



# 個人投資風險管家

台大經濟四-郭奕廷。東吳巨資二-賴威博。台大財金四-李佳盈。  
台大生機四-詹佳蓓。台大資管二-歐崇愷  
指導教授-張智星。指導業師-詹益安

A man with a beard and short brown hair is looking down at a white smartphone in his right hand. He is wearing a light blue button-down shirt with rolled-up sleeves and a dark backpack strap is visible over his shoulder. The background is a blurred city street at night with warm, out-of-focus lights.

# 目錄

1

專案介紹

2

資料分析

3

機器學習模型

4

系統呈現

5

附錄

# 1



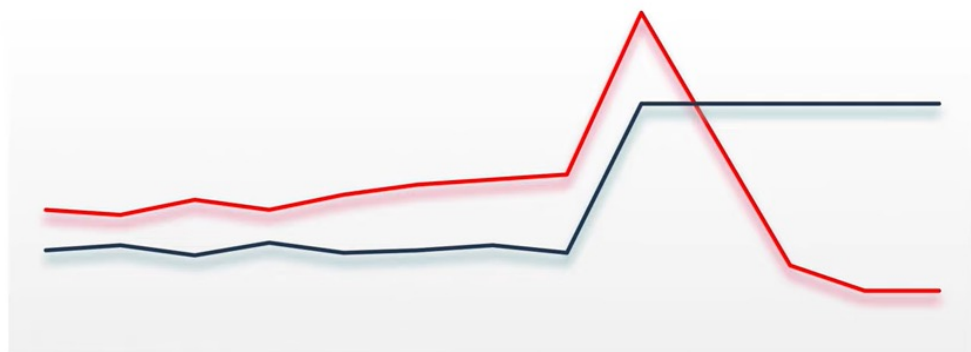
嗨！我是您的  
投資風險管家

## 專案介紹

# 業務痛點及專案目標

保護新手投資人：找出異常的投資行為

- 避免投資人跟著市場風向，導致過度投資
- 避免投資人違約交割



— 投資成本 — 投資部位市值



## 股市小白愈來愈多 黃天牧對年輕人示警

陳林幸虹採訪

2021年1月12日 · 2 分鐘 (閱讀時間)



金管會主委黃天牧。(資料照片/陳林幸虹 攝)

根據台灣證券交易所的統計，去年2020年在疫情期間，台灣20至30歲年輕族群在台北股市的開戶人數占比愈來愈高，已大幅成長至123萬人。股市小白愈來愈多，金管會主委黃天牧今天(12日)在立法院財政委員會備詢時也特別對年輕人示警，呼籲年輕人剛進入股市，必須先加強自我的金融知識和教育，金管會週邊單位也都有推出相關課程，希望年輕人多多參與。



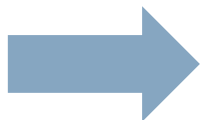


# 投資模式偵測系統

建立偵測投資行為改變的AI，輔助投資人意識到自己的投資模式改變



投資人投資紀錄



投資模式偵測系統



通知投專進行關懷



對顧客溫馨提醒

2

資料分析

## 特徵工程

## 機器學習模型

## 實際應用

資料整合

標準化

特徵選擇

隨機下採樣

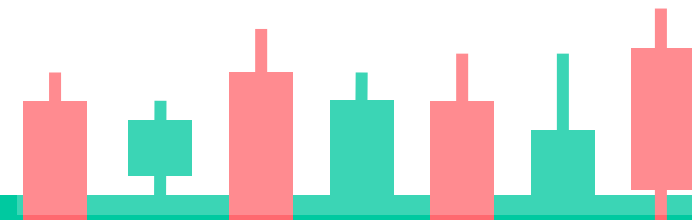
Random Forest

XGBOOST

LSTM

分類

機率預測



# 資料描述

交易紀錄

+

股票資訊

+

顧客資訊

**Train(2019)**

( 2604391, 44 )

**Test(2020-2021)**

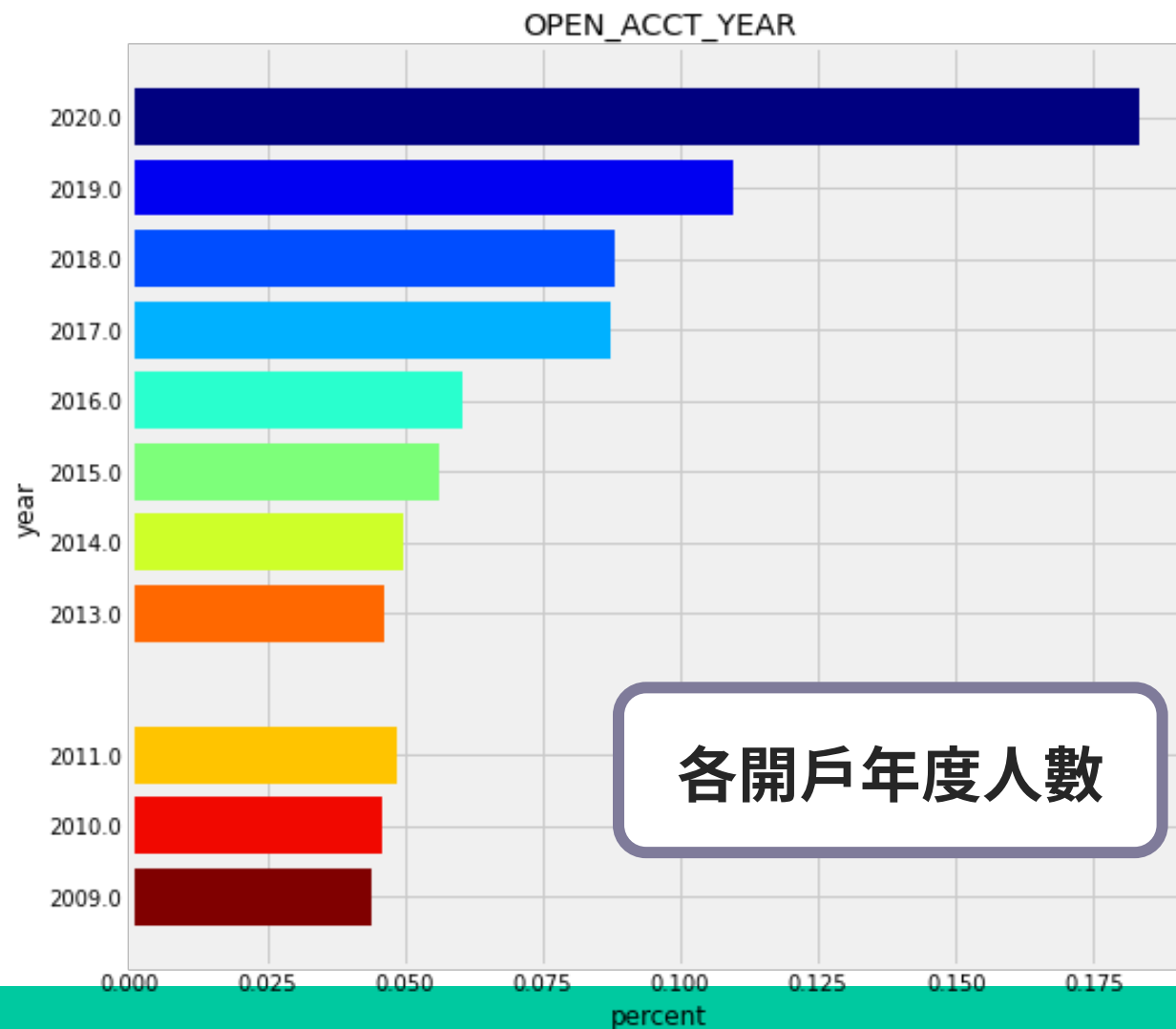
( 1439297, 44 )

**Train(2019)**

( 762723, 44 )

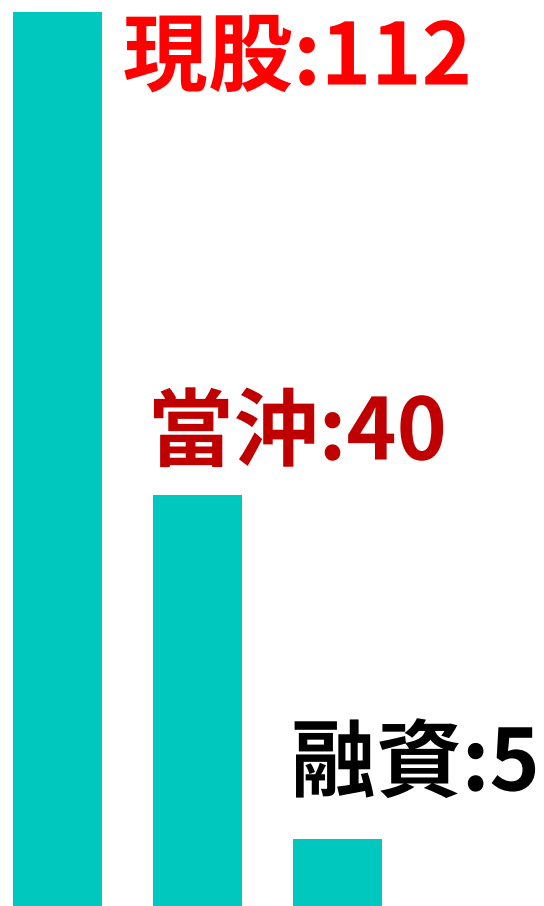
**Test(2020-2021)**

( 583693, 44 )

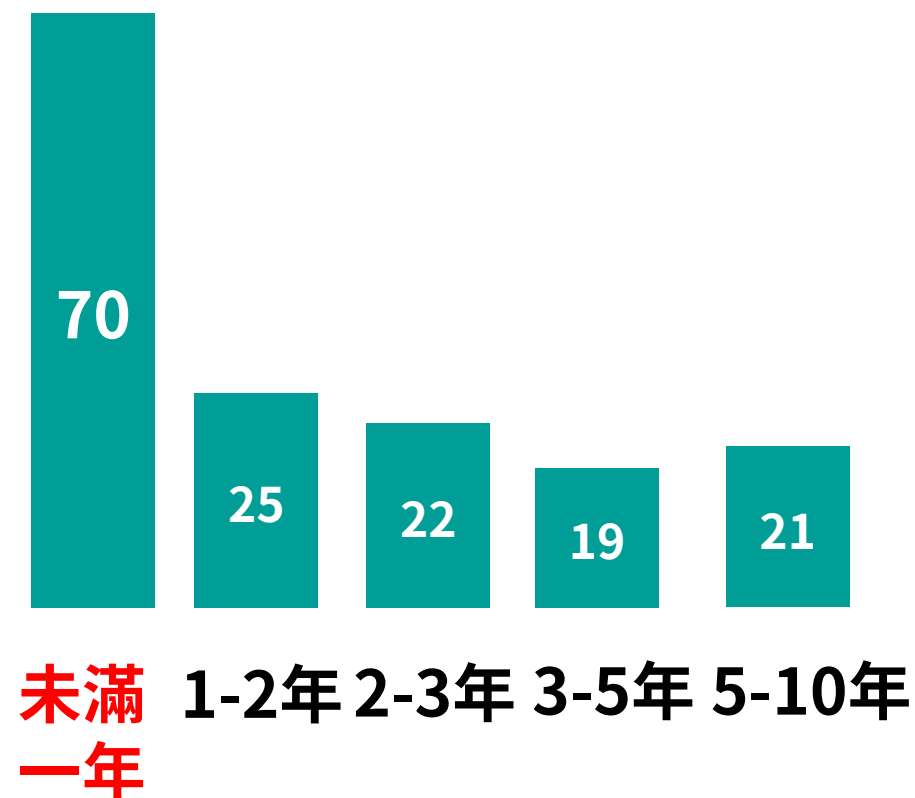




## 交易種類



## 股市交易經驗



## 隨機下採樣

# Under-Sampling

```
from imblearn.under_sampling import RandomUnderSampler
```

正常交割：違約交割



1016905 : 59

59000 : 59

Train data



583672 : 21

21000 : 21

Test data

## 特徵選擇

買賣價格的波動 (MA)

Beta值的波動 (MA)

電子類股、其他產業類股

Columns: 81 → 44

開戶年度 目標模型

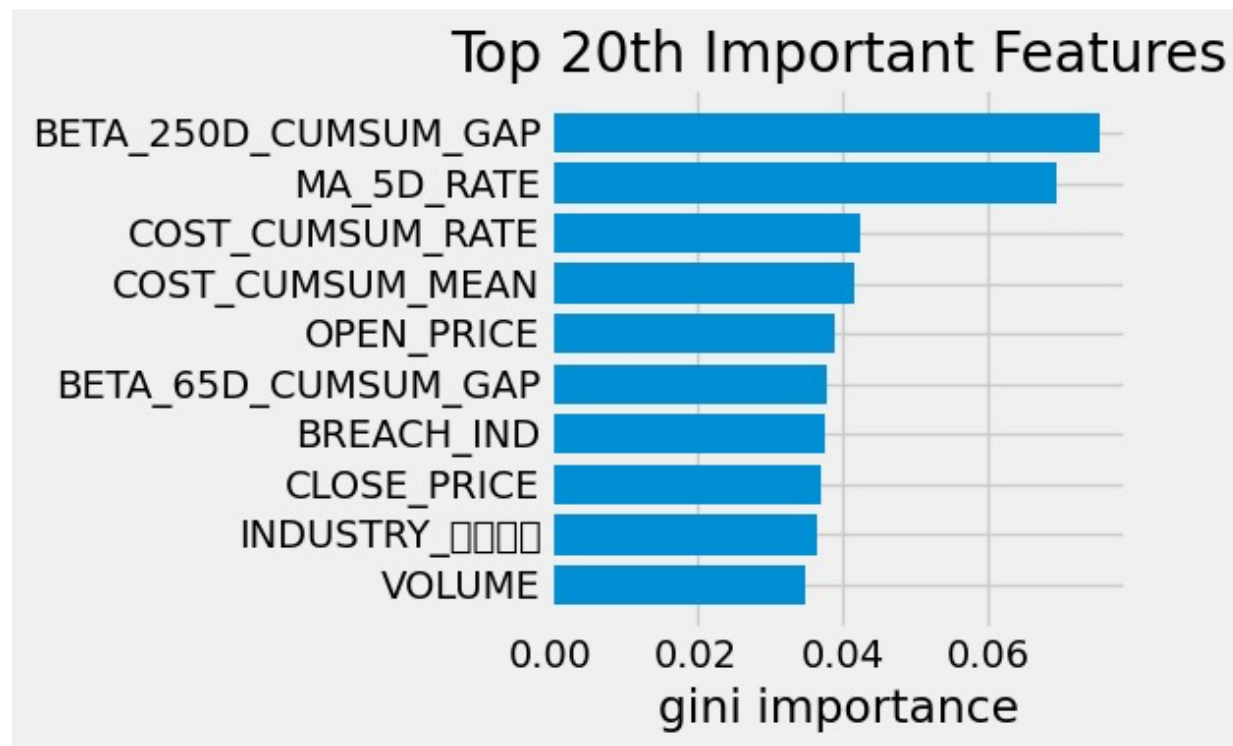
2019-2021年

2019年以前

Newbie Model

Old Model

Random Forest 找出的重要特徵



3

機器學習模型

# 機器學習模型

1

XGBOOST

2

LSTM

3

Random Forest



# XGBOOST

## 極限梯度提升

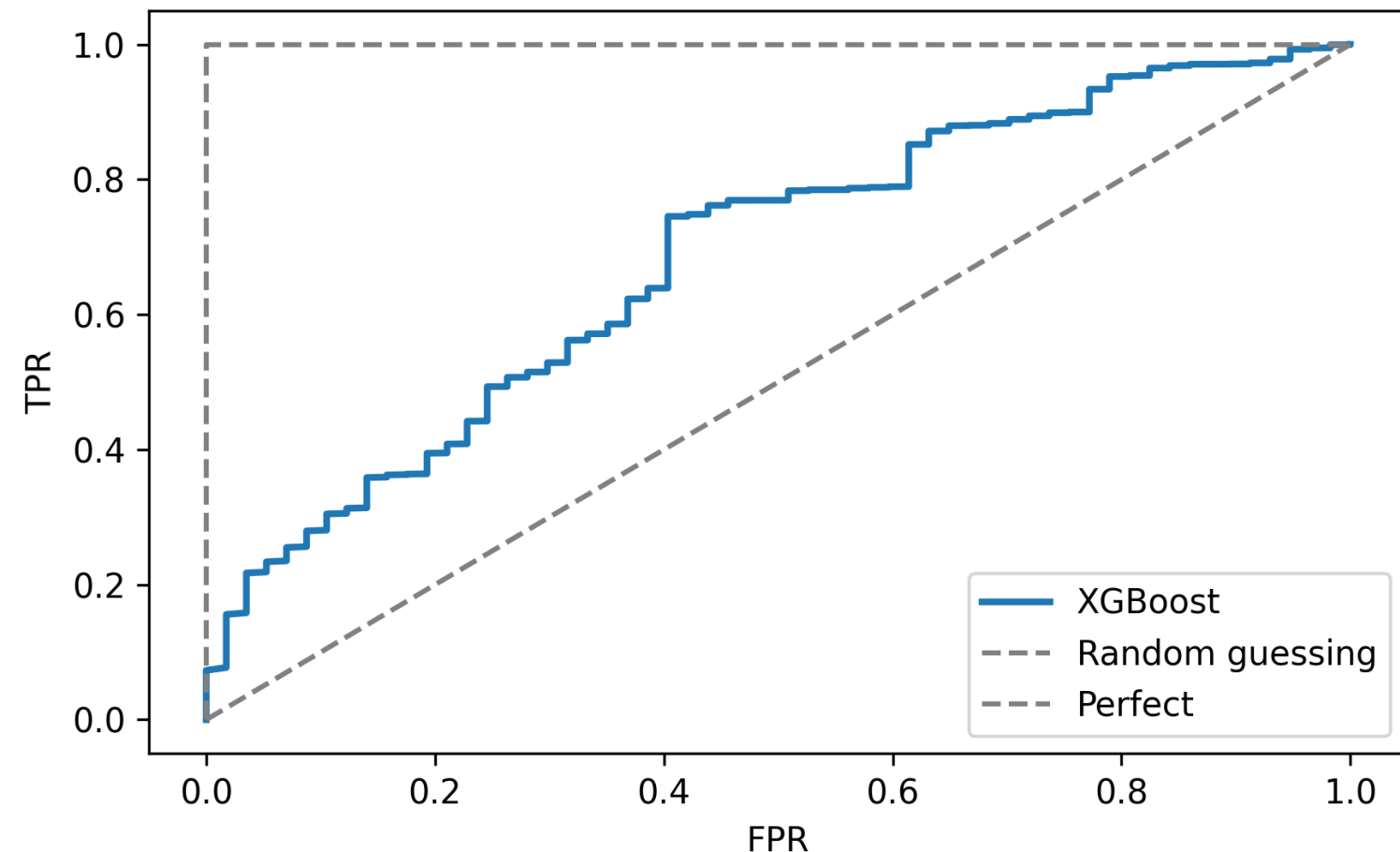
```
from sklearn.ensemble import RandomForestClassifier
```

### 特色

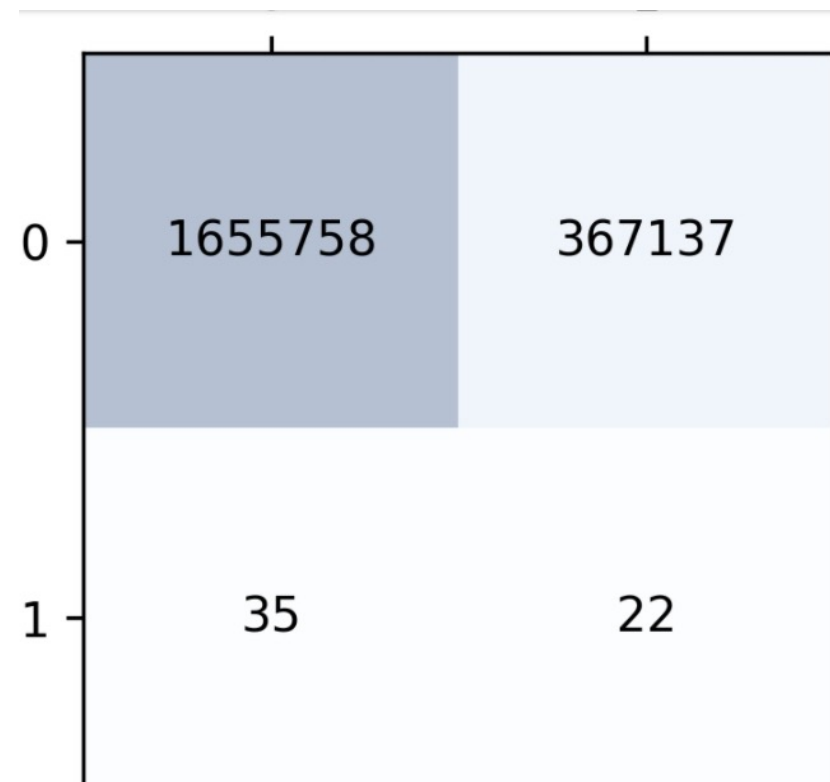
- 預防overfitting
- 可以調整樣本權重
- 平行處理

# 模型表現

ROC Curve



AUC = 0.814



Accuracy = 0.8185

F1\_score = 0.0001

Precision = 0.0005

Recall = 0.3860



# 機器學習模型

1

XGBOOST

2

LSTM

3

Random Forest



# LSTM

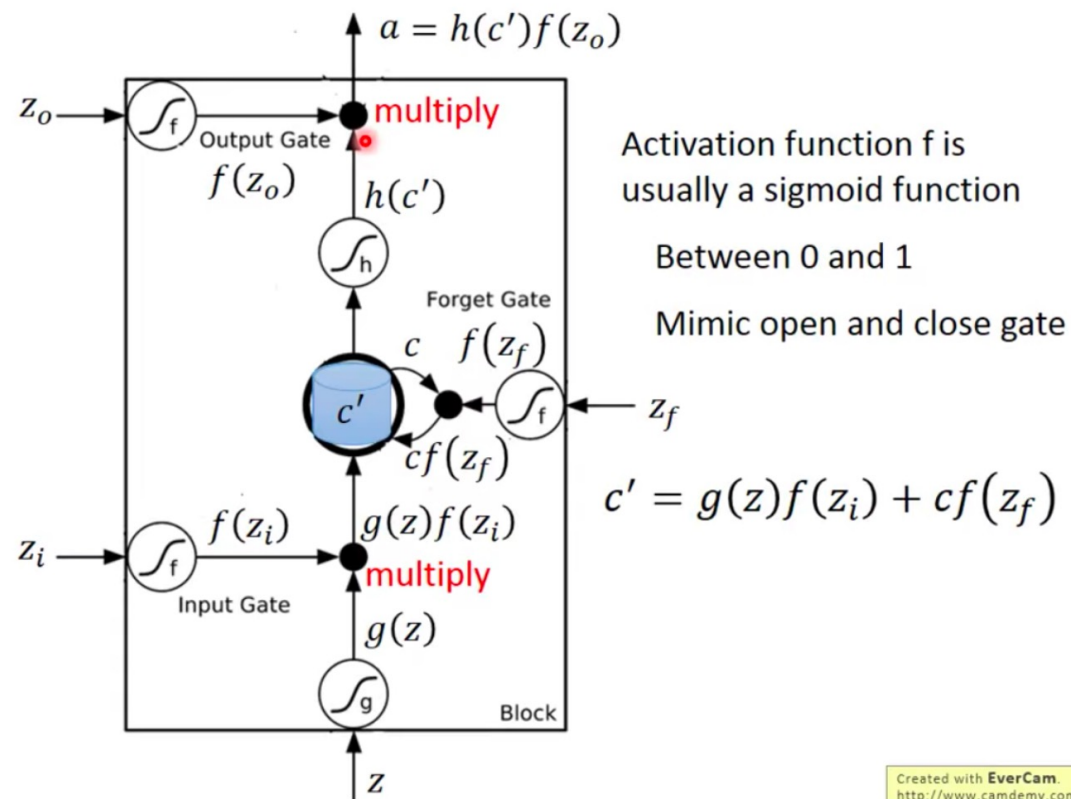
## 長短期記憶模型

前5日、前21日、前65日Beta值

前30天交易紀錄

預測

第31天發生違約交割機率



# LSTM

## 長短期記憶模型

模型  
架構

LSTM

LSTM

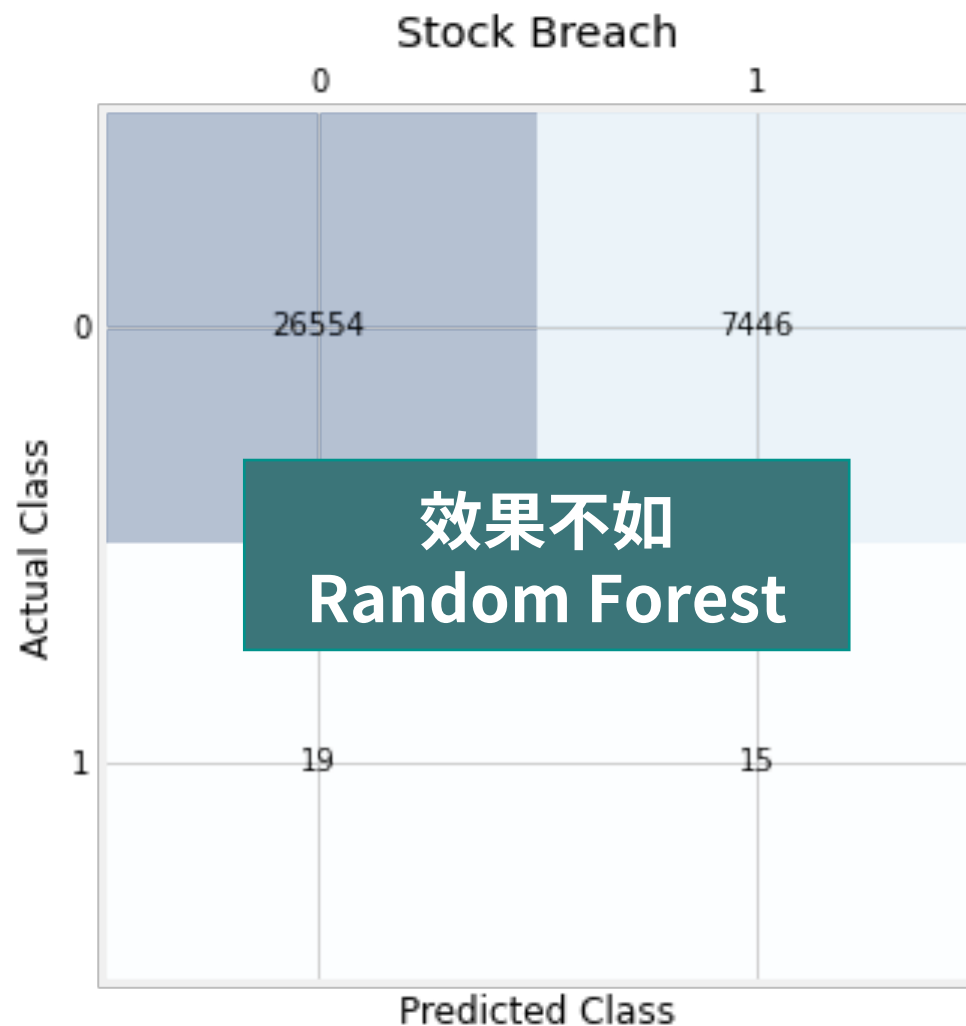
Dense

Dense

模型  
表現

Loss: Binary Cross-entropy

Optimizer: RMSProp



accuracy = 0.7806  
recall = 0.35

# 機器學習模型

1

XGBOOST

2

LSTM

3

Random Forest



# Random Forest

## 隨機森林演算法

```
from sklearn.ensemble import RandomForestClassifier
```

適合

- 不平衡資料 Imbalanced data
- 異常偵測 Outlier detection

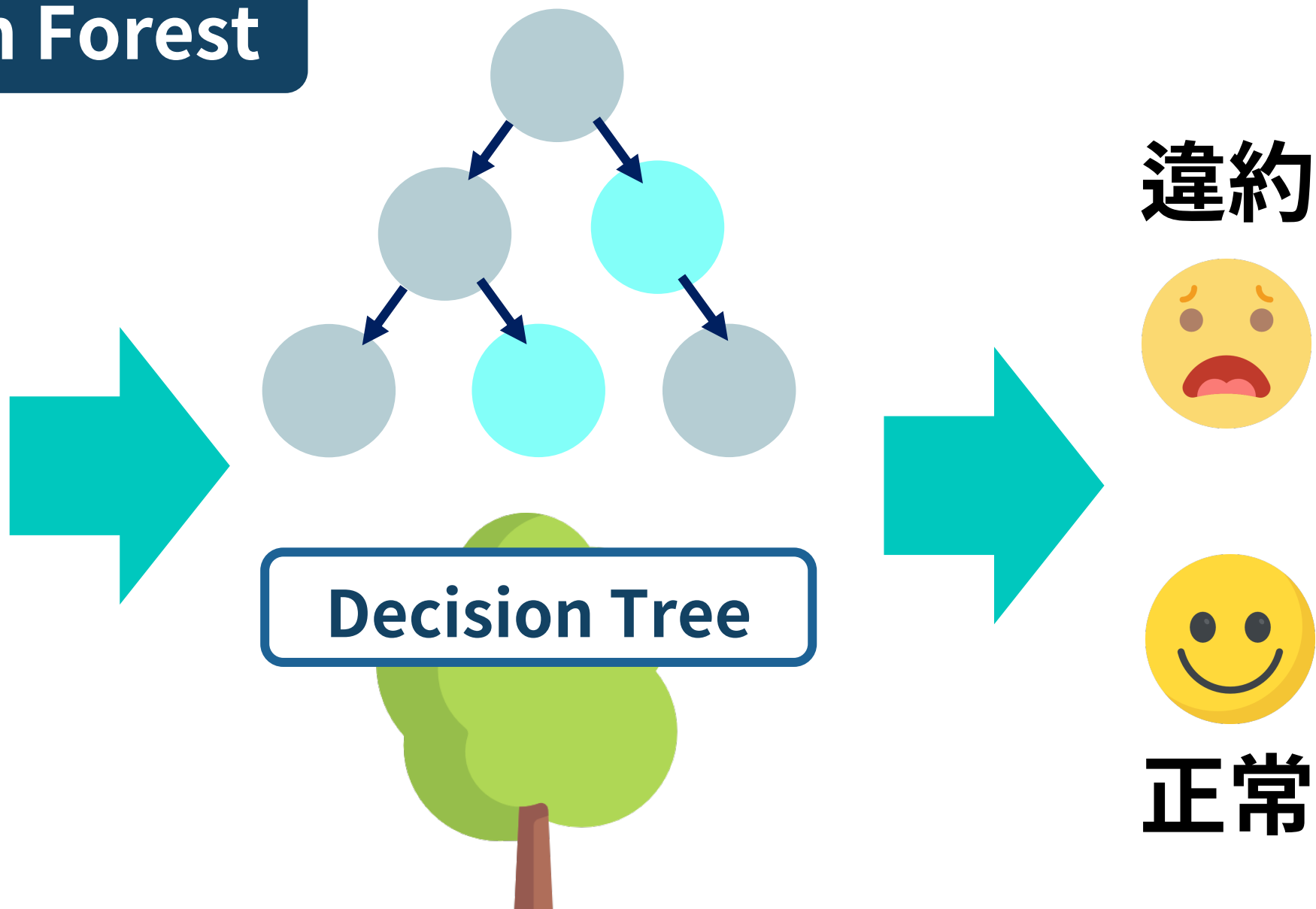
$$I_G(t) = \sum_{i=1}^c p(i|t)(1 - p(i|t)) = 1 - \sum_{i=1}^c p(i|t)^2$$



# Random Forest

## 資料

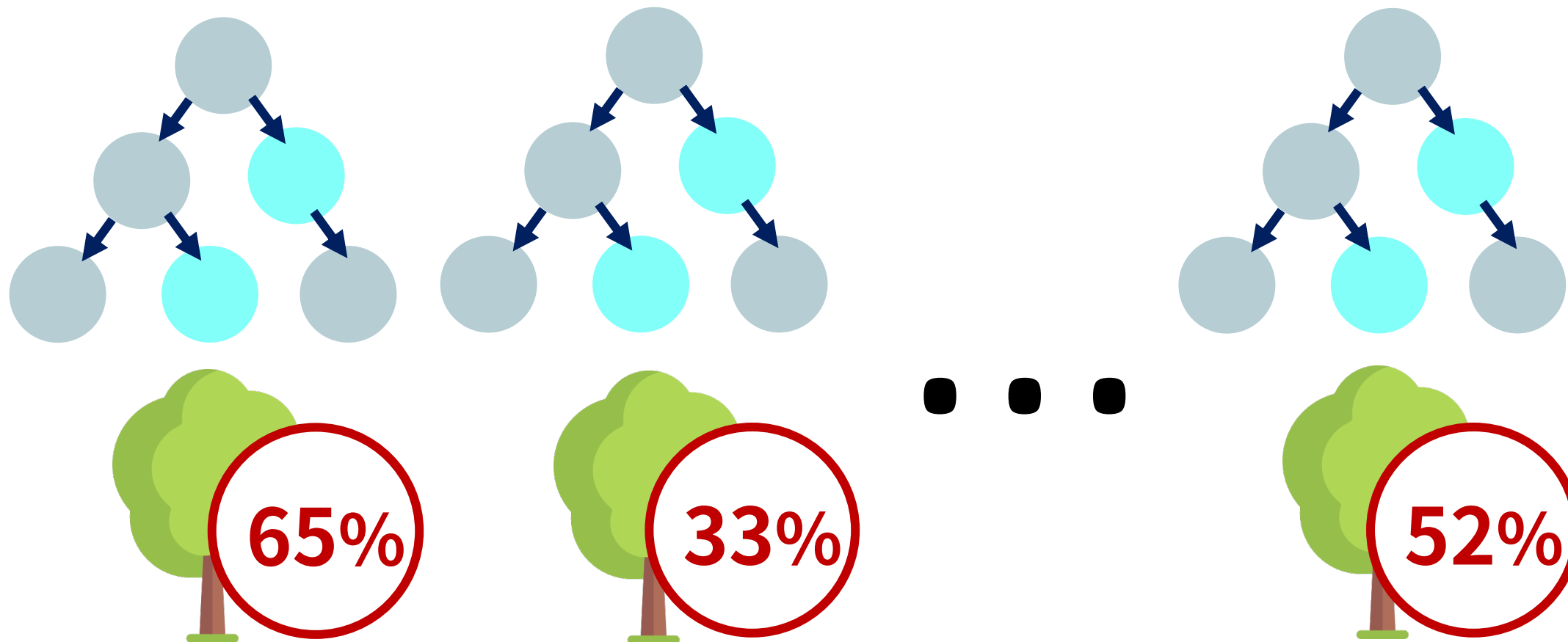
開戶年度  
交易股數  
交易價格  
Beta值  
產業別  
...



# Random Forest

## Ensemble Learning

## 防止過擬合





# Hyper-parameter 超參數

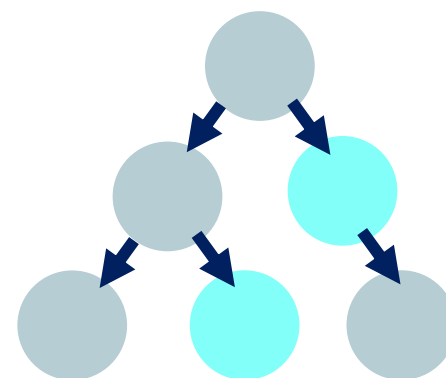
## Grid Search CV 調整參數

最佳參數組合：



= 105

n\_estimators



= 10

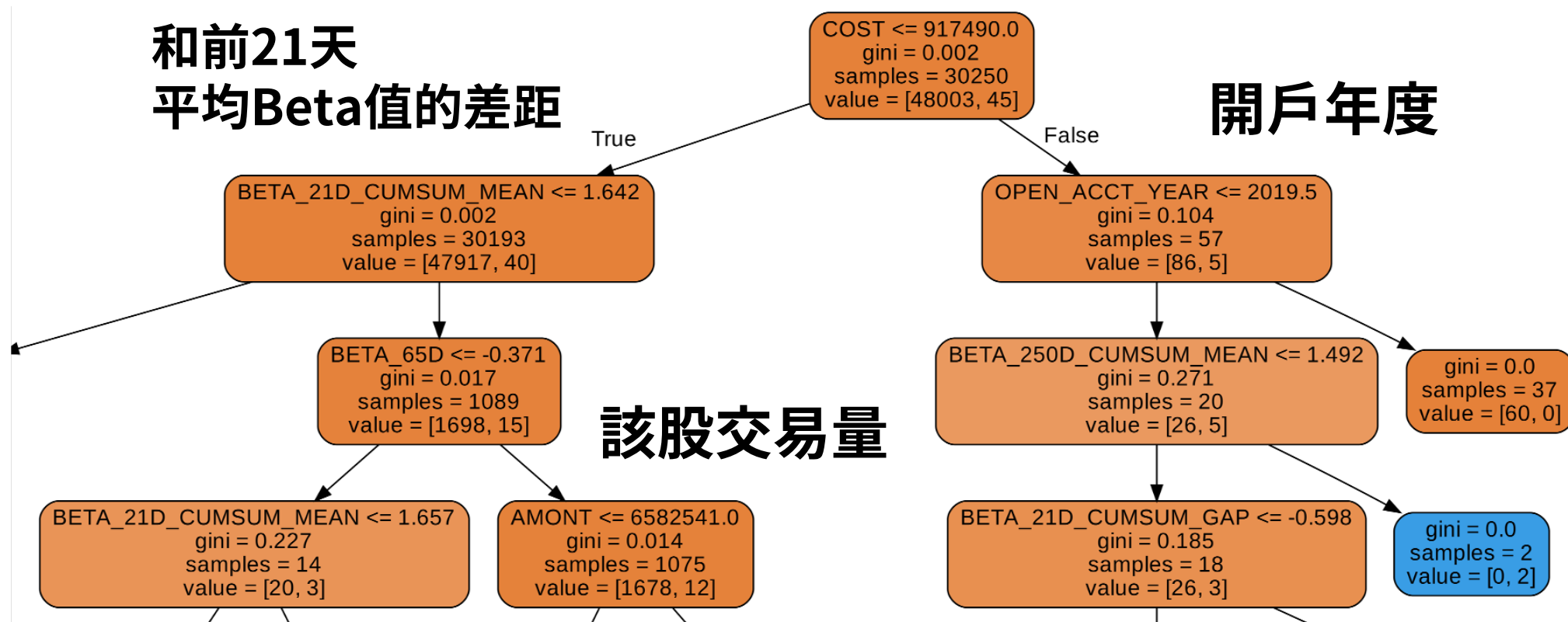
max\_depth

# 實際決策樹長相 (深度=10)

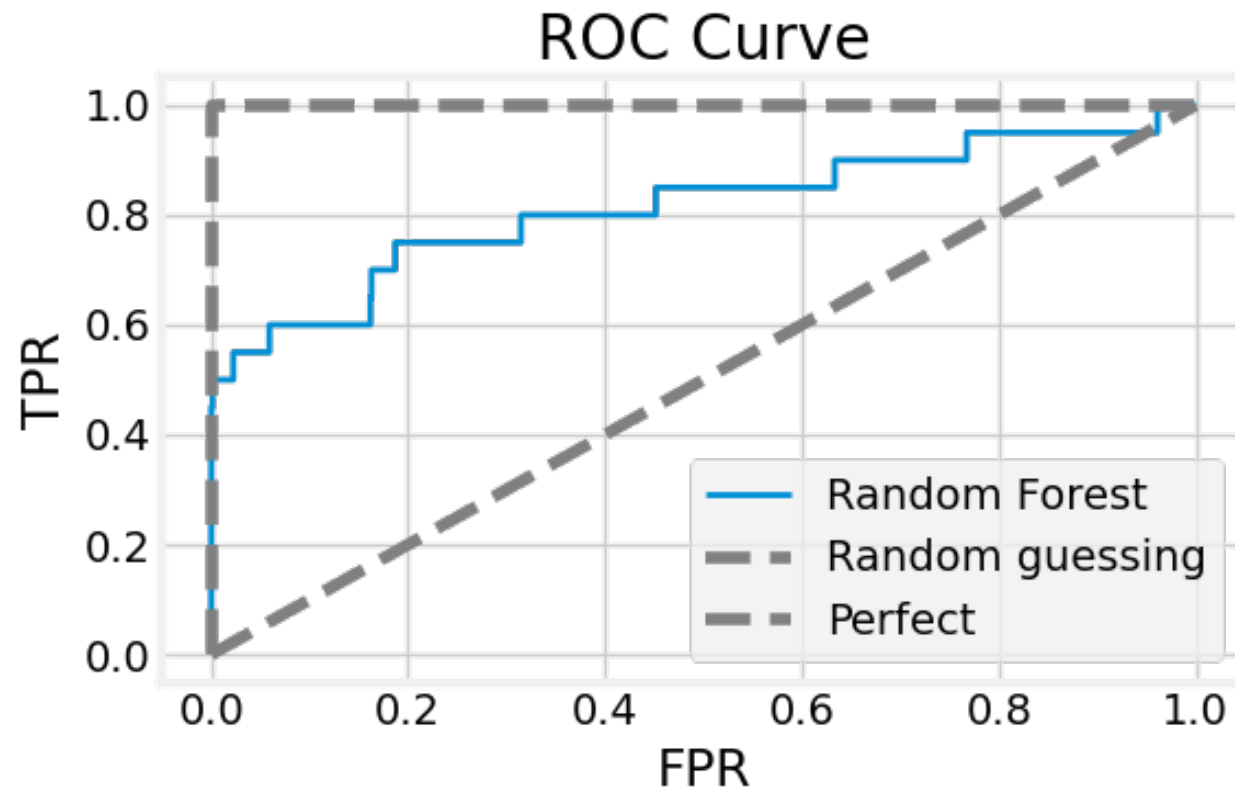
單筆交易金額 (價格\*股數)

和前21天  
平均Beta值的差距

開戶年度



# 模型表現



**AUC = 0.814**

## 模型預測表現

Accuracy = 99%

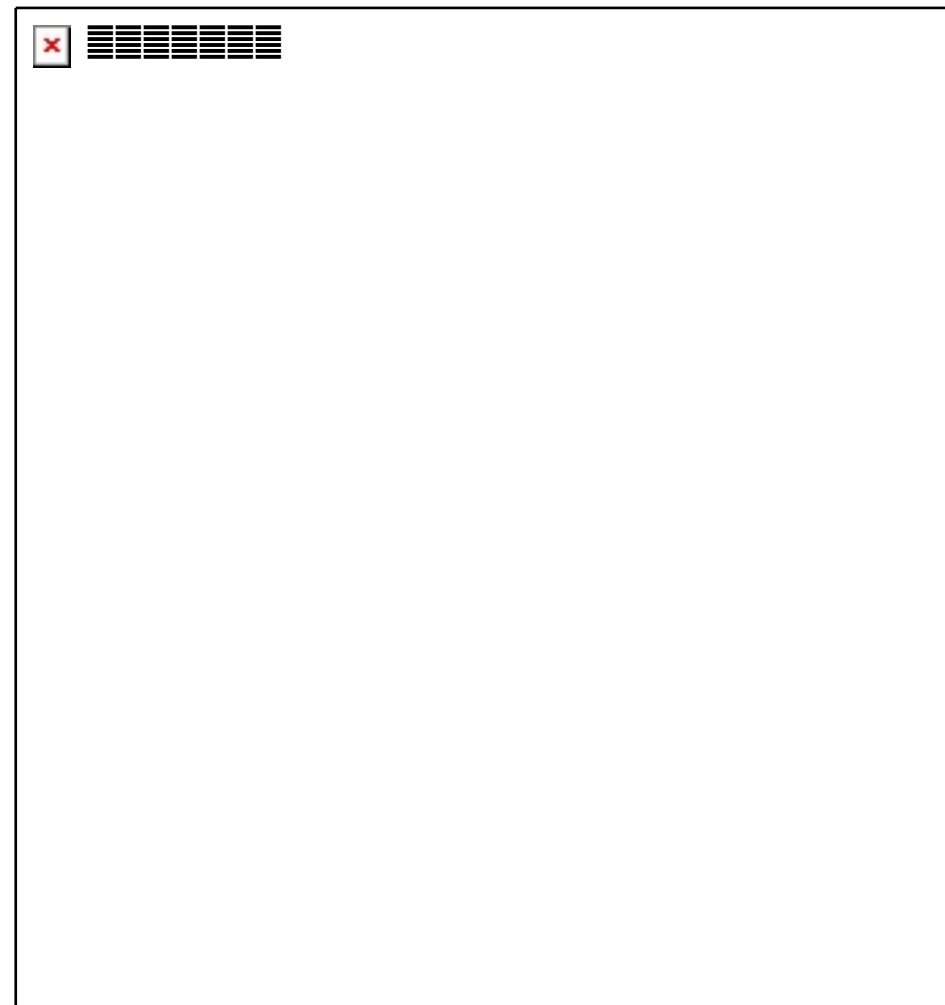
Precision = 82%

Recall = 45%

F1\_score = 52%



可以有效  
減少**45%**違約情況發生



測試資料

# 預測個人違約機率

違約機率 人數

0-1 15442

1 395

2 83

3 33

4 11

5 8

6 5

7 8

8 3

9 6

10 3

11 6

20 4

30 1

40 1

50 0

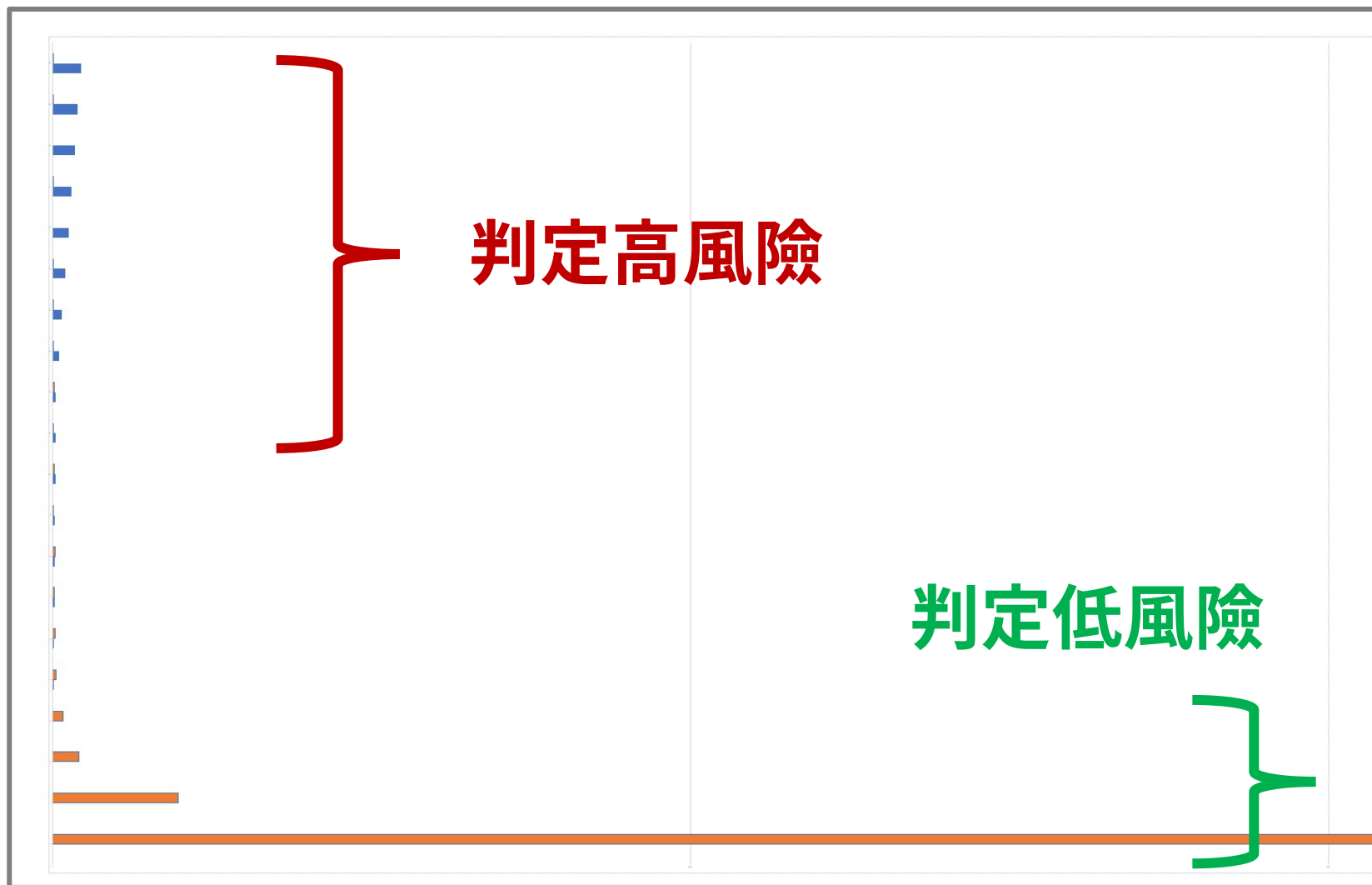
60 3

70 0

80 2

90 2

100%



人數

0%

4

系統呈現



# Telegram連結

一起試試看吧！

[https://t.me/bird\\_01\\_bot](https://t.me/bird_01_bot)

SCAN ME

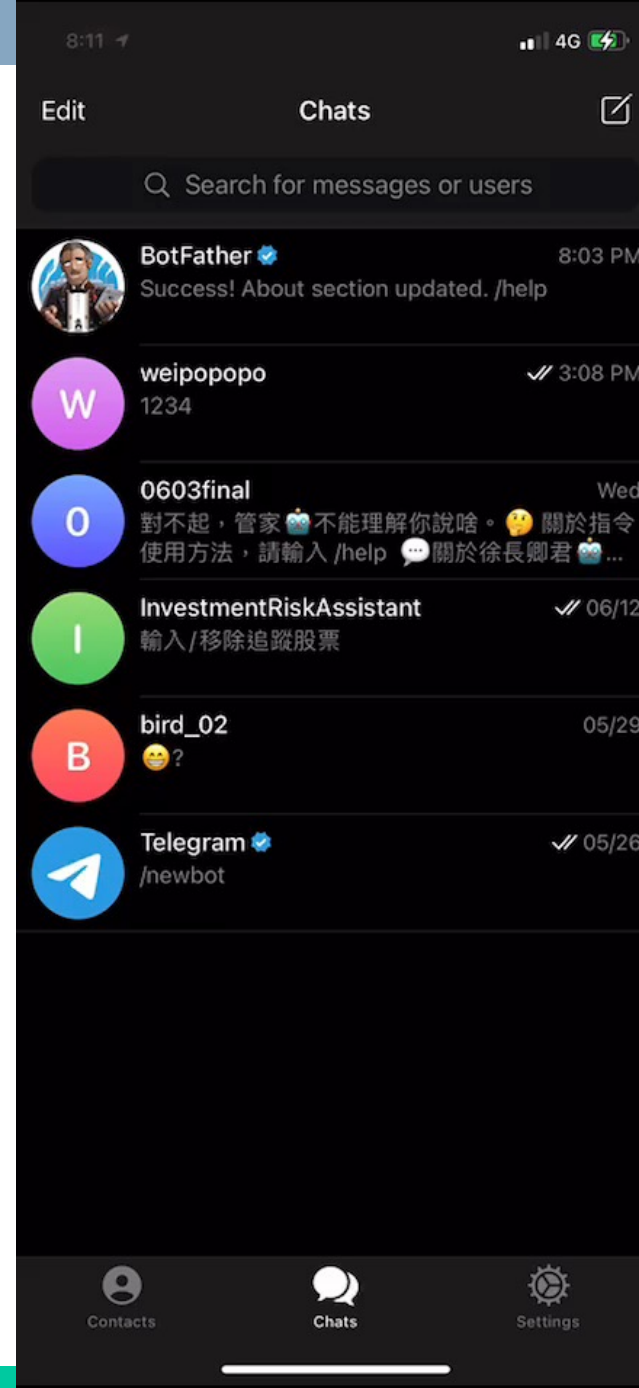




# DEMO影片

## 輸入投資的資訊，給你數據上的結果

- 以第二人稱的方式跟使用者交談
- 立即的查找股票資訊
- 給股市小白一些最簡單的名詞解釋，讓投資小白更容易進入市場
- 能夠以更親切的方式接近顧客，並依顧客需求，給予有關的網站資訊
- 擁有防呆機制，精確地給使用者想要的資訊
- 使用者免付費、安全機制良好，不怕聊天紀錄外洩。
- Telegram 擁有複雜的加密機制

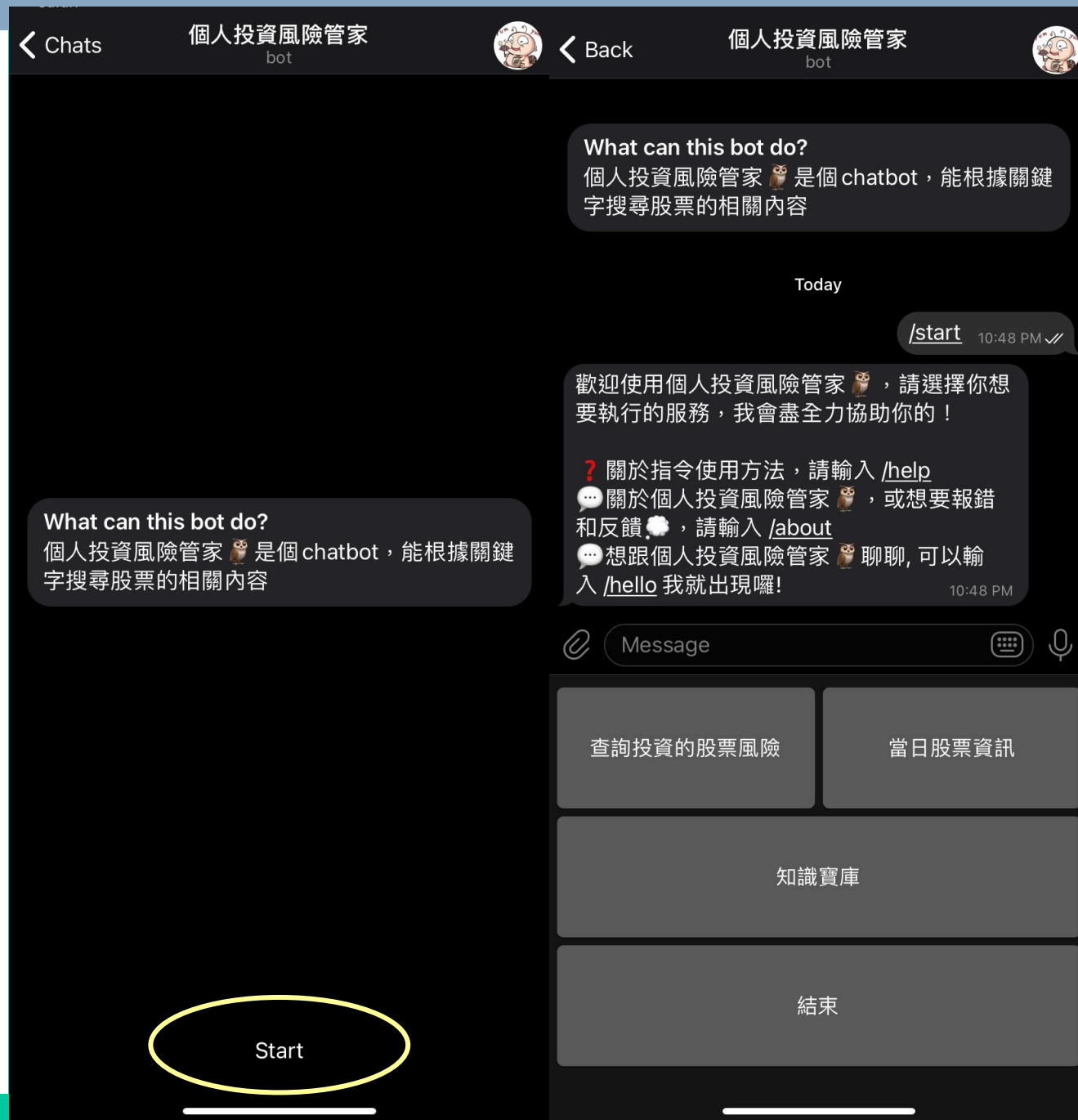


# 開始功能

## 簡單介紹

- 按下 Start 即可使用
- 自動跳出表選單供選擇

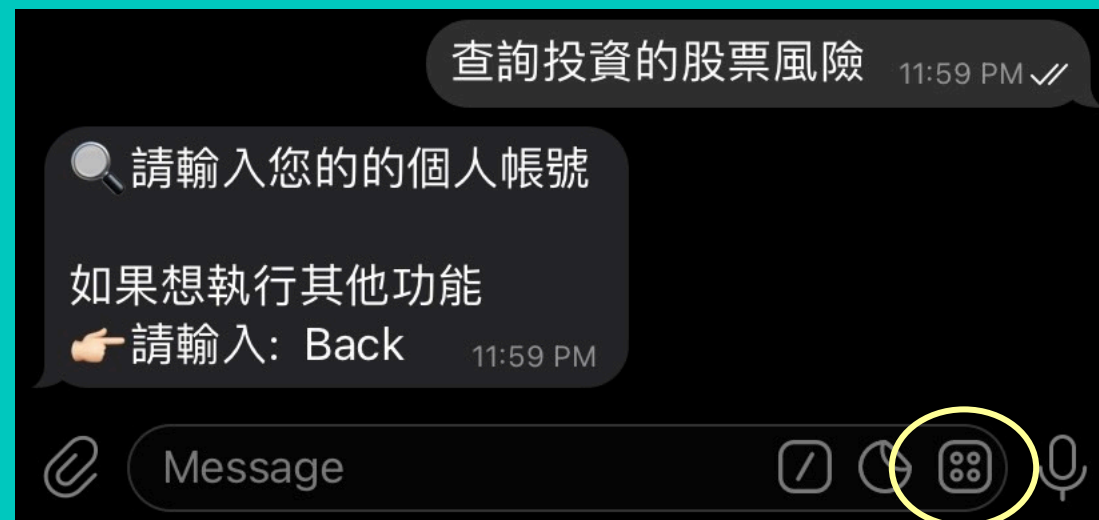
SCAN ME



# 股票違約風險

跑後臺模型並給予適當提醒

- 依使用者輸入的**帳號 股票名稱 交易價格 數量 日期** 來判斷此行為風險的大小
- 欲使用其他功能可按旁邊的選單鍵
- 回傳**風險等級**，依照顧客風險貼心提醒



# 股票資訊

可依不同日期、代碼給予  
當日的股票資訊

- 爬取**台灣證券所**的股票資料
- 若有查找不到的股票，給予**連結**
- 提供重點資訊



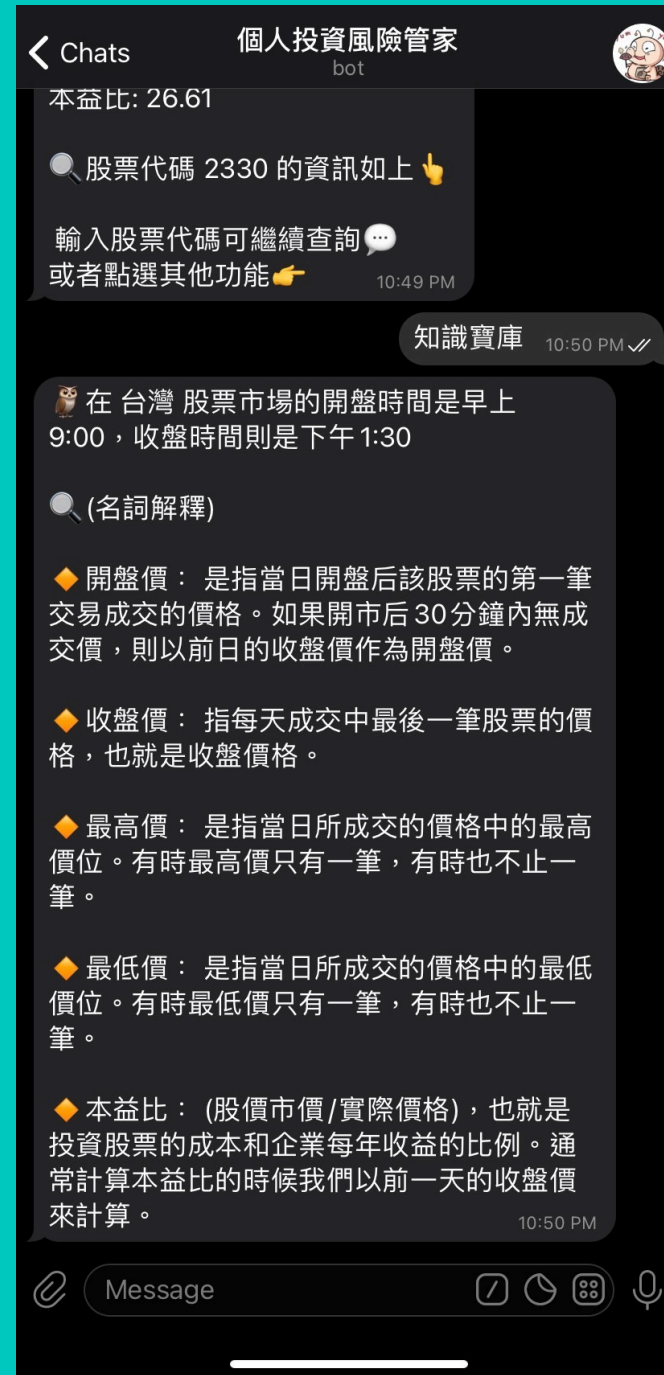
臺灣證券交易所



# 知識寶庫

專門為股市小白設計

- 給予些許的台股小知識
- 股票資訊的名詞做更进一步的解釋

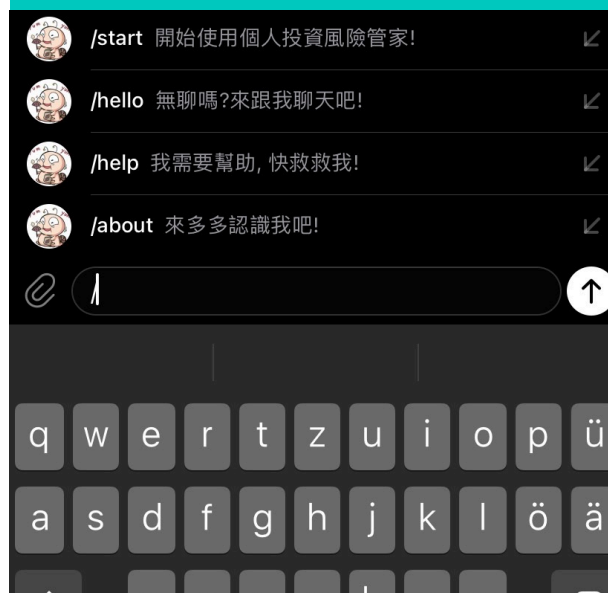


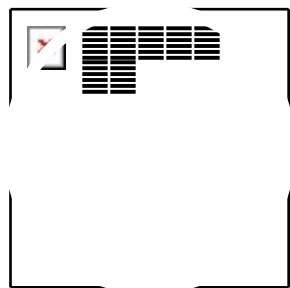
# 其他功能



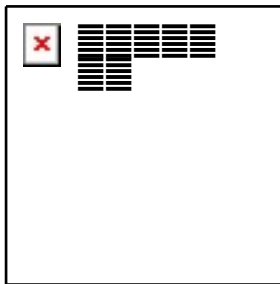
打出 / 即可叫出

- HELLO 與使用者簡單對話
- HELP 使用方法說明
- ABOUT 機器人自我介紹

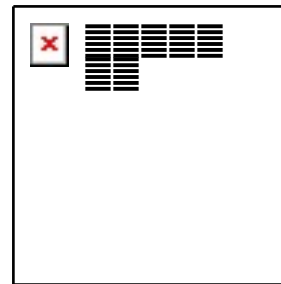




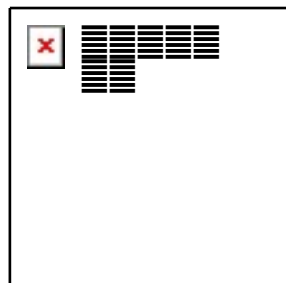
親切對談



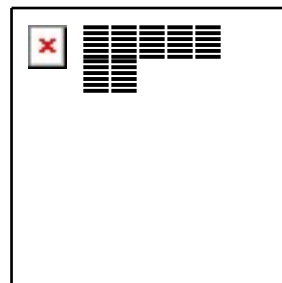
適當提醒



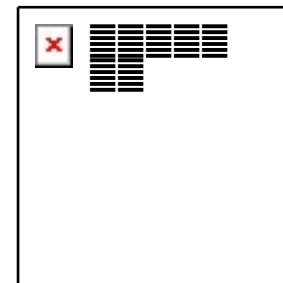
全部免費



即時資訊



防呆機制



加密機制

服務優點





5

附錄

# 工作分配

---

## 郭奕廷

前期資料分析  
Random Forest模型  
LSTM模型  
違約機率預測功能  
簡報和影片設計

## 賴威博

資料前處理  
Trello更新  
Telegram製作  
影片製作

## 李佳盈

資料前處理  
爬蟲  
Telegram製作  
簡報設計  
影片製作

## 詹佳蓓

前期資料分析  
Logistic regression  
違約機率預測功能  
Telegram製作  
程式碼統整

## 歐崇愷

前期資料分析  
XGBOOST模型  
decision tree模型  
違約機率預測功能

## 短片連結

<https://drive.google.com/file/d/1LH326OUPSMotHFRXlfjxaO9XaEB2JY1j/view?usp=sharing>



# 感謝聆聽

祝福大家投資順利

## 玉山-第四組