

# 人工智慧與金融科技

## 期末報告第四組

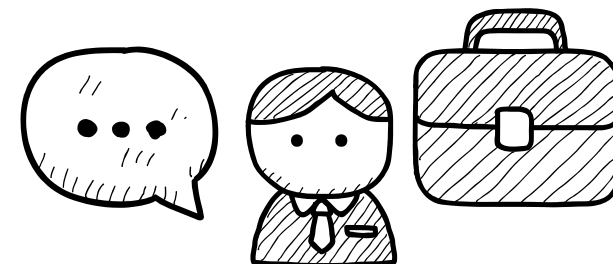


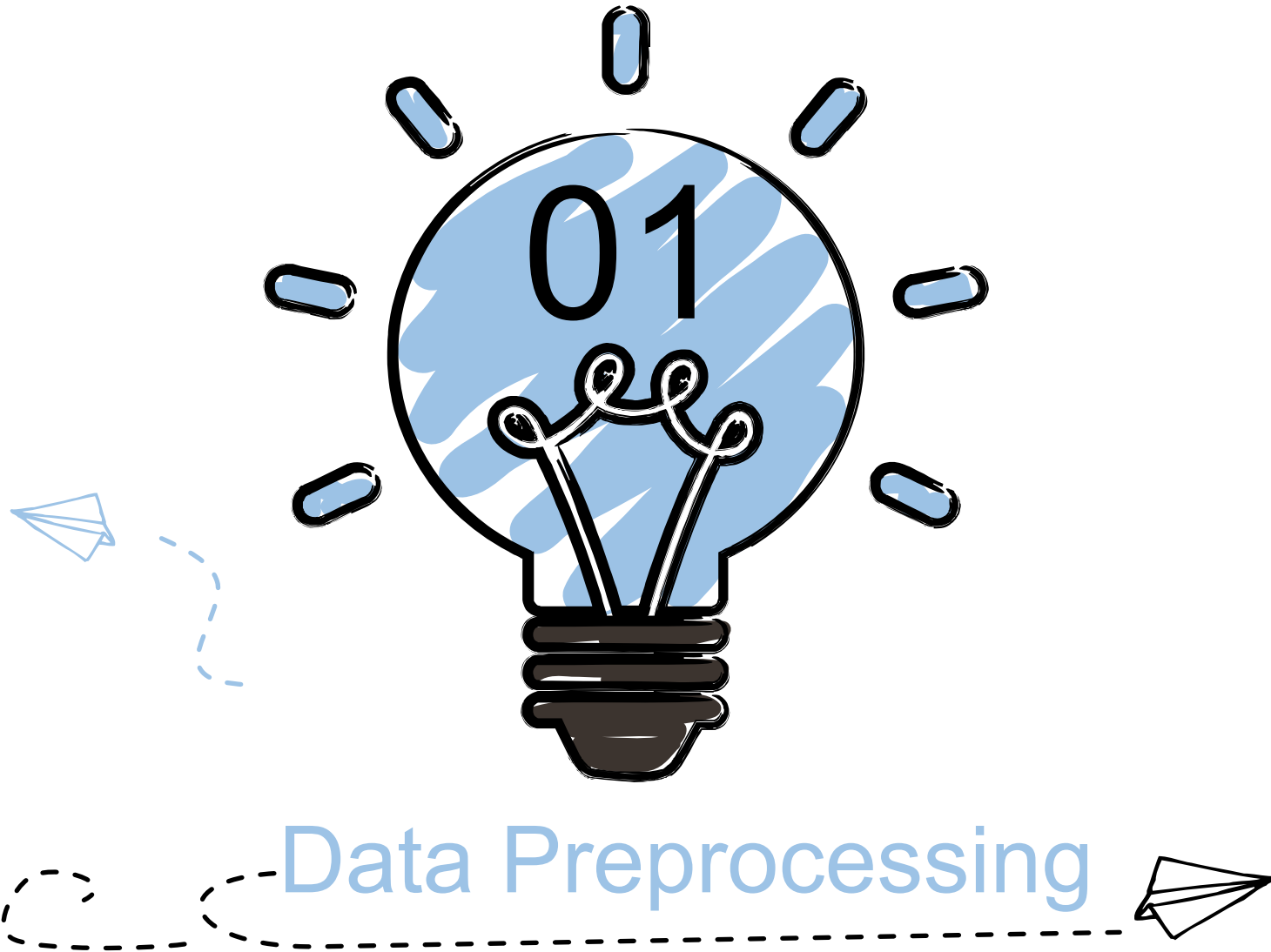
A1095502 沈品豪  
A1095517 黃泓茗  
A1095521 鄭惠心  
A1095527 廖習驊  
A1095557 陳柏安  
A1095560 歐津萍  
A1095565 張庭瑋



# 目錄

1. Preprocessing
2. Strategy
3. GA
4. KNN
5. Decision tree
6. Random forest
7. XGBoost
8. Demo
9. Feature importance







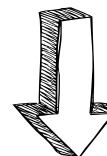
本金:10000元



買進:Return\_Label = 1

賣出:Return\_Label = -1

每年都有做買賣

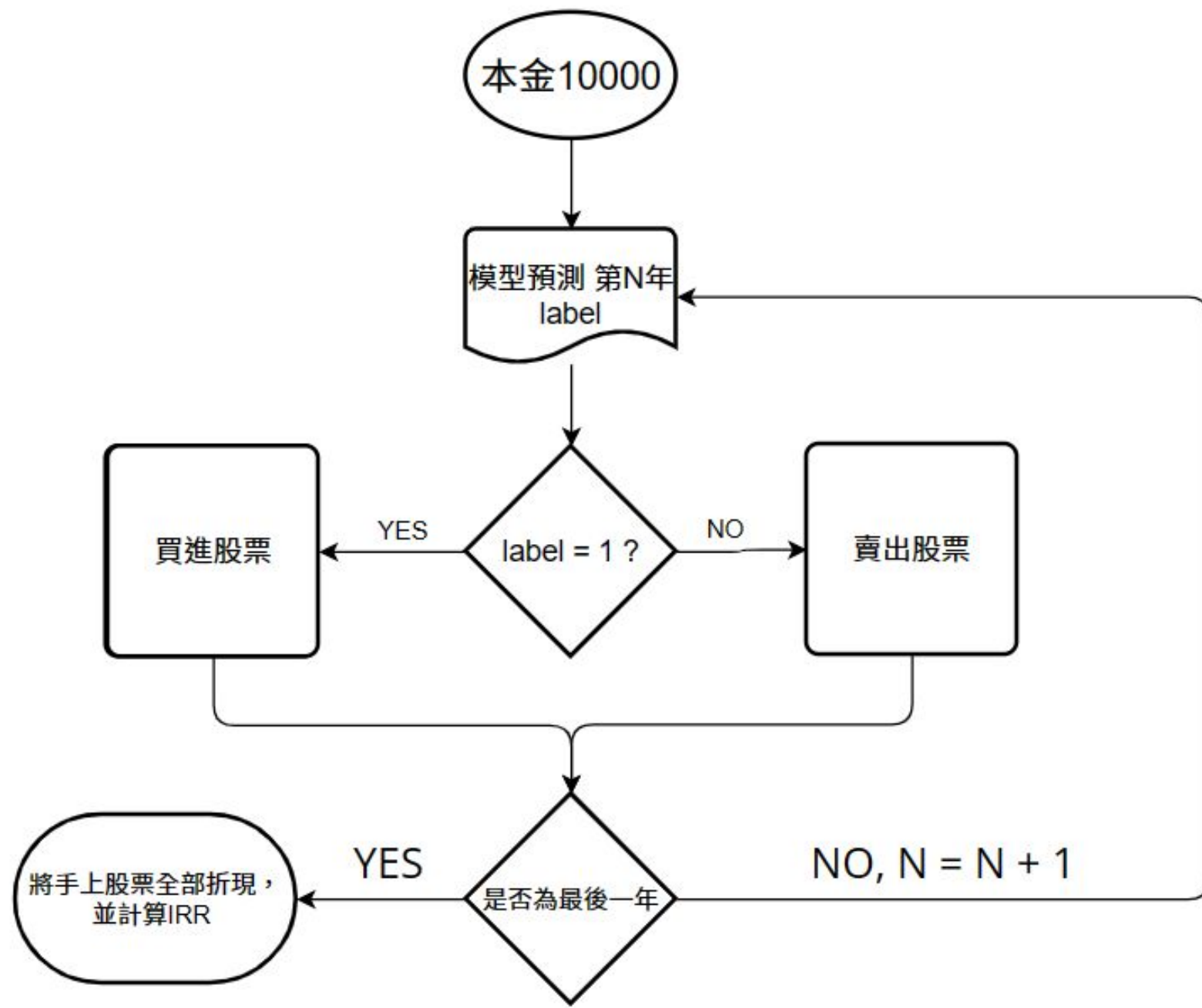


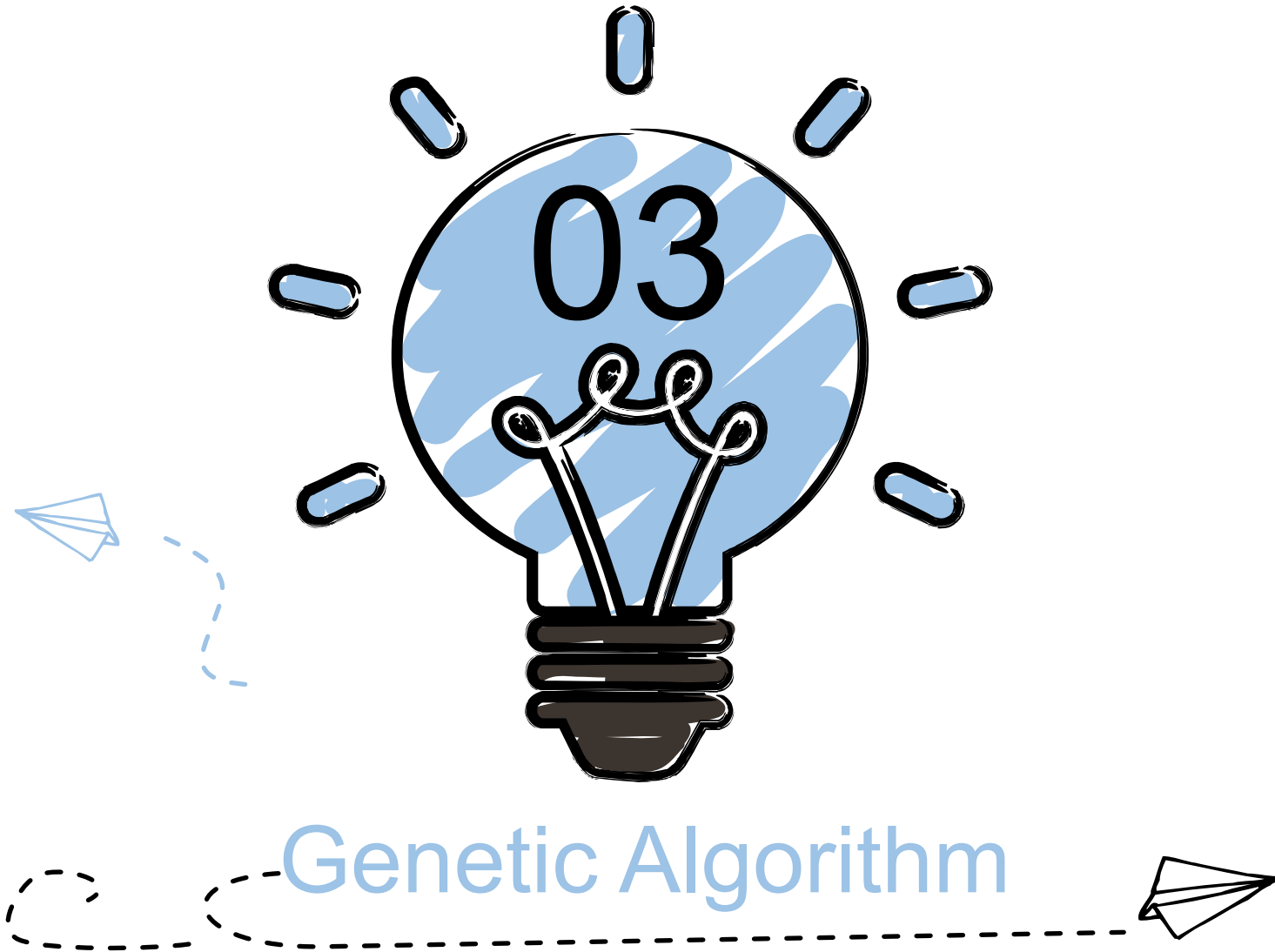
(1)紀錄要買哪幾支股票

(2) $\text{current\_money} / \text{買的股票數量} = \text{每隻股票要花多少錢去買}$

(1) $\text{current\_money} += \text{收盤價} * \text{每隻股票的數量}$

最終折現: 將現有股數\*2009年的股票收盤價 = 賺多少錢





## [GA\_KNN]

- **attribute**: 16bits
- **K**: 4bits + 1 (1 ~ 16)

## [GA\_DST , GA\_RF]

- **attribute**: 16bits

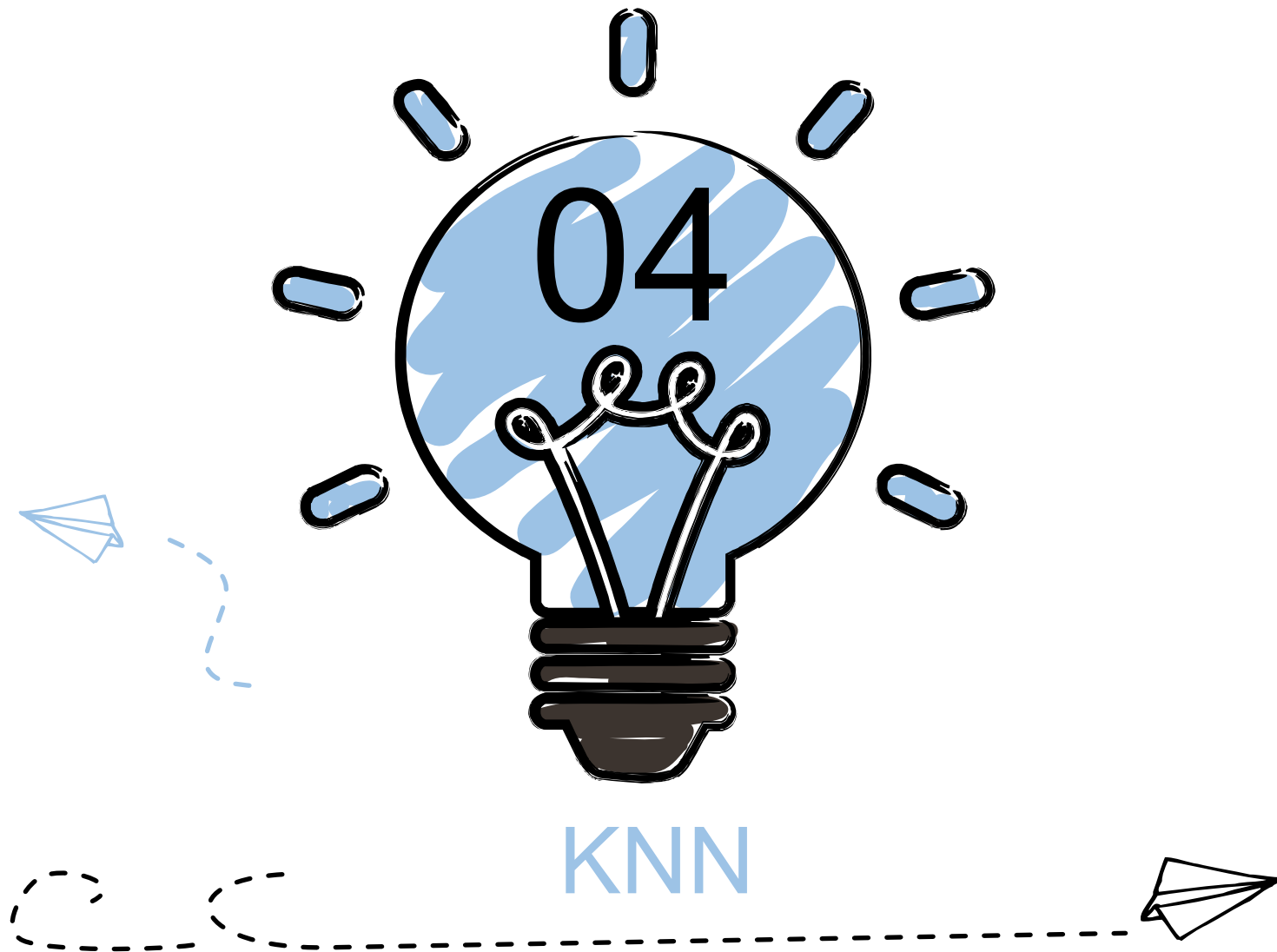
## [GA\_XGBoost]


- **attribute**: 16bits
- **parameter**: 24bits -> learning\_rate[0.01~1]: 4bits, eta[0.01~0.2]: 4bits, gamma[0.01~10]: 4bits, max\_depth[1~10]: 3bits, subsample[0.01~1]: 3bits, min\_child\_weight[0.01~10]: 3bits, colsample\_bytree[0.01~1]: 3bits

## ▲ Demo

-> **attribute**: 21bits







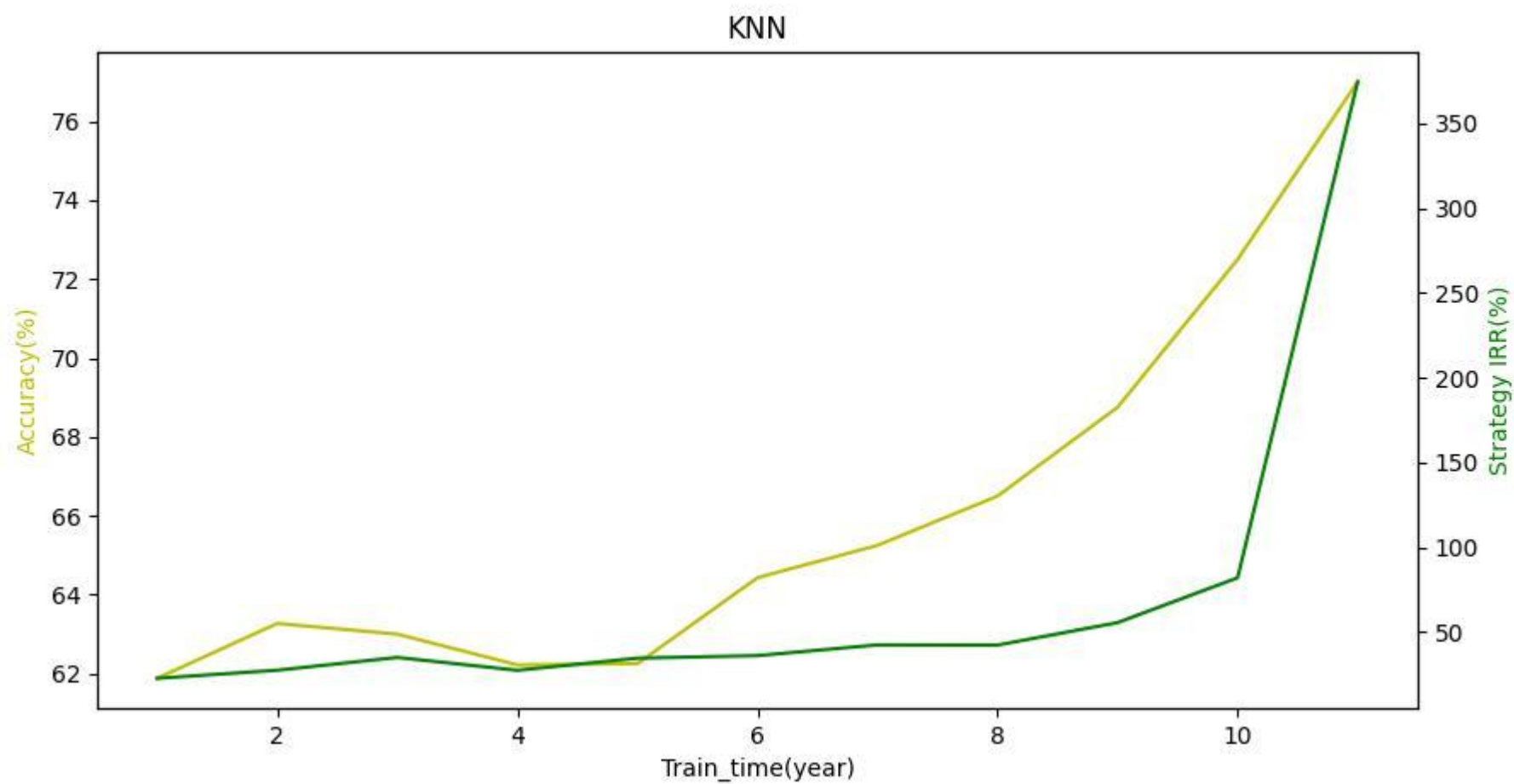
## KNN Accuracy

TV1	61.88%
TV2	63.27%
TV3	63%
TV4	62.22%
TV5	62.25%
TV6	64.43%
TV7	65.25%
TV8	66.5%
TV9	68.75%
TV10	72.5%
TV11	77%



TV1	22.65%
TV2	27.42%
TV3	34.96%
TV4	27.37%
TV5	34.53%
TV6	36.03%
TV7	42.35%
TV8	42.24%
TV9	55.48%
TV10	81.94%
TV11	374.44%

# KNN Result







## Decision\_Tree Accuracy

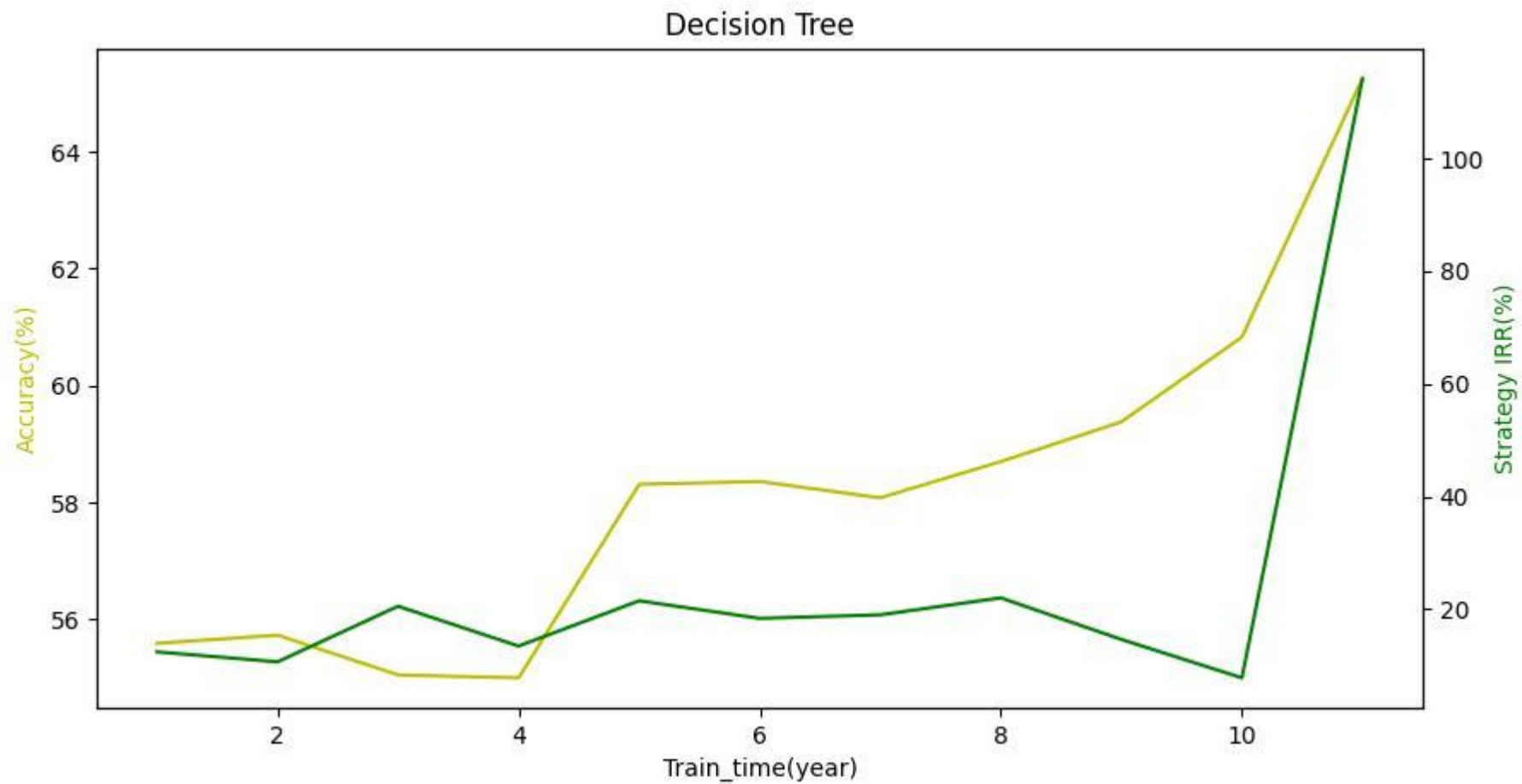
TV1	<b>55.59%</b>
TV2	<b>55.73%</b>
TV3	<b>55.05%</b>
TV4	<b>55%</b>
TV5	<b>58.31%</b>
TV6	<b>58.36%</b>
TV7	<b>58.08%</b>
TV8	<b>58.7%</b>
TV9	<b>59.38%</b>
TV10	<b>60.83%</b>
TV11	<b>65.25%</b>



## Decision\_Tree IRR

TV1	12.43%
TV2	10.66%
TV3	20.54%
TV4	13.46%
TV5	21.5%
TV6	18.38%
TV7	19.01%
TV8	22.02%
TV9	14.64%
TV10	7.82%
TV11	114.19%

# Decision\_Tree Result







定義

“decision tree  
+  
bagging”

雖然資料重新放回後  
資料會重複  
但訓練出來的tree  
結果仍不同

## bagging

1.

從train\_data中隨意  
抽取N筆資料  
( $N < \text{total data num}$ )  
(取出後放回)

2.

再從此N筆資料中  
選出K種特徵  
訓練出M棵  
decision tree

3.

每個decision tree  
權重相同

4.

最終用Majority vote  
(多數決)  
得出結果

## 優點



01

已多棵樹取代深度大，避免過度擬合

02

透過bagging抽樣來避免選出噪音資料  
穩定度增高

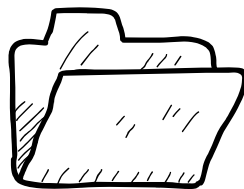
03

隨機抽取資料及特徵

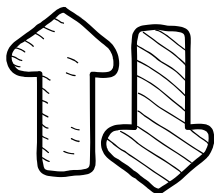
04

每棵decision tree皆獨立

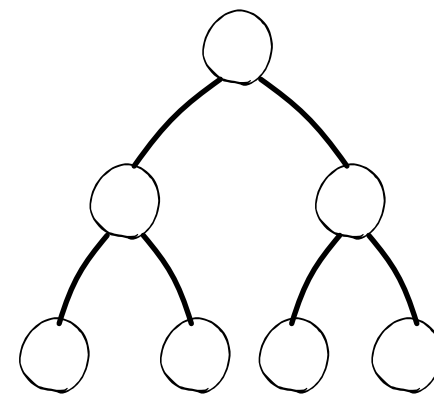
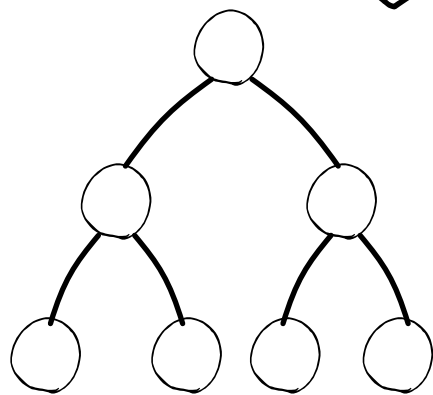
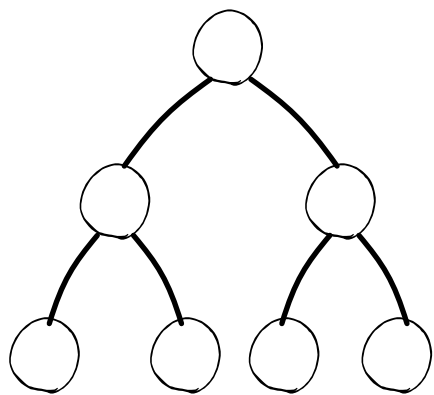
重複  
隨機  
選取  
K  
筆  
特徵  
做出  
決策  
樹



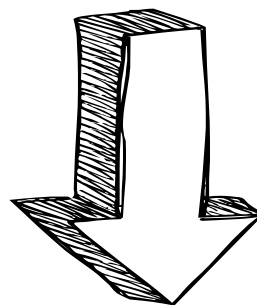
用完放回



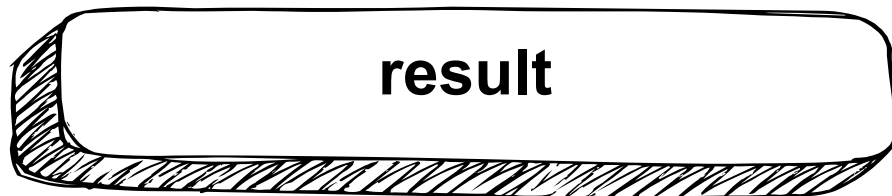
隨機選出N筆資料



majority vote



result





## Random\_Forest Accuracy

TV1 58.12%

TV2 57.36%

TV3 58.30%

TV4 59.77%

TV5 63.12%

TV6 63.07%

TV7 63.00%

TV8 65.60%

TV9 66.62%

TV10 68.50%

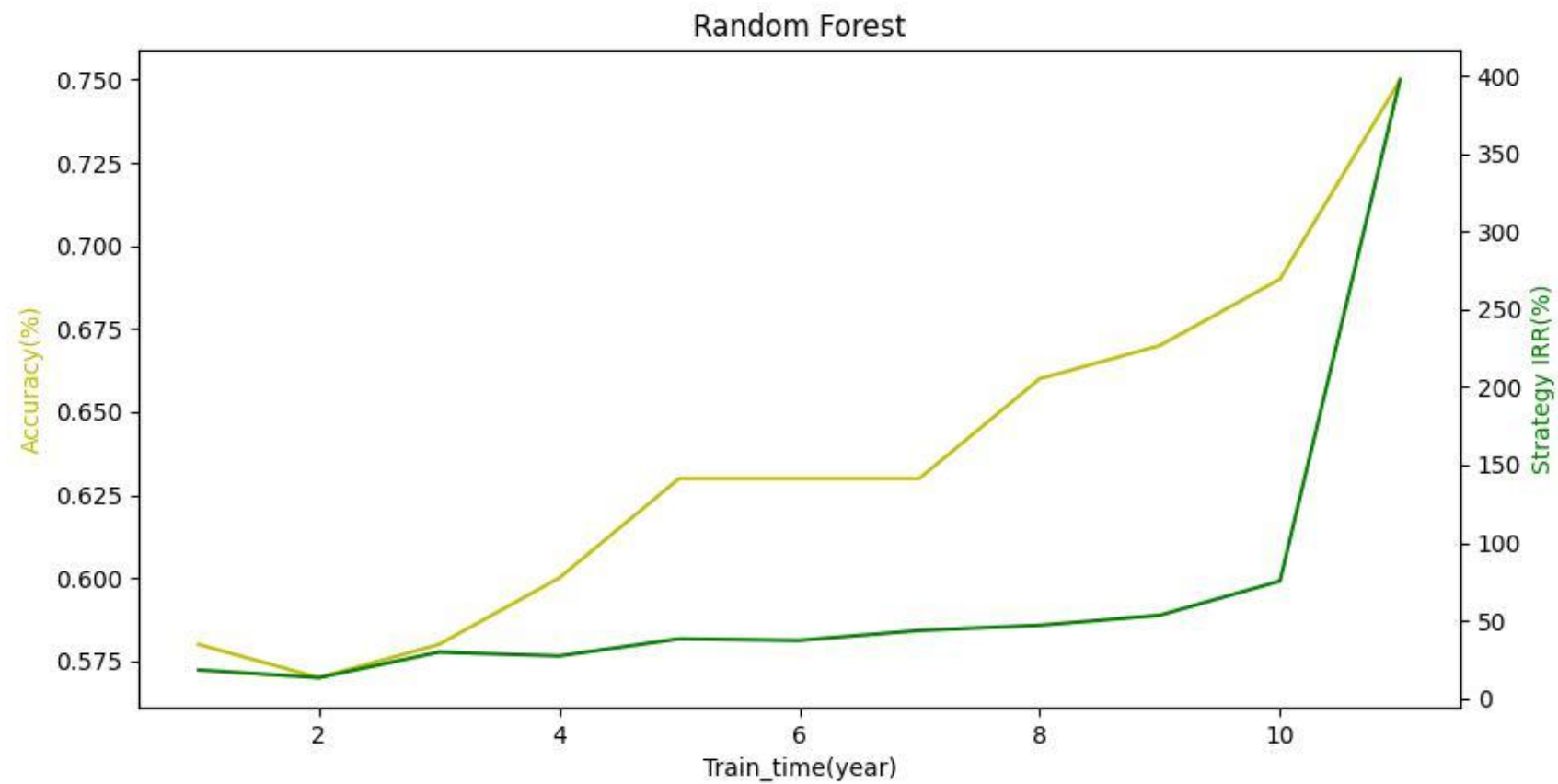
TV11 75.25%



## Random\_Forest IRR

TV1	18.27%
TV2	13.35%
TV3	29.76%
TV4	27.32%
TV5	38.28%
TV6	37.28%
TV7	43.67%
TV8	47%
TV9	53.44%
TV10	75.44%
TV11	397.4%

# Random\_Forest Result



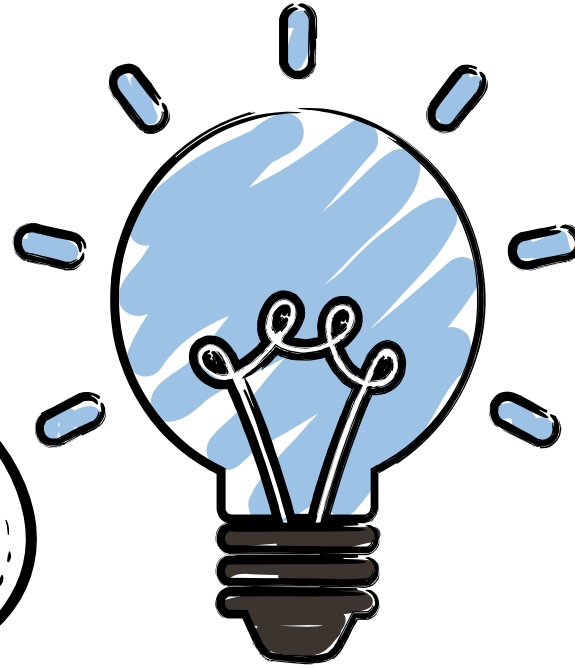




XGBoost

(Extreme Gradient Boosting)

# Boosting



將許多weak classifier  
結合起來變成一個  
一個strong classifier

把舊classifier的錯誤資料權重提高  
來訓練新的classifier  
以讓新classifier學習到  
錯誤資料的特性  
以提升分類結果

因此若boosting的訓  
列資料中有太多噪音  
會嚴重影響分類性能

在做boosting前  
需先告知該weak  
classifier的準確率  
下限

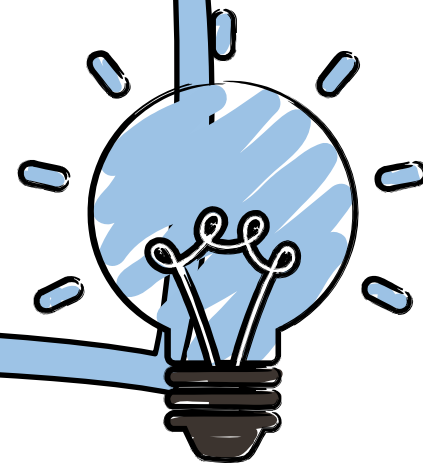
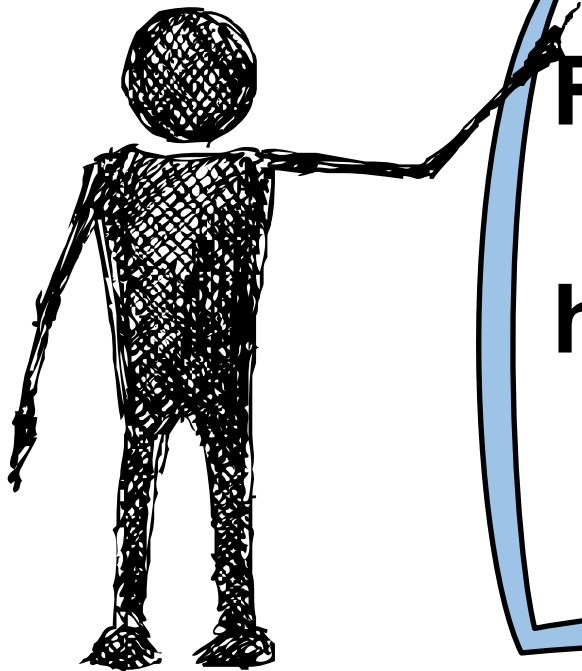
# Gradient Boosting

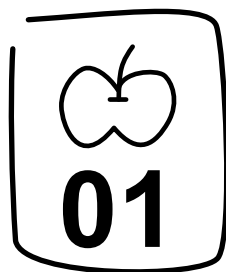
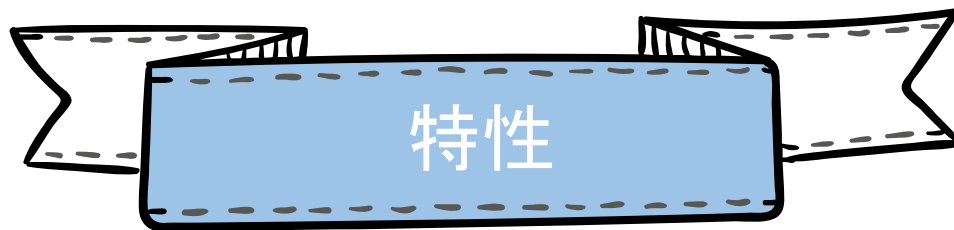
目的:

$F(x)$ : 原先的預測模型

$h(x)$ : 再用分類器訓練出  
新的預測模型

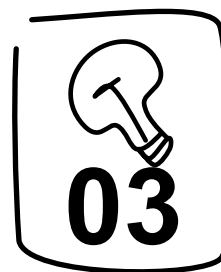
$F(x) + h(x) \rightarrow$   
實際值





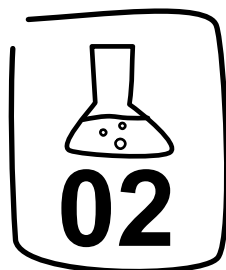
### 樹和樹之間關聯

後面的樹會修正前面樹的錯誤



### 標準化

使用L1/L2 Regularization  
降低過度擬合機率



### 特徵隨機採樣

不會一次拿全部的特徵來生成樹



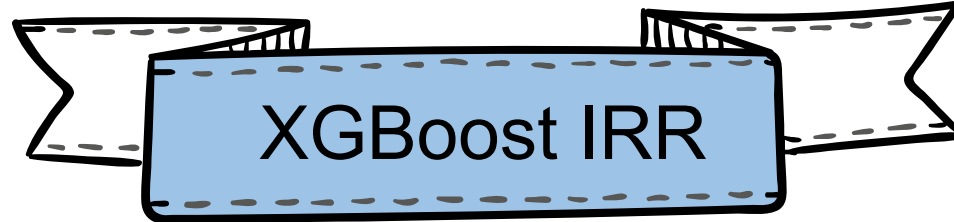
### 生成新的樹

一階導數: Gradient  
二階導數: Hessian



## XGBoost Accuracy

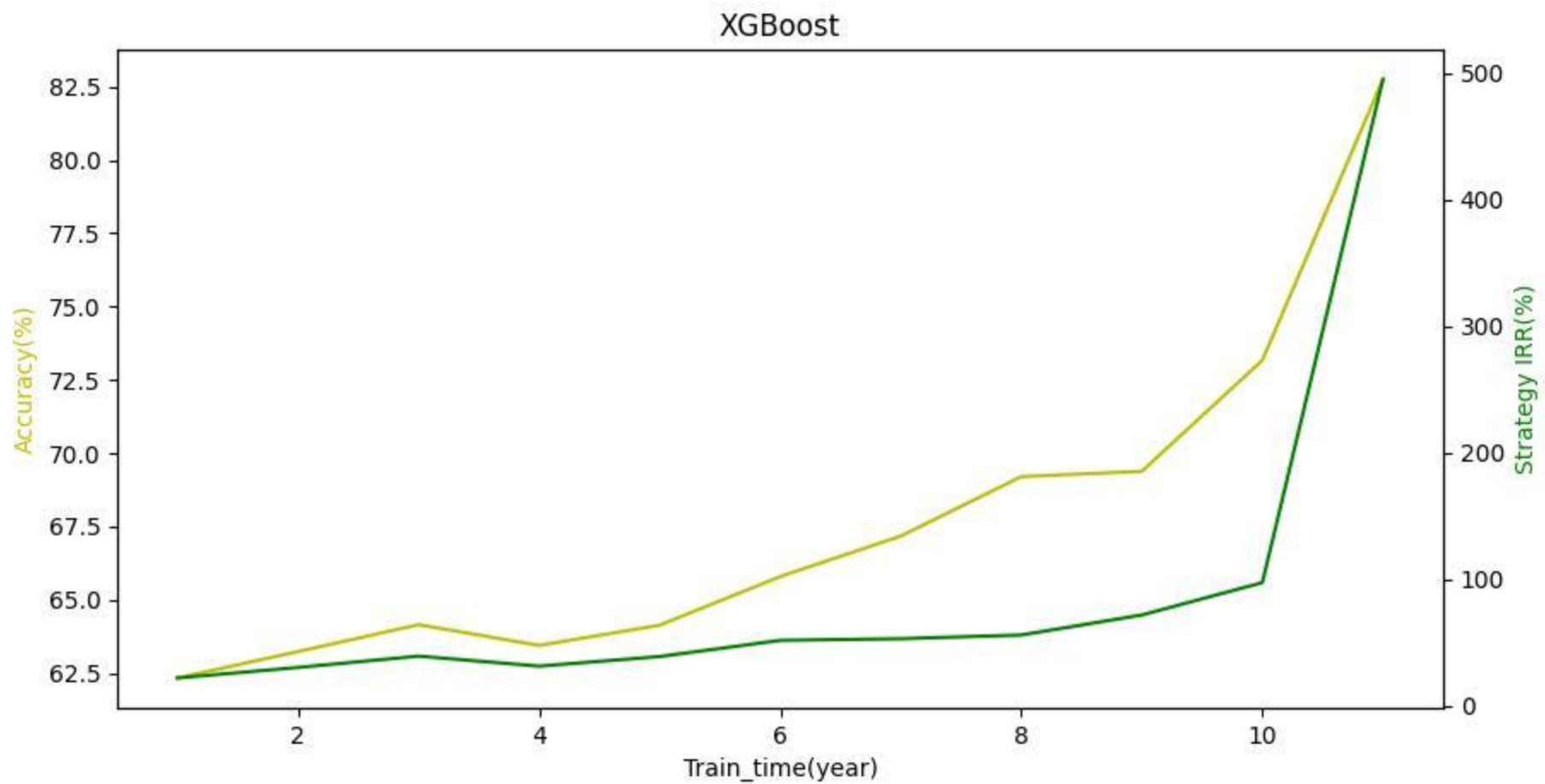
TV1	62.33%
TV2	63.23%
TV3	64.15%
TV4	63.44%
TV5	64.13%
TV6	65.79%
TV7	67.17%
TV8	69.2%
TV9	69.38%
TV10	73.17%
TV11	82.75%

A blue banner with a black outline and dashed lines, featuring the text "XGBoost IRR" in black.

## XGBoost IRR

TV1	21.93%
TV2	30.23%
TV3	39.19%
TV4	31.26%
TV5	38.85%
TV6	51.63%
TV7	52.95%
TV8	55.87%
TV9	71.67%
TV10	97.28%
TV11	494.77%

# XGBoost Result









## KNN Accuracy

TV1	56.6%
TV2	59.5%
TV3	59.17%
TV4	61.75%
TV5	62%

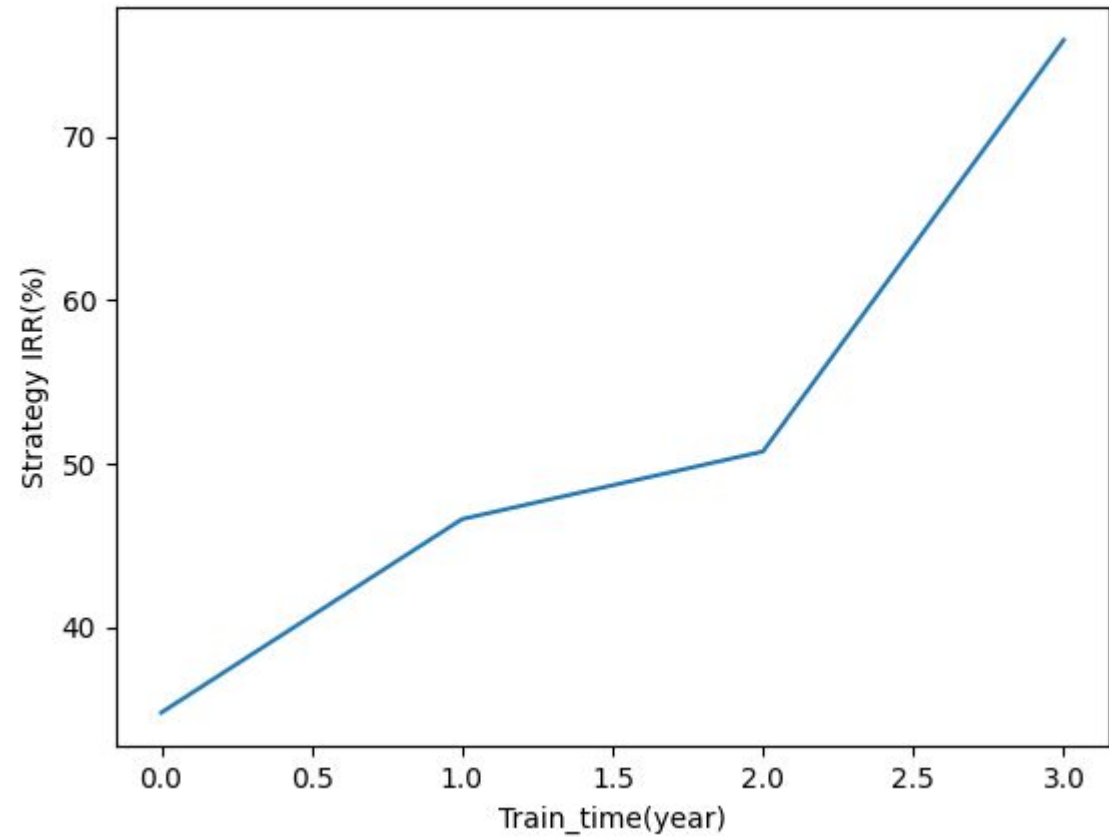


**TV1**      **34.78%**

**TV2**      **46.62%**

**TV3**      **50.75%**

**TV4**      **75.93%**





## Decision\_Tree Accuracy

TV1	<b>56.5%</b>
TV2	<b>58.38%</b>
TV3	<b>59%</b>
TV4	<b>58.25%</b>
TV5	<b>63.5%</b>

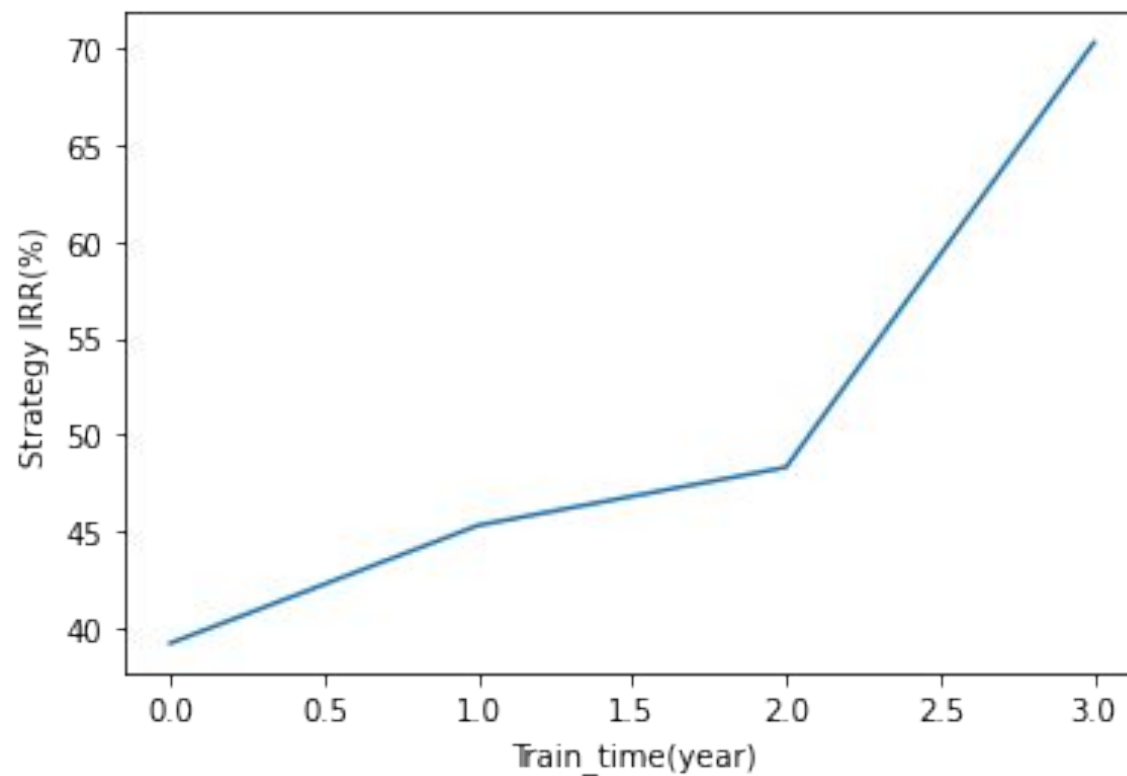
## Decision\_Tree IRR

**TV1 39.2%**

**TV2 45.3%**

**TV3 48.31%**

**TV4 70.3%**



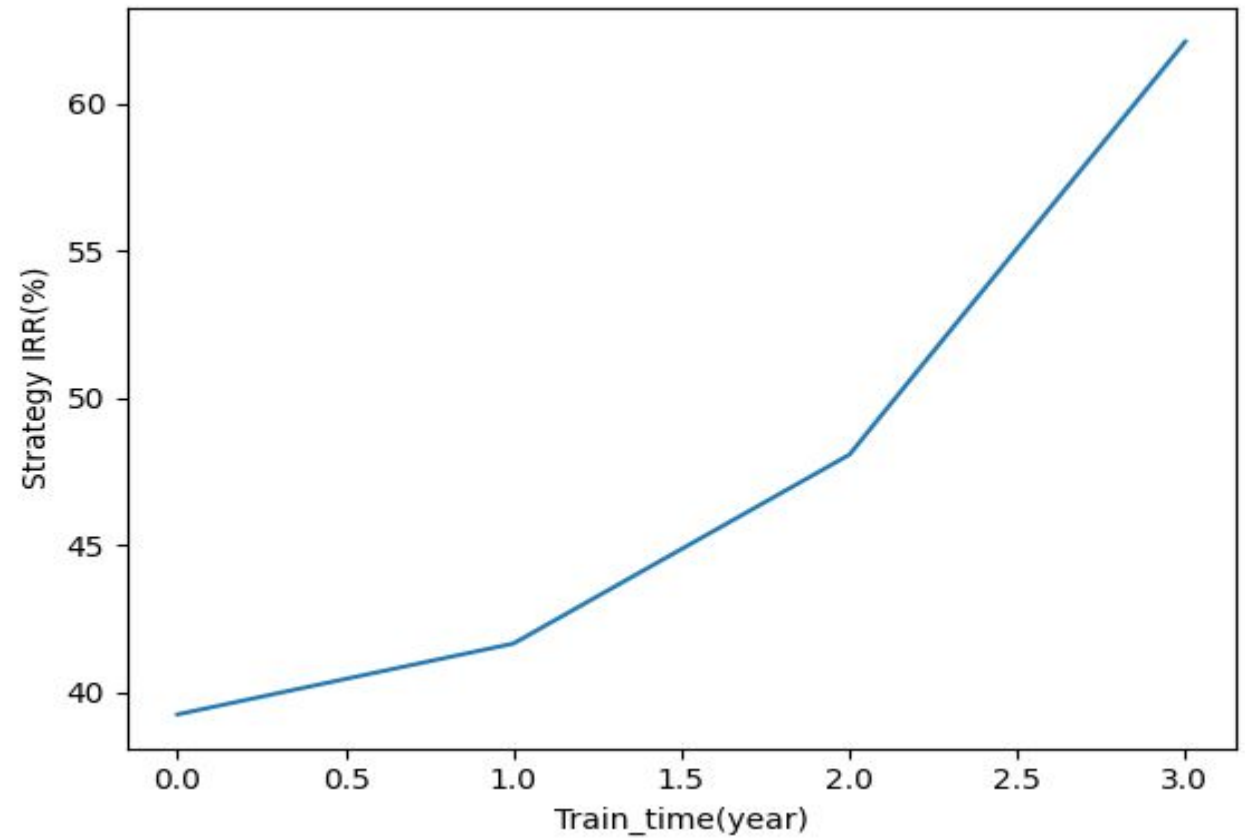


## Random\_Forest Accuracy

TV1	59.9%
TV2	59.13%
TV3	61.83%
TV4	59.5%
TV5	60%

## Random\_Forest IRR

TV1	39.26%
TV2	41.66%
TV3	48.08%
TV4	62.1%





## XGBoost Accuracy

TV1	58.7%
TV2	60.63%
TV3	62.5%
TV4	61.25%
TV5	62.5%

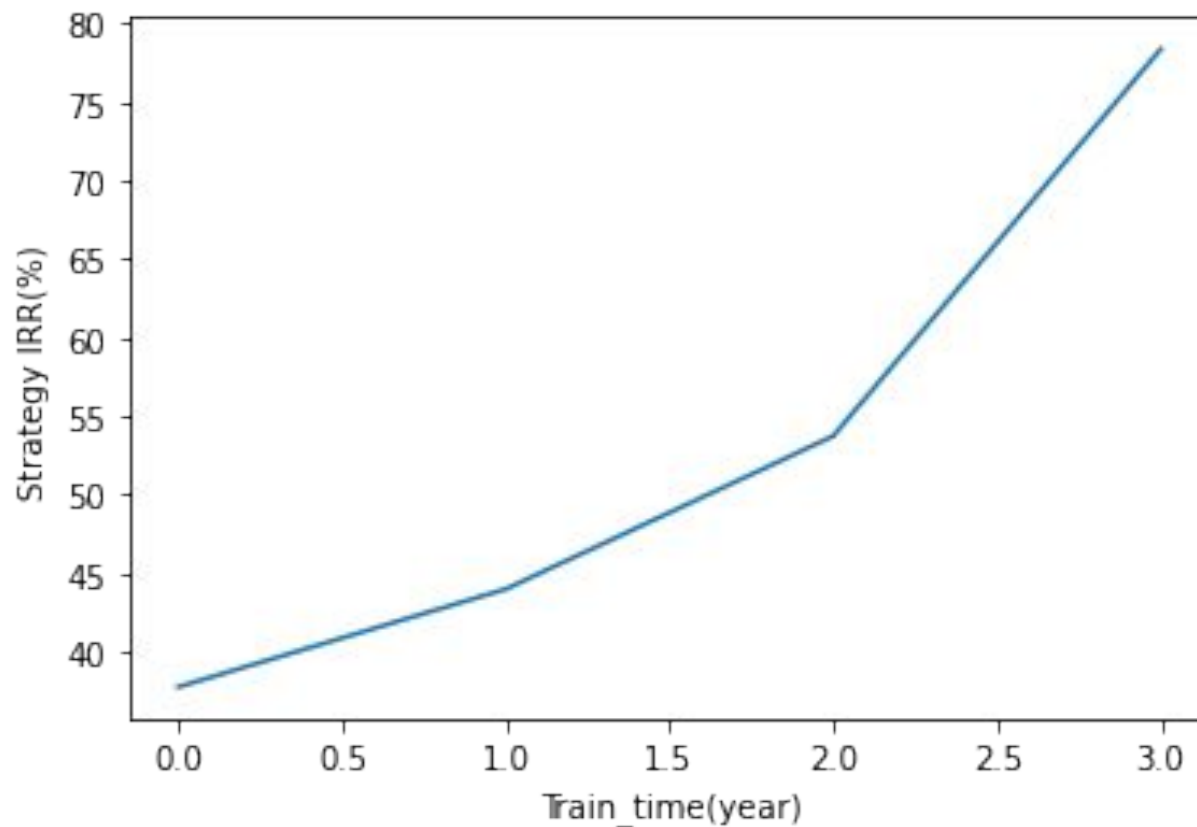
# XGBoost IRR

TV1 37.78%

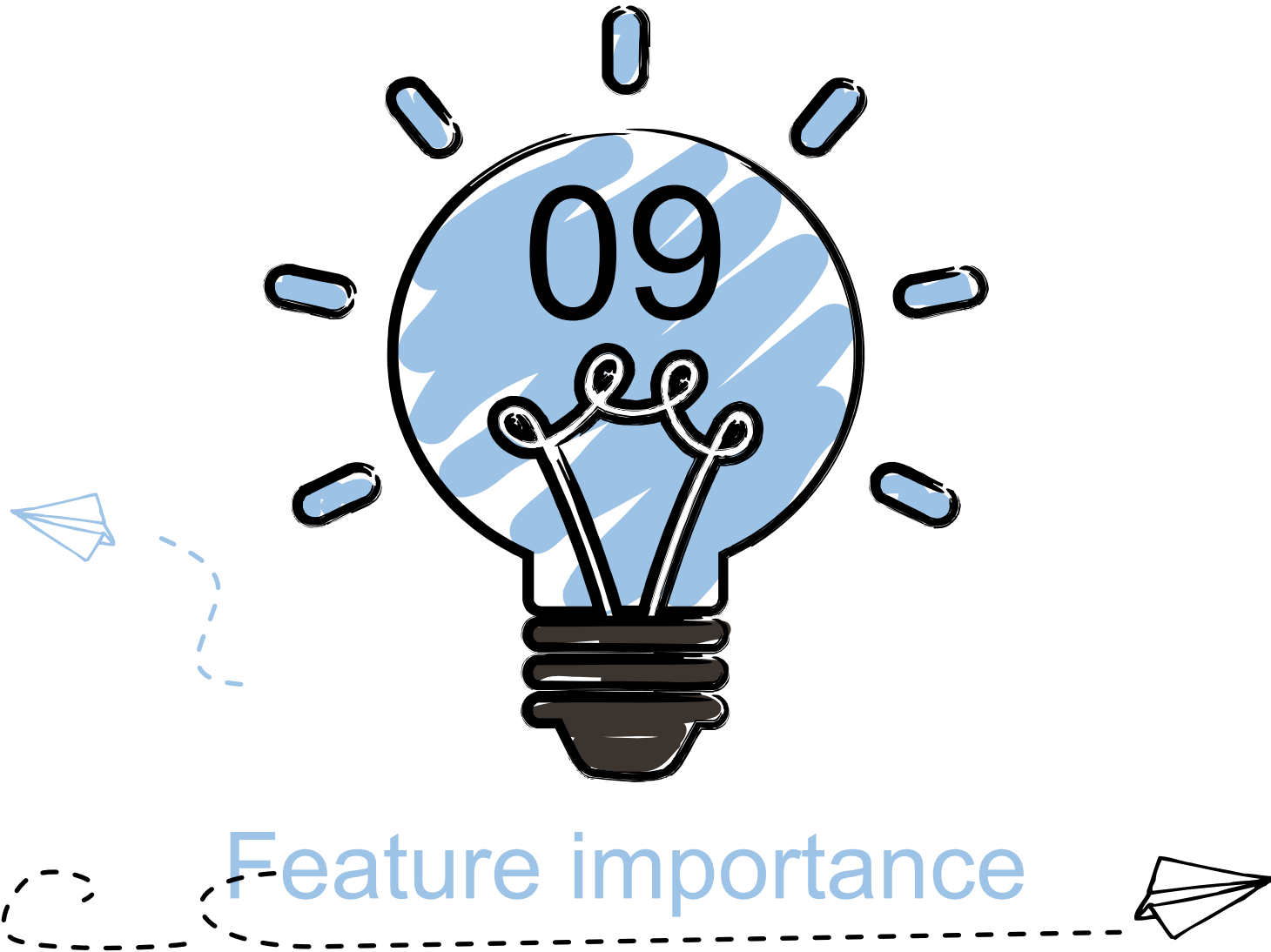
TV2 44%

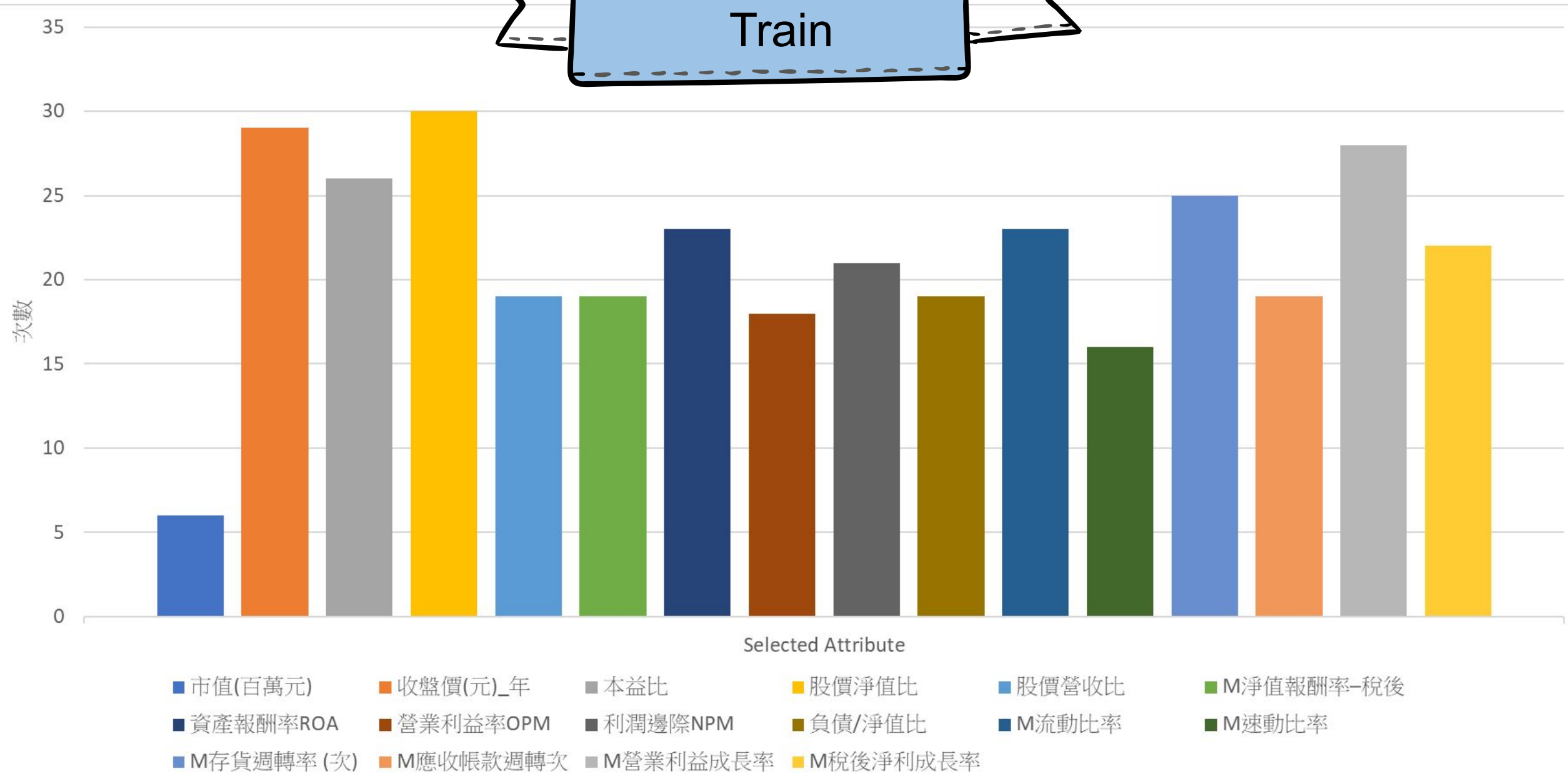
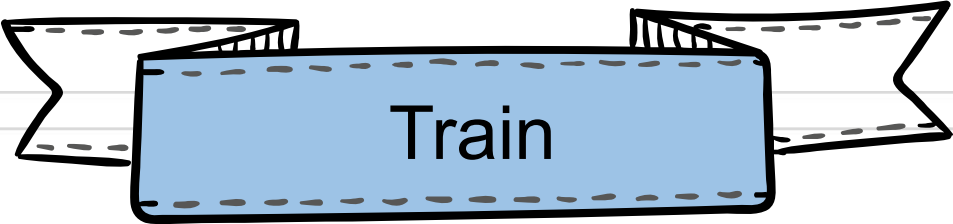
TV3 53.74%

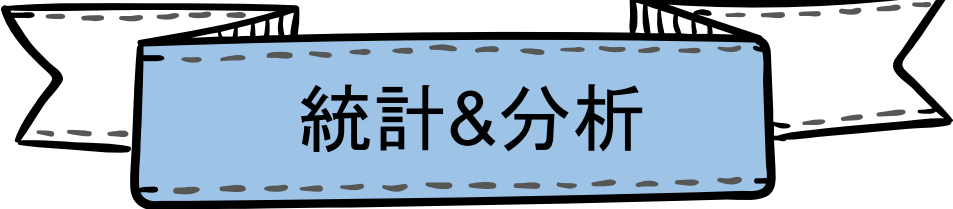
TV4 78.37%





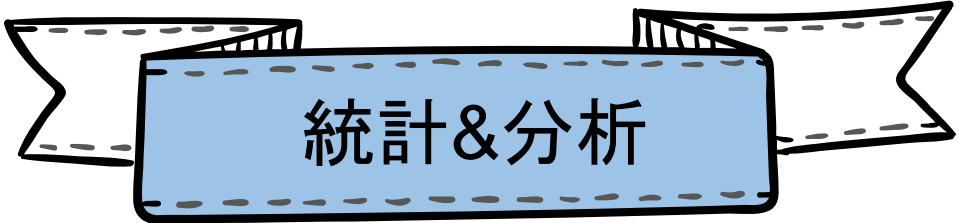






# 統計&分析

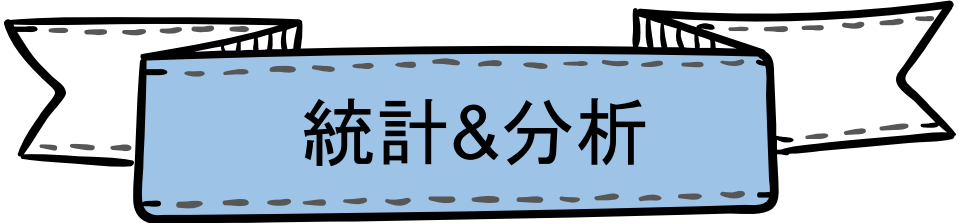
較為重要	較不重要
股價淨值比、收盤價(元)_年、M營業利益成長率、 本益比	市值(百萬元)、M速動比率



## 統計&分析

較重要的attribute:

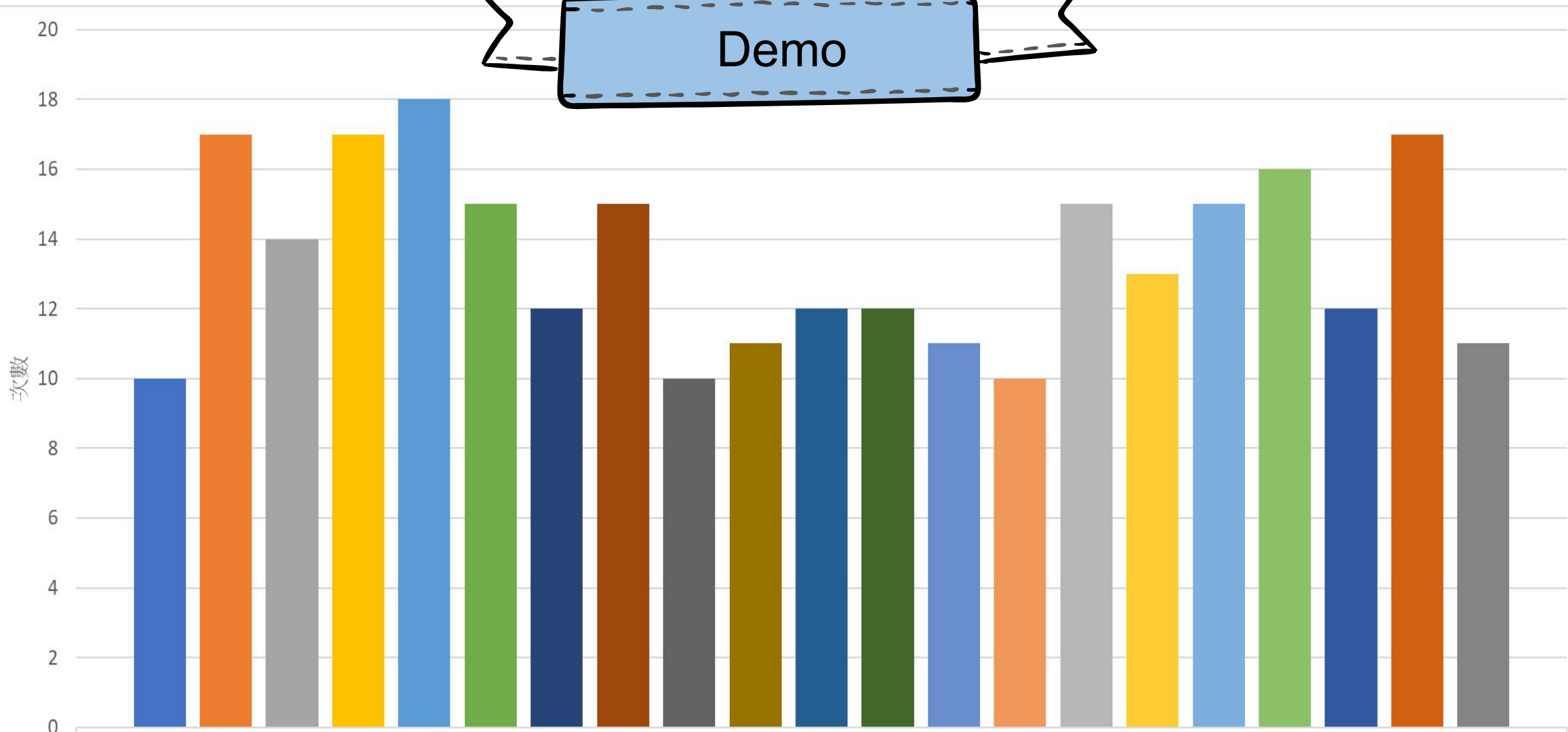
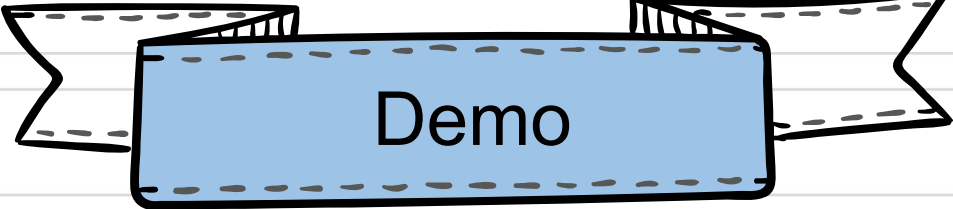
1. **股價淨值比**=股價/(每股淨值 (總資產-總負債)):願意花幾倍的每股淨值去投資這支股票。當股價淨值比 $<1$ :股價較低, 潛在報酬率高, 可考慮買進;反之( $>1$ ):股價較貴, 潛在報酬率低, 可考慮賣出。在投資股票時, 是一個重要的attribute!
2. **收盤價**:可與前一交易日的收盤價格, 或下一交易日的開盤價格做比較, 作為評估該資產在單一交易日的價格變動之方式。分析股票價格長期變化的關鍵性attribute!
3. **營業利益成長率** = (當期的營業利益 - 去年同期的營業利益) / 去年同期的營業利益 \* 100%:可得知與去年相比, 此公司的營收狀況是進步或是退步。例如:當要買賣股票時, 若我們發現有一間公司連續5年營業利益成長率 $>0$ , 就代表此公司正在成長當中, 所以我們可以考慮買入, 是一個重要的attribute!
4. **本益比** = 股價 / (每股盈餘 (EPS)):可評估一股票是貴或便宜, 並且預估買此股票的回本時間, 亦是用來評估企業股價與獲利潛力的判斷標準。股票的潛力對於是否要購買一股票是非常重要的attribute!



## 統計&分析

較**不重要**的attribute:

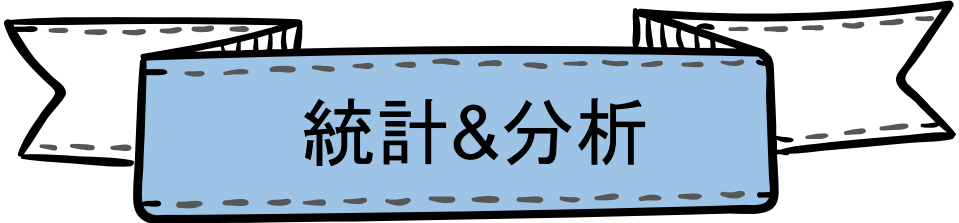
1. **市值**: 一絕對的數值, 並無法直接地呈現股價的變動情況。而股票交易的重點在於賺取買賣的價差, 並非是看一股票價格高/低, 例如: 一股票市值低, 但卻有潛力, 對比一股票市值高而上漲的幅度較低。雖然後者市值較高, 但買賣股票應選擇前者較佳。因此該attribute較不重要。
2. **速動比率**: 速動資產(貨幣資金、短期投資、應收票據、應收帳款等)與流動比率的比例。用來評估一企業的流動資產, 能因應緊急情況, 立刻兌現用來償還負債的能力。在買賣股票時, 價格的變動與償還負債能力較無直接的關係, 因此這個attribute較不重要。



- |           |            |            |            |           |           |           |
|-----------|------------|------------|------------|-----------|-----------|-----------|
| ■ 市值(百萬元) | ■ 收盤價(元)_年 | ■ 本益比      | ■ 股價淨值比    | ■ 股價營收比   | ■ 每股盈餘    | ■ 每股淨值(C) |
| ■ 每股營業額   | ■ ROA-稅後   | ■ ROE-稅後   | ■ 營業利益率    | ■ 邊際利潤    | ■ 常續性稅後淨利 | ■ 總負債/總淨值 |
| ■ 流動比率    | ■ 速動比率     | ■ 存貨週轉率(次) | ■ 應收帳款週轉次數 | ■ 營業利益成長率 | ■ 稅後淨利成長率 | ■ 加權平均股數  |

統計&分析

較為重要	較不重要
股價營收比、收盤價(元)_年、股價淨值比、稅後淨利成長率	市值(百萬元)、ROA－稅後、總負債/總淨值

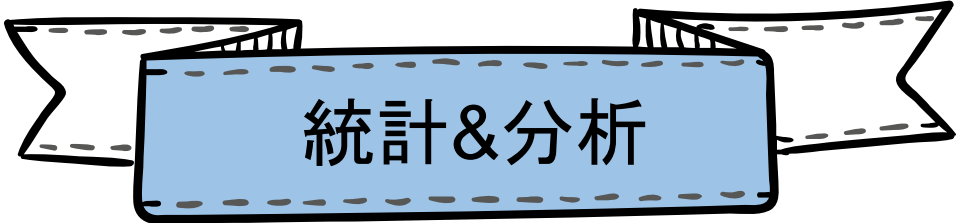


## 統計&分析

較重要的attribute:

1. **股價營收比** = 股價/每股營收: 代表目前公司市值是總營收的n倍, 即表示投資人願意用多少金額購買此公司每一元的營收。當我們要購買股票時, 股價營收比越少越好, 這樣就可以用相對便宜的價格購得高營收的股票, 是一個重要的attribute!
2. **收盤價**: 可與前一交易日的收盤價格, 或下一交易日的開盤價格做比較, 作為評估該資產在單一交易日的價格變動之方式。分析股票價格長期變化的關鍵性attribute!
3. **股價淨值比** = 股價 / (每股淨值): 願意花幾倍的每股淨值去投資這支股票。
  - 當股價淨值比 < 1: 股價較低, 潛在報酬率高, 可考慮買進
  - 反之(>1): 股價較貴, 潛在報酬率低, 可考慮賣出。在投資股票時, 是一個重要的attribute!
4. **稅後淨利成長率** = (當年稅後淨利 - 前一年稅後淨利) ÷ 前一年稅後淨利: 可評估出一公司稅後淨利成長的情況。公司營收高未必稅後淨利成長率也高, 因為可能因為搶市場而削價競爭, 所以這個重要的attribute可以看出企業經營能力和效率。





## 統計&分析

較**不重要**的attribute:

1. **市值**: 一絕對的數值, 並無法直接地呈現股價的變動情況。而股票交易的重點在於賺取買賣的價差, 並非是看一股票價格高/低, 例如: 一股票市值低, 但卻有潛力, 對比一股票市值高而上漲的幅度較低。雖然後者市值較高, 但買賣股票應選擇前者較佳。因此該attribute較不重要。
2. ROA—稅後 = 公司資產運用的效率: 用來衡量一間公司營運績效、賺錢的效率, 但對不同產業類型來說, 靠資產能創造的價值也不大一樣, 所以無法衡量輕資產的行業(例如: 科技、軟體公司), 所以這個attribute的重要性較低。
3. **總負債/總淨值** (負債權益比): 負債權益比高, 對投資人和企業主來說, 利息成本更高、還款壓力更大、資金成本更高、對現金周轉與融資能力要求更高。此attribute與股價的波動較無直接關係, 所以較不重要。

