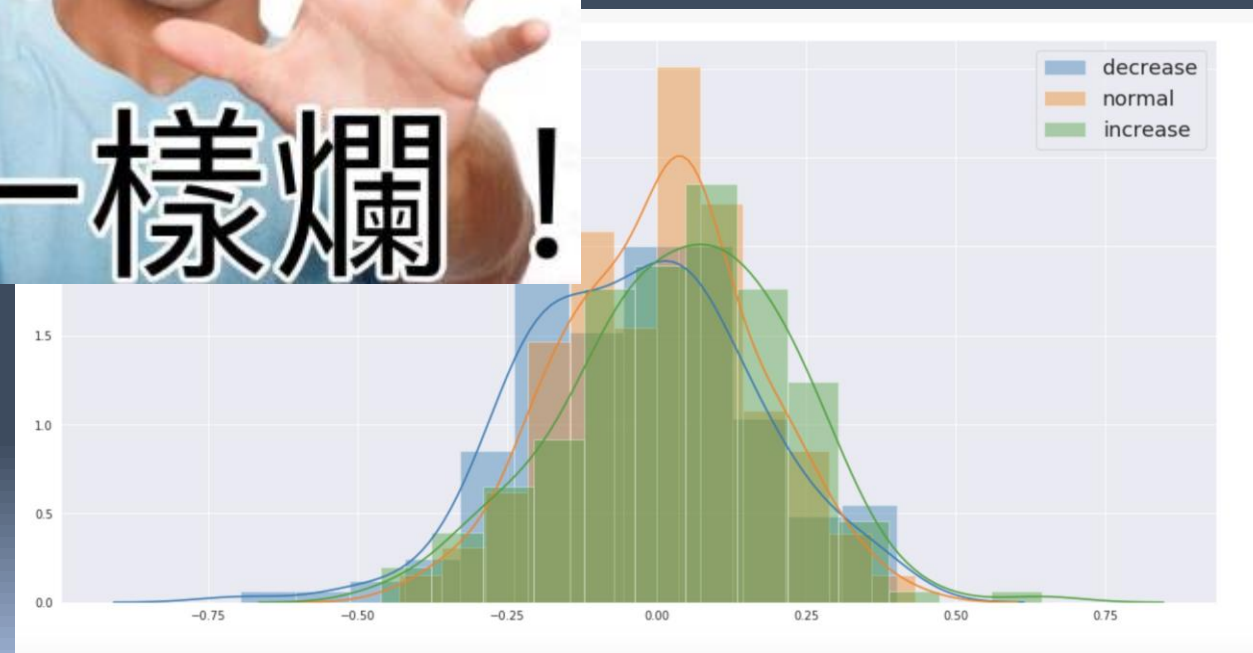
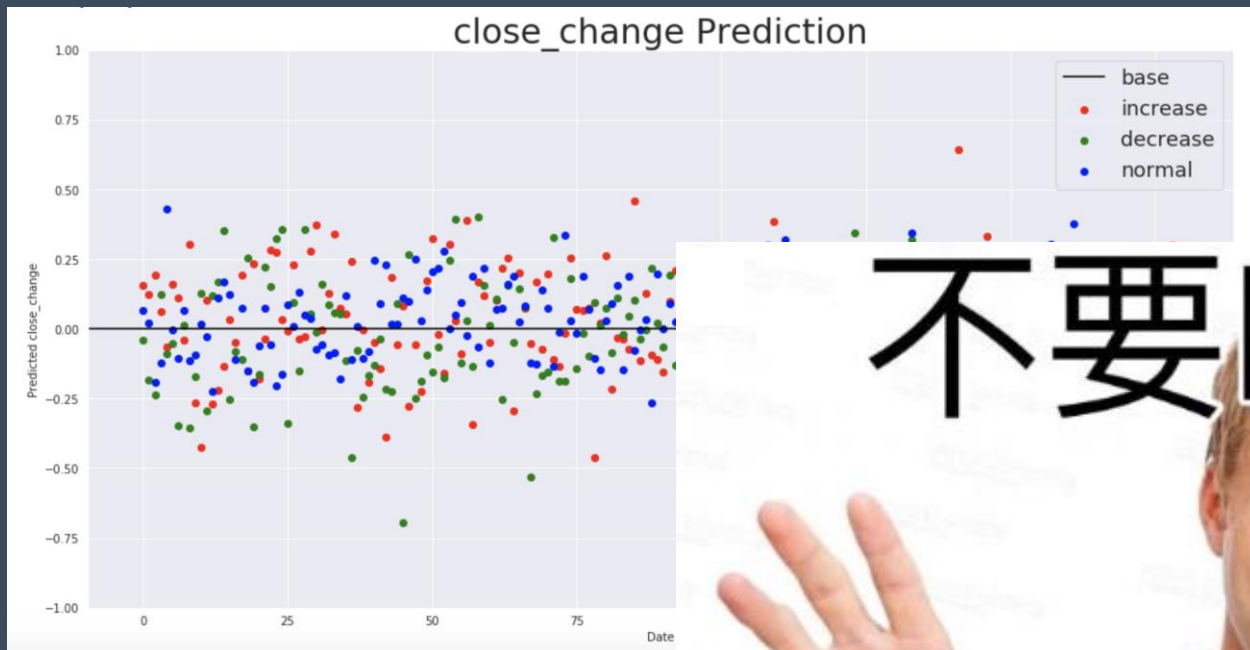


Autoencoder

隔日異常漲跌檢測

Autoencoder





問題 解決方法

無法幫助我們辨認趨勢

把指數改成類別

Res Family

ResNet

Dense
Self-
attention

Conv1D
Self-
attention

**Loss
function**

Multi-Loss

Focal-Loss

Accuracy

0.4001

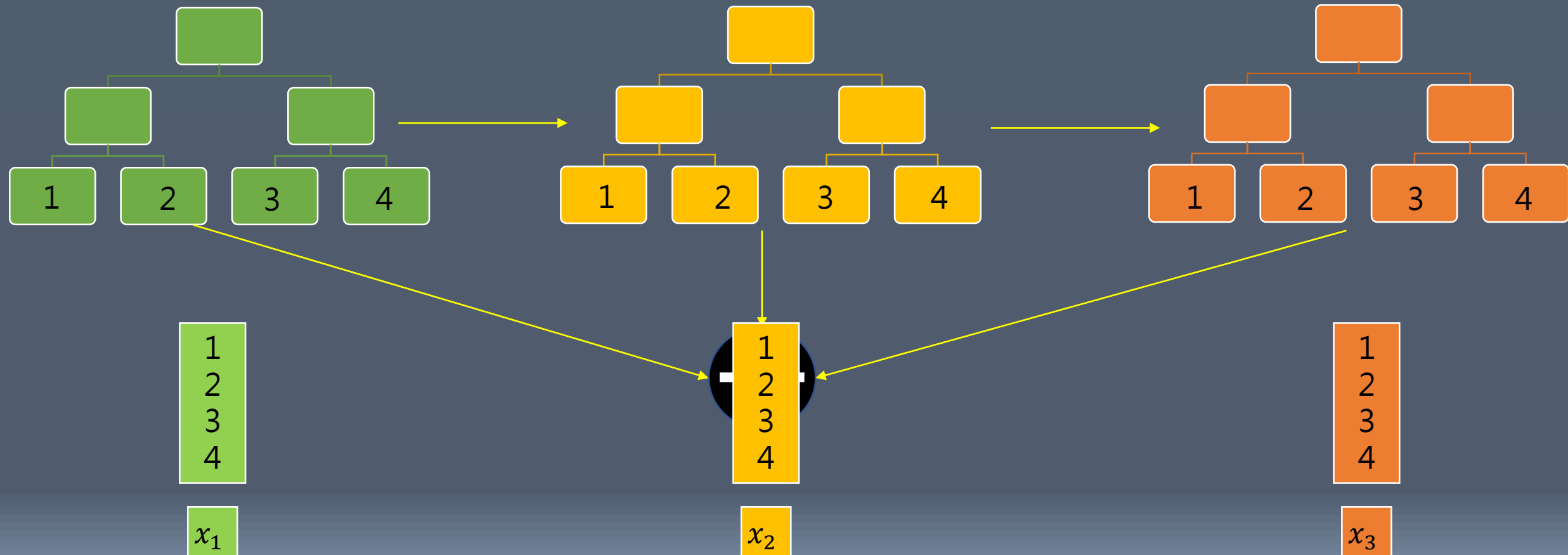
0.4471

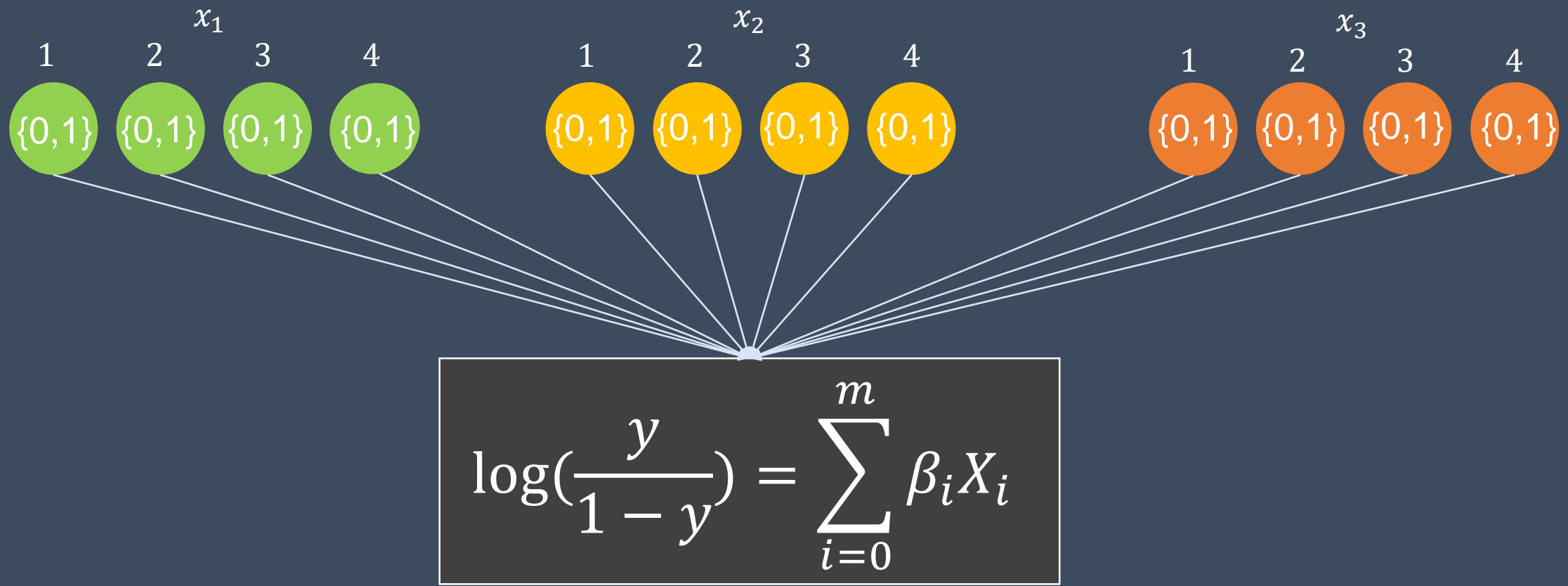
Loss

0.0911

0.0861

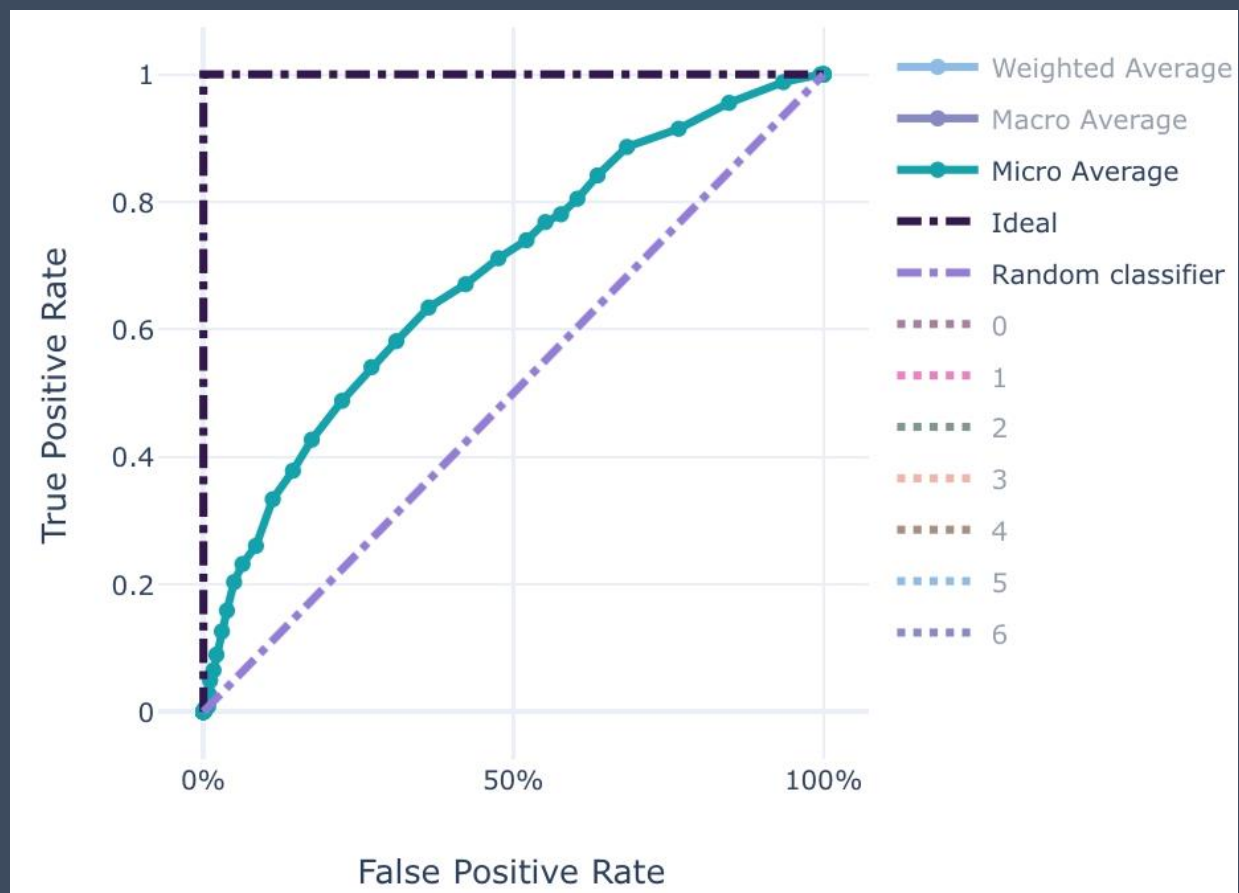
eXtreme Gradient Boosting



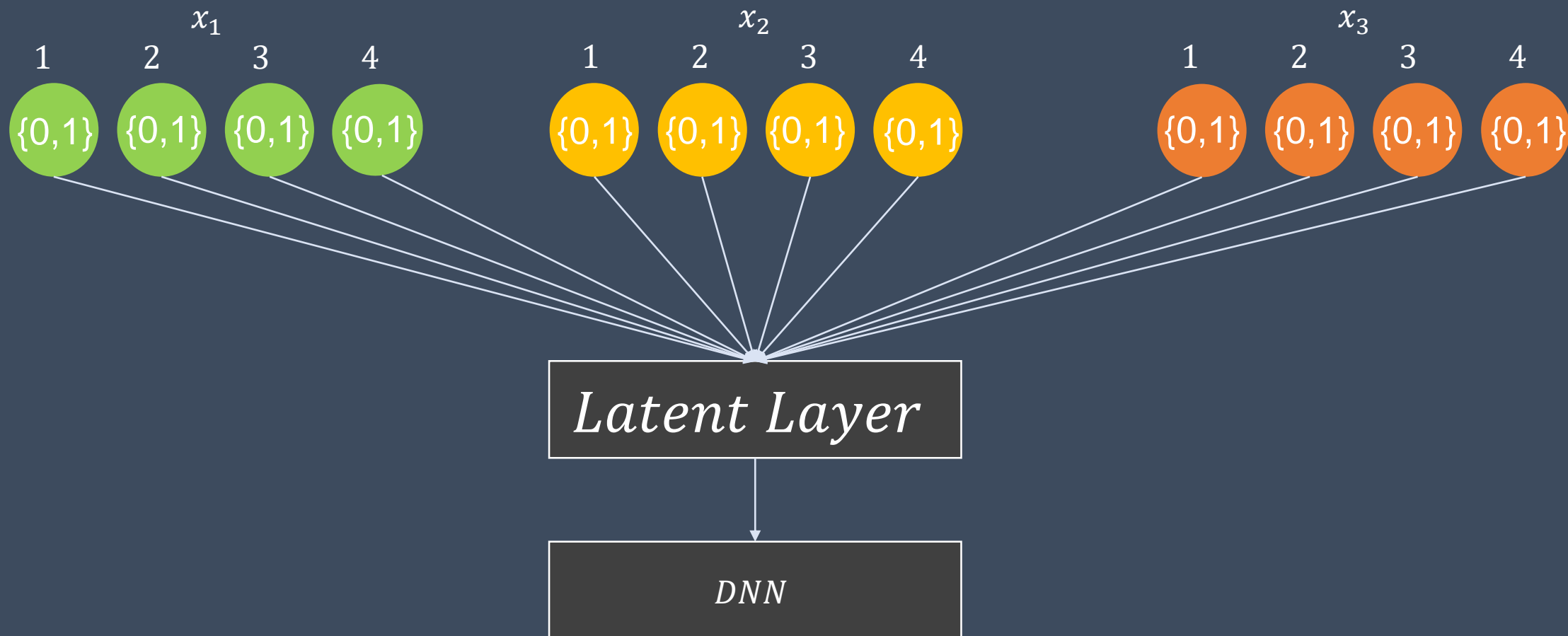


Ordinal logistic regression

	大跌	中跌	小跌	平盤	小漲	中漲	大漲
pred	0.0	1.0	2.0	3.0	4.0	5.0	6.0
label							
0.0	36	8	0	1	0	5	4
1.0	33	28	2	2	2	9	1
2.0	14	19	1	0	3	5	0
3.0	9	11	0	3	2	4	1
4.0	14	10	2	4	0	12	4
5.0	19	14	0	2	2	27	8
6.0	8	0	0	0	0	14	28
LR AUC: 0.68874							



實驗方向



洪子惟

報告人：



第三步 當日買賣點

Reinforcement Learning



以資料量選擇適合的演算法

Double Deep Q-learning Network

利用強化學習來判斷當日當沖買賣點



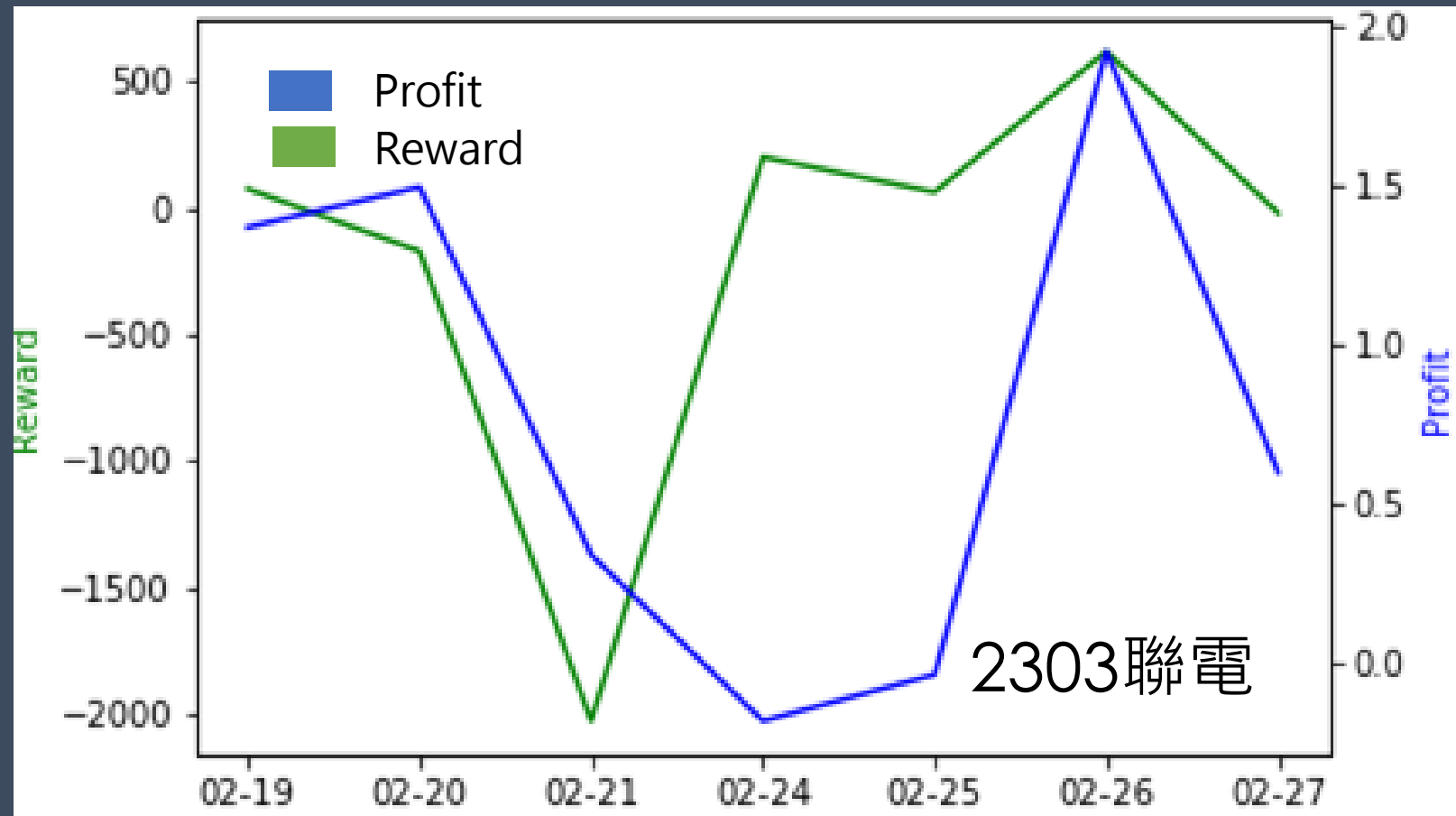
Double DQN (Training)



Double DQN (Training)

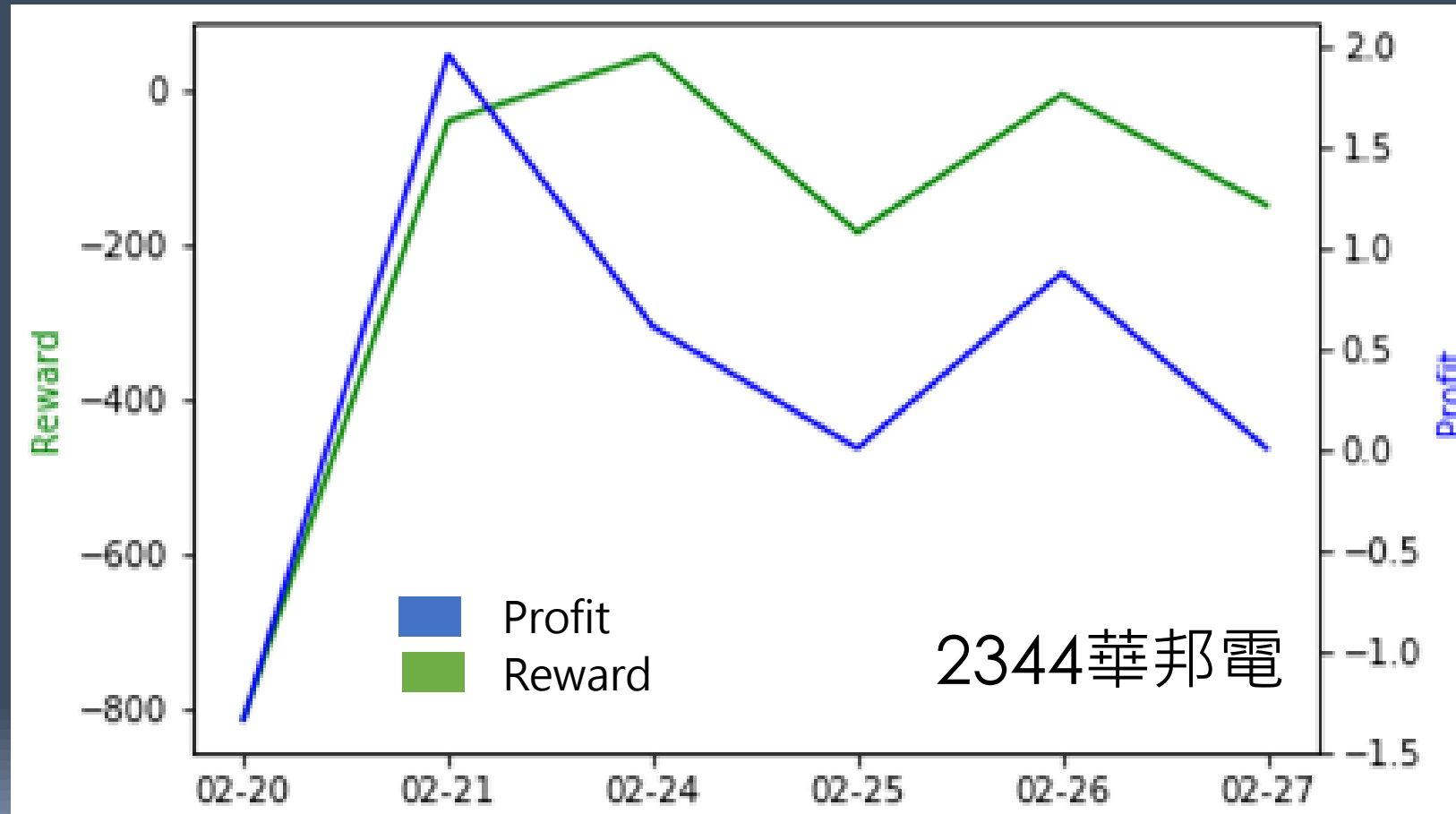


Double DQN (Testing)



股價-5%

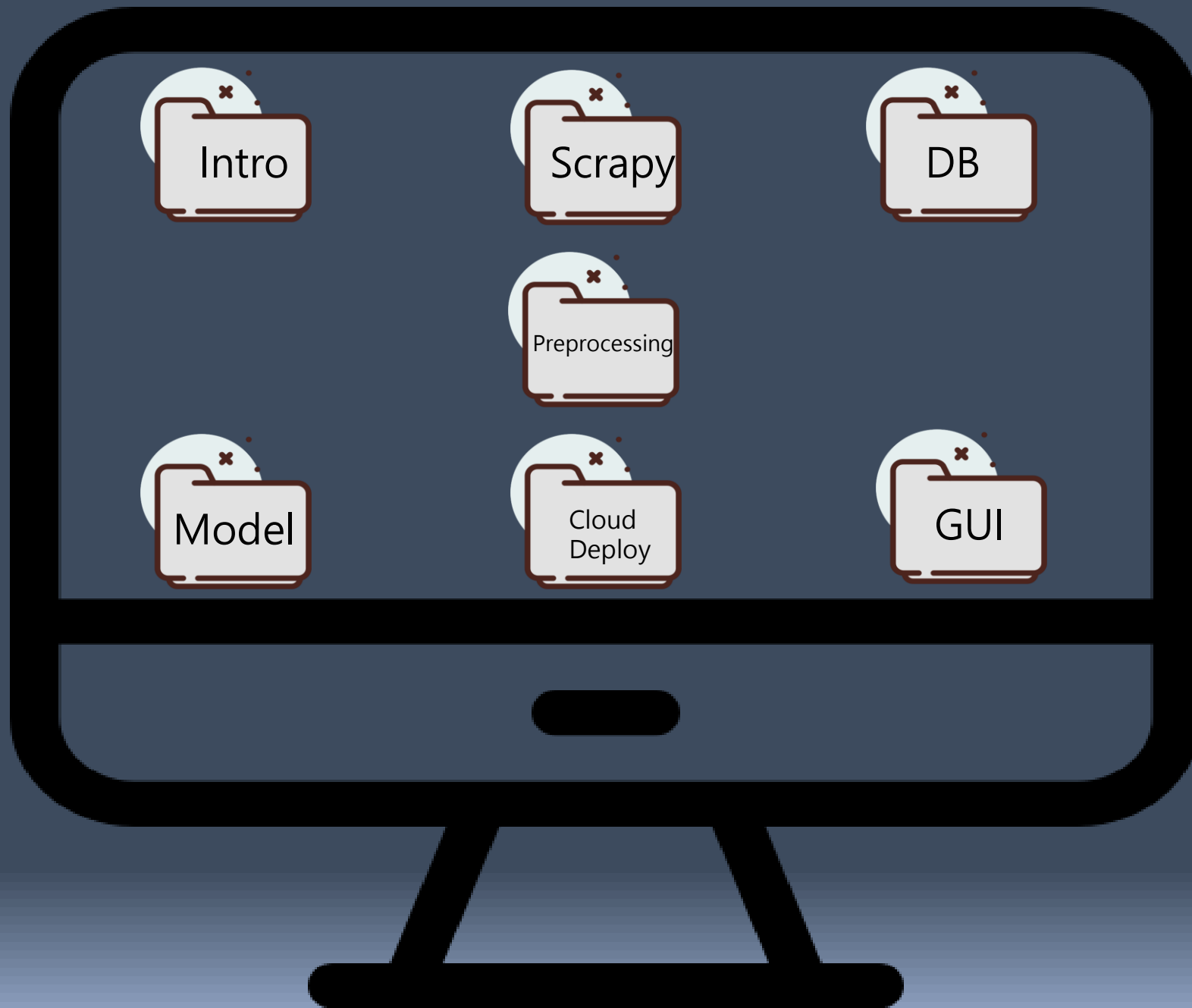
Double DQN (Testing)



股價-7%

問題

1. 模擬交易跟實際下單有差異



Intro

Scrapy

DB

Preprocessing

Model

Cloud
Deploy

GUI

雲端架構



由本地端爬蟲改為雲端爬蟲
並採用Kubernetes技術

Model部屬至雲端

splitter-k8s - 深入解析

Kubernetes 服務

搜尋 (Cmd +/)

- 概觀
- 活動記錄
- 存取控制 (IAM)
- 標籤
- 診斷並解決問題

設定

- 節點集區
- 升級
- 縮放
- 網路
- Dev Spaces
- 部署中心 (預覽)
- 原則 (預覽)
- 屬性
- 鎖定
- 匯出範本

監視

重新整理

檢視所有叢集

View Workbooks

說明

意見反應

時間範圍 = Last 6 hours

加入篩選器

- 新增功能
- 叢集
- 健全狀況 (預覽)
- 節點
- 控制器
- 容器
- 部署 (預覽)

依名稱搜尋...

計量: CPU 使用量 (millicores)

最小值 平均 第 50 第 90 個 第 95 最大值

1 個項目

名稱	狀態	第 ...	第 95	容器	運作時間	控制器	趨勢 第 95 百分位數 (1 列 = 15 分鐘)
aio-scrap...	✓	38%	762 mc	1	1 小時	aio-scrapy-...	
aio-s...	✓	38%	762 mc	1	1 小時	aio-scrapy-...	
scrapy-7...	✓	29%	585 mc	1	1 小時	scrapy-789...	

Pod 名稱: aio-scrapy-8668b855b8-j9q26 (aio-scrapy)

- 記錄
- 事件
- (沒有新資料)

搜尋...

0/0

移動 暫停 清除

2020-03-10T02:21:51.804261921Z retry request, times: 1 total_error: 1 tse_2344 https://mis.twse.com.tw/stock/api/getStockInfo.jsp?ex_ch=tse_2344.tw&json=1&delay=0

2020-03-10T02:21:51.804266821Z retry request, times: 1 total_error: 1 tse_2363 https://mis.twse.com.tw/stock/api/getStockInfo.jsp?ex_ch=tse_2363.tw&json=1&delay=0

2020-03-10T02:21:51.804271721Z retry request, times: 1 total_error: 1 tse_2449 https://mis.twse.com.tw/stock/api/getStockInfo.jsp?ex_ch=tse_2449.tw&json=1&delay=0

2020-03-10T02:21:51.804276421Z Added 26 records to db session, committed

2020-03-10T02:21:51.804280921Z retry request, times: 2 total_error: 1 tse_4968 https://tw.quote.finance.yahoo.net/quote/q?type=tick&perd=5s&mkt=10&sym=4968

2020-03-10T02:21:51.804286621Z retry request, times: 2 total_error: 1 tse_2449 https://tw.quote.finance.yahoo.net/quote/q?

aio-scrapy 容器

檢視即時資料 (預覽)

在分析中檢視

容器名稱

aio-scrapy

容器識別碼

ba8b5418f3887dfec00e667b2b37ae8eb65de53f30148ec18b50cf3f073c94d9

容器狀態

running

Container Status Reason

-

映像

aio_scrapy

影像標籤

v1

▲ aio-scrap... 38% 762 mc 1 1 小時 aio-scrapy...

證交所爬蟲

▲ aio-s... 38% 762 mc 1 1 小時 aio-scrapy...

▷ scrapy-7... 29% 585 mc 1 1 小時 scrapy-789...

新聞爬蟲

名稱: aio scrapy 3668b855b8-j9q26 (aio-scrapy)

事件

0/0 X ^ v 捲動

-10T02:21:51.804261921Z retry request, times: 1 total_error: 1 tse_2344 https://mis.twse.com.tw/stock/api/g
se_2344.tw&json=1&delay=0

-10T02:21:51.804266821Z retry request, times: 1 total_error: 1 tse_2363 https://mis.twse.com.tw/stock/api/g
se_2363.tw&json=1&delay=0

-10T02:21:51.804271721Z retry request, times: 1 total_error: 1 tse_2449 https://mis.twse.com.tw/stock/api/g
se_2449.tw&json=1&delay=0

-10T02:21:51.804276421Z Added 26 records to db session, committed

-10T02:21:51.804280921Z retry request, times: 2 total_error: 1 tse_4968 https://tw.quote.finance.yahoo.net/q
k&perd=5s&mkt=10&sym=4968

-10T02:21:51.804286621Z retry request, times: 2 total_error: 1 tse_2449 https://tw.quote.finance.yahoo.net/q

雲端部屬



Google Cloud

將資料庫部屬至雲端



設定

vCPU

1

記憶體

1.7 GB

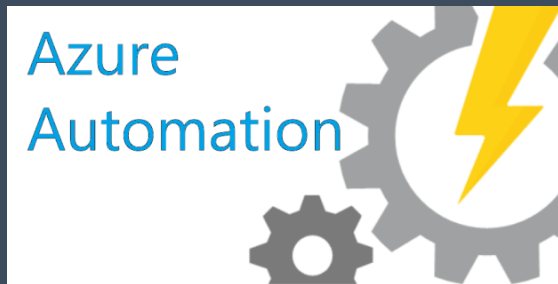
SSD的儲存空間

10 GB



資料庫版本是 PostgreSQL 11

節省約80%成本
\$2800/月 → \$470/月



Azure Runbook
Python Script

8:50
自動開機

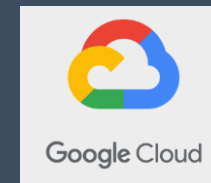


13:40
自動關機

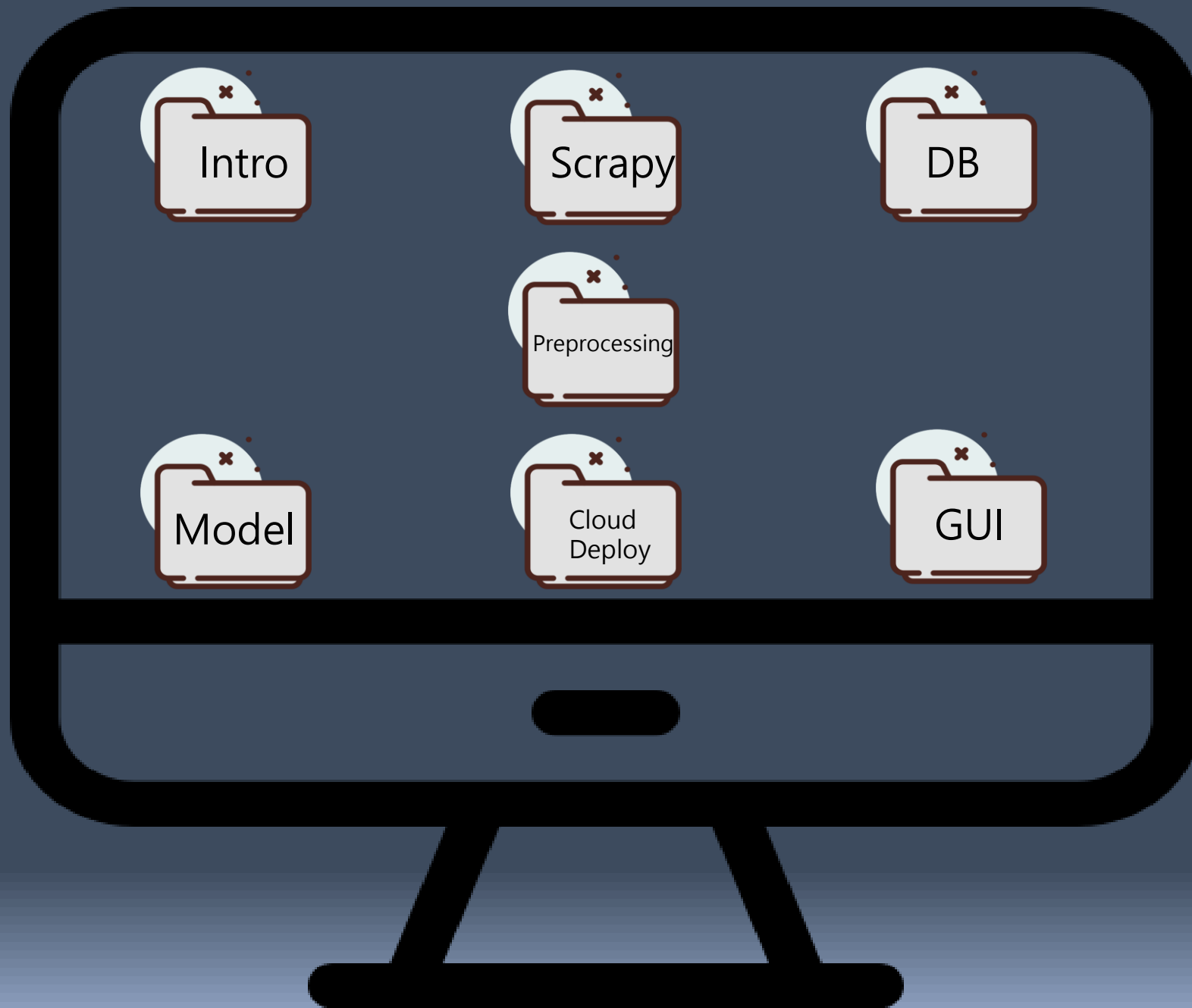


k8s爬蟲

5秒存入資料



PostgreSQL



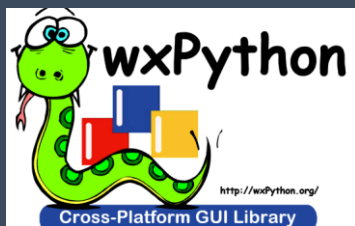
李家豪

報告人:



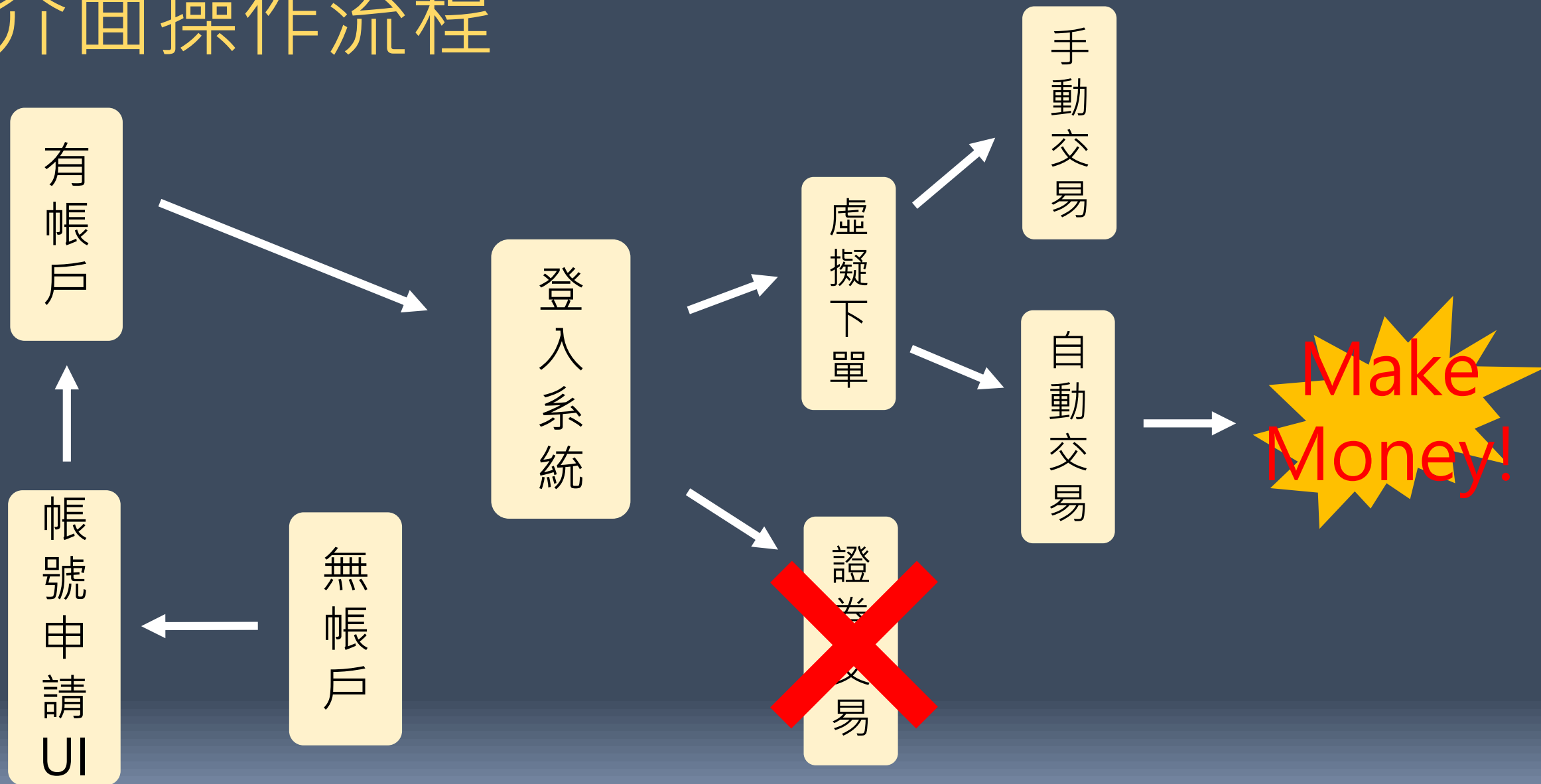
呈現樣式

What



為何選擇  ?

介面操作流程



問題 解決方法

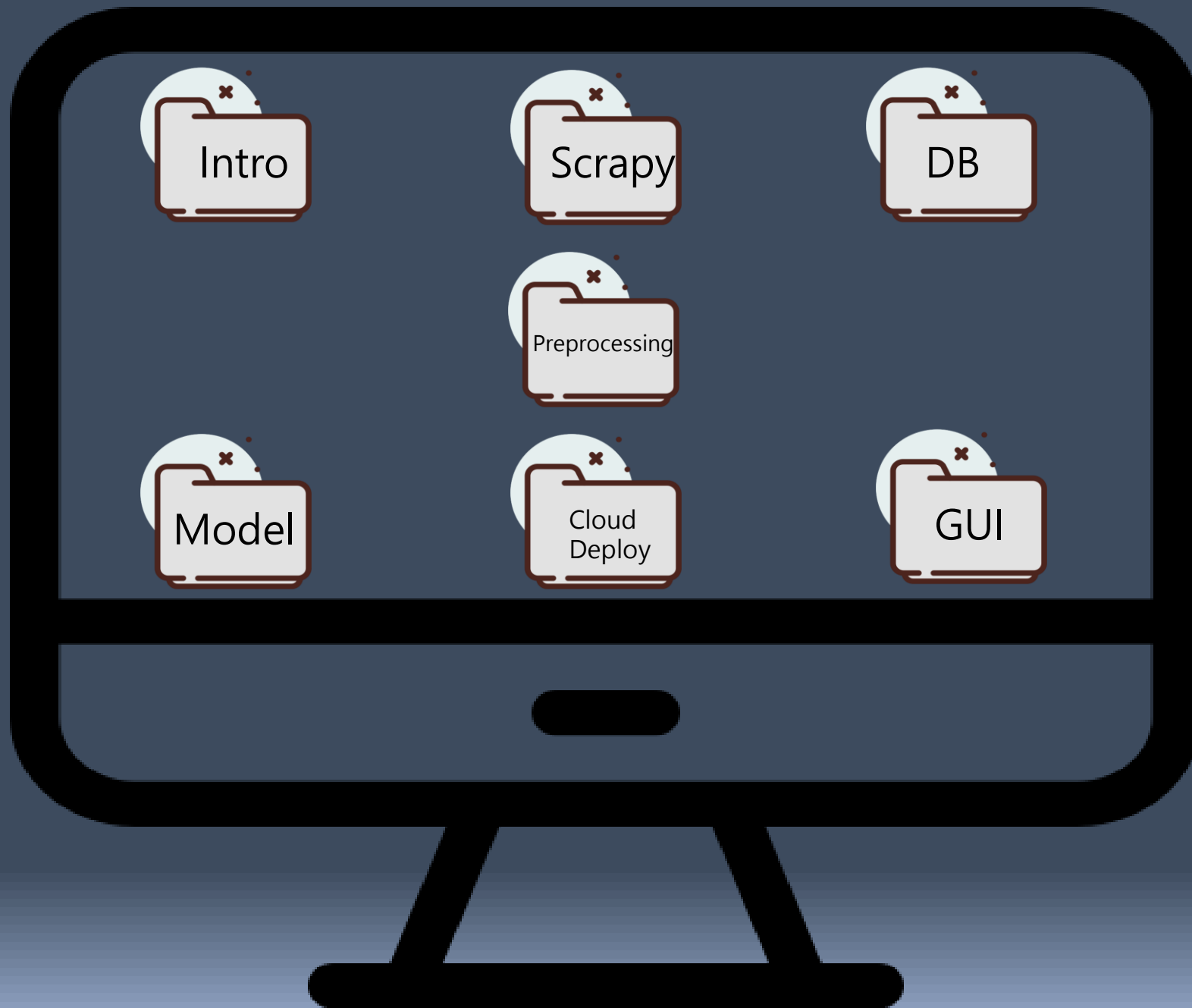
Selenium串接導致程式過慢反應

利用requests功能，以達到介面及資料同步

程式是單執行緒

建立多執行緒

預測結果呈現



Intro

Scrapy

DB

Preprocessing

Model

Cloud
Deploy

GUI

感謝聆聽