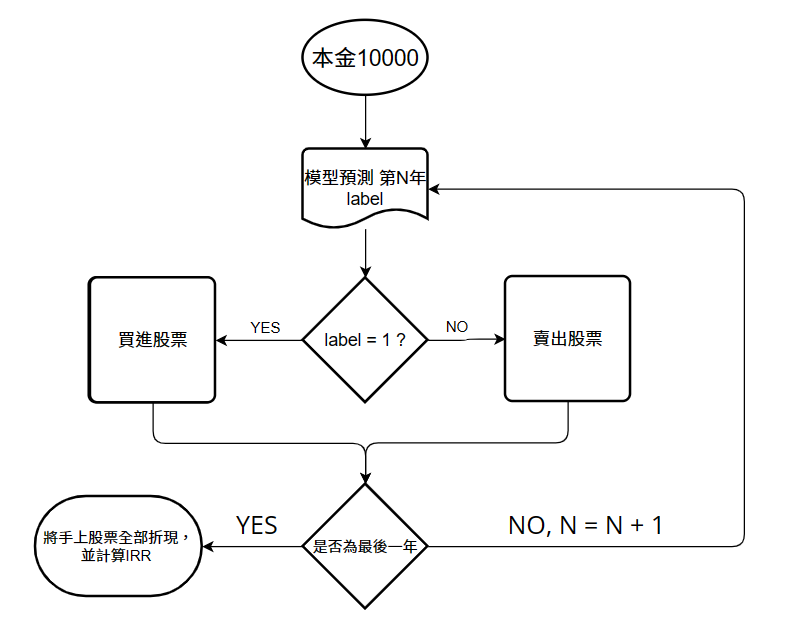
**人工智慧與金融科技 期末報告 第四組**

**Stragety**

假設本金為10000，買進Return\_Labellabel被標記為1的股票，並將現有現金平均買入每支股票

，若股票的Return\_Label被標記為-1，且現在持有該股票，則全數賣出；



**KNN**

Step1 [ 資料前置處理 ]

1. 因為最後一年資料的Reurn為-100，因此將其去除

2. 訓練model的時候只需要特徵值所以也要把一些不需要的欄位去除，比如股票代碼、簡稱

3. 將資料的label取出

4. 使用minmax的方式正規化，將所有資料特徵範圍調整至0~1之間，減少離群值的影響

5. 依Temporal Validation將資料分成TV1~TV11

Step2 [ GA挑選資料特徵與模型參數 ]

染色體:

1. 16個特徵，每個特徵1bit，表示是否挑選此特徵

2. K值: 4bits，decode + 1 -> 1~16

◆ Total: 20 bits

GA參數:

交配機率= 0.7

變異機率= 0.001

種群大小= 80

迭代次數= 30

Early Stopping= 3 (When Best Fitness is equal to Average Fitness three times)

適應度:

Model Accuracy

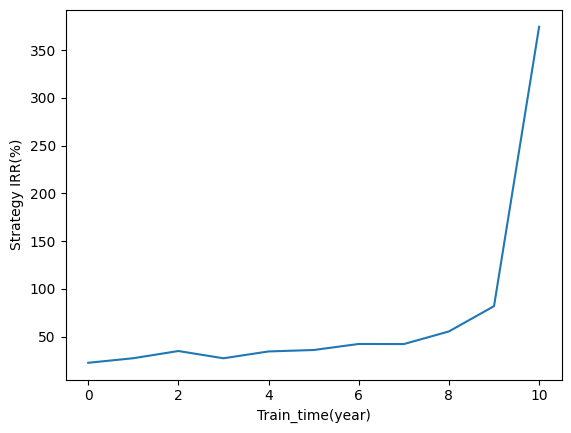
Step3 [ 訓練模型與驗證成效 ]

1. 將GA挑擇之特徵與參數代入模型訓練

2. 以test資料測試，產生Accuracy與預測結果

3. 將預測結果帶入策略，計算出ROI、IRR來驗證其成效

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TV | K | Selected Attributes | | | | | | | | | | | | | | | | Best accuracy | IRR | | |
|  |  | 市值(百萬元) | 收盤價(元)\_年 | 本益比 | 股價淨值比 | 股價營收比 | M淨值報酬率─稅後 | 資產報酬率ROA | 營業利益率OPM | 利潤邊際NPM | 負債/淨值比 | M流動比率 | M速動比率 | M存貨週轉率 (次) | M應收帳款週轉次 | M營業利益成長率 | M稅後淨利成長率 |  | |  |
| 1 | 16 |  | **v** | **v** | **v** |  |  | **v** |  | **v** |  |  |  |  | **v** | **v** |  | 0.61875 | 22.64956396618291 | |
| 2 | 16 |  |  | **v** | **v** | **v** |  | **v** |  |  |  |  |  | **v** | **v** | **v** | **v** | 0.63272727 | 27.41973125812331 | |
| 3 | 16 |  | **v** |  | **v** | **v** |  | **v** |  |  | **v** | **v** |  |  |  | **v** |  | 0.63 | 34.96264931633914 | |
| 4 | 16 |  | **v** | **v** | **v** | **v** |  | **v** |  |  | **v** | **v** |  |  | **v** | **v** | **v** | 0.62222222 | 27.36684519123202 | |
| 5 | 14 | **v** | **v** | **v** | **v** | **v** | **v** |  |  |  | **v** |  | **v** | **v** | **v** | **v** | **v** | 0.6225 | 34.52575668780009 | |
| 6 | 16 |  | **v** | **v** | **v** | **v** | **v** |  |  |  |  | **v** |  | **v** | **v** | **v** |  | 0.644285714 | 36.02785950923351 | |
| 7 | 16 |  | **v** | **v** | **v** |  | **v** | **v** |  | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** | 0.6525 | 42.35419717047626 | |
| 8 | 16 |  | **v** | **v** | **v** | **v** | **v** |  |  |  | **v** | **v** | **v** |  | **v** |  | **v** | 0.665 | 42.2441149859661 | |
| 9 | 16 |  | **v** |  | **v** |  | **v** | **v** | **v** |  |  | **v** | **v** | **v** |  | **v** | **v** | 0.6875 | 55.47892105793291 | |
| 10 | 16 |  | **v** | **v** | **v** | **v** |  |  |  | **v** |  | **v** | **v** | **v** |  |  |  | 0.725 | 81.9368022143953 | |
| 11 | 16 |  | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** |  |  | **v** |  | **v** | **v** | 0.77 | 374.4400000000005 | |
| 統計 |  | 1 | 10 | 9 | 11 | 8 | 6 | 7 | 2 | 4 | 6 | 6 | 5 | 7 | 7 | 9 | 7 |  |  | |



**Decision Tree**

Step1 [ 資料前置處理 ]

1. 因為最後一年資料的Reurn為-100，因此將其去除

2. 訓練model的時候只需要特徵值所以也要把一些不需要的欄位去除，比如股票代碼、簡稱

3. 將資料的label取出

4. 使用minmax的方式正規化，將所有資料特徵範圍調整至0~1之間，減少離群值的影響

5. 依Temporal Validation將資料分成TV1~TV11

Step2 [ GA挑選資料特徵 ]

染色體:

16個特徵，每個特徵1bit，表示是否挑選此特徵

◆ Total: 16 bits

GA參數:

交配機率= 0.7

變異機率= 0.001

種群大小= 80

迭代次數= 30

Early Stopping= 3 (When Best Fitness is equal to Average Fitness three times)

適應度:

Model Accuracy

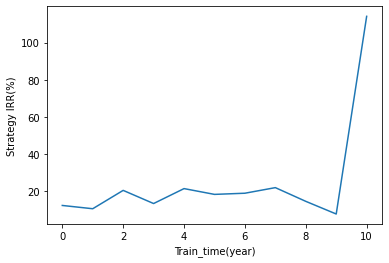
Step3 [ 訓練模型與驗證成效 ]

1. 將GA挑擇之特徵代入模型訓練

2. 以test資料測試，產生Accuracy與預測結果

3. 將預測結果帶入策略，計算出ROI、IRR來驗證其成效

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TV | Selected Attributes | | | | | | | | | | | | | | | | Best accuracy | IRR | | |
|  | 市值(百萬元) | 收盤價(元)\_年 | 本益比 | 股價淨值比 | 股價營收比 | M淨值報酬率─稅後 | 資產報酬率ROA | 營業利益率OPM | 利潤邊際NPM | 負債/淨值比 | M流動比率 | M速動比率 | M存貨週轉率 (次) | M應收帳款週轉次 | M營業利益成長率 | M稅後淨利成長率 |  | |  |
| 1 |  | **v** | **v** | **v** | **v** | **v** | **v** |  | **v** |  | **v** | **v** | **v** | **v** | **v** |  | 0.5558333333 | 12.426874318369752 | |
| 2 |  | **v** | **v** | **v** | **v** |  | **v** | **v** | **v** |  | **v** |  | **v** | **v** | **v** | **v** | 0.5572727273 | 10.662664857865423 | |
| 3 |  | **v** | **v** |  |  | **v** | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** | 0.5505 | 20.53637810406157 | |
| 4 | **v** | **v** |  | **v** |  | **v** | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** |  | 0.55 | 13.46018754831151 | |
| 5 |  | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** |  | **v** |  | **v** | **v** | **v** |  | 0.583125 | 21.49549465658971 | |
| 6 | **v** | **v** | **v** | **v** |  |  | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** | **v** | **v** | 0.5835714285 | 18.384754154464655 | |
| 7 |  | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** |  |  |  | **v** | 0.5808333333 | 19.009436707939663 | |
| 8 |  | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** |  |  | **v** | **v** | **v** | **v** | 0.587 | 22.01741843238363 | |
| 9 | **v** | **v** |  | **v** | **v** | **v** | **v** |  | **v** |  |  |  | **v** |  | **v** | **v** | 0.59375 | 14.639070036983615 | |
| 10 |  | **v** |  | **v** |  | **v** | **v** |  |  | **v** |  | **v** | **v** |  | **v** | **v** | 0.608333333 | 7.823930553472214 | |
| 11 |  |  | **v** | **v** |  |  | **v** |  |  |  | **v** | **v** | **v** |  | **v** | **v** | 0.6525 | 114.18999999999997 | |
| 統計 | 3 | 10 | 8 | 9 | 6 | 8 | 10 | 7 | 8 | 5 | 8 | 6 | 10 | 7 | 10 | 8 |  |  | |



**Random Forest**

Step1 [ 資料前置處理 ]

1. 因為最後一年資料的Reurn為-100，因此將其去除

2. 訓練model的時候只需要特徵值所以也要把一些不需要的欄位去除，比如股票代碼、簡稱

3. 將資料的label取出

4. 使用minmax的方式正規化，將所有資料特徵範圍調整至0~1之間，減少離群值的影響

5. 依Temporal Validation將資料分成TV1~TV11

Step2 [ GA挑選資料特徵 ]

染色體:

16個特徵，每個特徵1bit，表示是否挑選此特徵

◆ Total: 16 bits

GA參數:

交配機率= 0.7

變異機率= 0.001

種群大小= 80

迭代次數= 30

Early Stopping= 3 (When Best Fitness is equal to Average Fitness three times)

適應度:

Model Accuracy

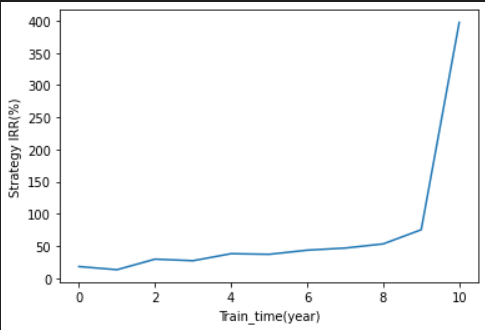
Step3 [ 訓練模型與驗證成效 ]

1. 將GA挑擇之特徵代入模型訓練

2. 以test資料測試，產生Accuracy與預測結果

3. 將預測結果帶入策略，計算出ROI、IRR來驗證其成效

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TV | Selected Attributes | | | | | | | | | | | | | | | | Best accuracy | IRR | | |
|  | 市值(百萬元) | 收盤價(元)\_年 | 本益比 | 股價淨值比 | 股價營收比 | M淨值報酬率─稅後 | 資產報酬率ROA | 營業利益率OPM | 利潤邊際NPM | 負債/淨值比 | M流動比率 | M速動比率 | M存貨週轉率 (次) | M應收帳款週轉次 | M營業利益成長率 | M稅後淨利成長率 |  | |  |
| 1 |  | **v** |  | **v** |  |  | **v** |  | **v** | **v** | **v** |  |  | **v** |  |  | 0.58125 | 18.26700907245762 | |
| 2 |  | **v** |  | **v** |  |  | **v** |  | **v** | **v** |  | **v** |  |  | **v** | **v** | 0.5736363636 | 13.353928077787213 | |
| 3 |  | **v** | **v** | **v** | **v** |  |  |  | **v** | **v** |  |  | **v** | **v** | **v** | **v** | 0.583 | 29.76318319368092 | |
| 4 |  | **v** |  | **v** |  |  |  |  | **v** | **v** |  |  | **v** | **v** |  | **v** | 0.5977777777 | 27.324266225691886 | |
| 5 | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** |  | **v** |  | **v** |  |  | **v** | 0.63125 | 38.28116161083442 | |
| 6 | **v** |  | **v** | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** |  | **v** | **v** |  | **v** | 0.6307142857 | 37.28392615536913 | |
| 7 | **v** | **v** | **v** | **v** | **v** | **v** |  | **v** |  |  | **v** | **v** | **v** |  | **v** | **v** | 0.63 | 43.673220708169104 | |
| 8 |  | **v** |  | **v** | **v** | **v** |  |  | **v** | **v** | **v** | **v** |  | **v** | **v** |  | 0.656 | 46.99772694311613 | |
| 9 | **v** | **v** | **v** | **v** | **v** |  | **v** |  | **v** |  | **v** | **v** | **v** | **v** | **v** | **v** | 0.66625 | 53.4433153388179 | |
| 10 |  |  | **v** | **v** | **v** |  |  |  | **v** | **v** | **v** | **v** | **v** | **v** | **v** |  | 0.685 | 75.44229820656135 | |
| 11 |  |  | **v** | **v** |  | **v** |  |  |  | **v** |  | **v** | **v** |  | **v** | **v** | 0.7525 | 397.46000000000004 | |
| 統計 | 4 | 8 | 7 | 11 | 7 | 5 | 4 | 3 | 9 | 8 | 7 | 6 | 8 | 7 | 7 | 8 |  |  | |



**XGBoost**

Step1 [ 資料前置處理 ]

1. 因為最後一年資料的Reurn為-100，因此將其去除

2. 訓練model的時候只需要特徵值所以也要把一些不需要的欄位去除，比如股票代碼、簡稱

3. 將資料的label取出

4. 使用minmax的方式正規化，將所有資料特徵範圍調整至0~1之間，減少離群值的影響

5. 依Temporal Validation將資料分成TV1~TV11

Step2 [ GA挑選資料特徵與模型參數 ]

染色體:

1. 16個特徵，每個特徵1bit，表示是否挑選此特徵

2. 參數設定: 選擇Tree Booster其中7種我認為影響最大之參數進行GA優化

(1) learning\_rate:

4bits，0.01 + decode \* 0.05 -> 0.01~0.76

控制模型中梯度下降的速度，也稱為步長

根據公式new\_weight = old\_weight - learning\_rate \* gradient

learning\_rate設置過小，會導致模型學習收斂較慢，容易過擬合 設置過大時，會導致模型學習過快，無法收斂，容易震盪或發散

(2) eta:

4bits，0.01 + decode \* 0.01 -> 0.01 ~ 0.16

為了防止過擬合，更新過程中用到的收縮步長

在每次梯度提升後，計算新特徵的權重，通過縮減權重使梯度提升過程更保守。

(3) max\_depth:

3bits，2 + decode -> 2 ~ 9

樹的最大深度

深度越大，對數據的擬合程度越高（過擬合程度也越高）

(4) min\_child\_weight:

3bits，1 + decode -> 1 ~ 8

子節點中最小的樣本權重和

若一個子節點的樣本權重和小於此數值則拆分過程結束

(5) gamma:

4bits，0.1 + decode \* 0.5 -> 0.1 ~ 7.6

模型對於一個節點的劃分只有在其損失函數得到的結果大於0的情況下才進行， 而gamma給定了所需的最低損失函數值

(6) subsample:

3bits，0.1 + decode \* 0.1 -> 0.1 ~ 0.8

用於訓練模型的子樣本佔整個樣本集合的比例

若設置為0.5則意味著XGBoost將隨機從整個樣本集合中抽取出50%的子樣本建立 樹模型，能防止過擬合

(7) colsample\_bytree:

3bits，0.1 + decode \* 0.1 -> 0.1 ~ 0.8

在建立樹時，對特徵隨機採樣的比例

◆ Total: 40 bits

GA參數:

交配機率= 0.7

變異機率= 0.001

種群大小= 80

迭代次數= 30

Early Stopping= 3 (When Best Fitness is equal to Average Fitness three times)

適應度:

Model Accuracy

Step3 [ 訓練模型與驗證成效 ]

1. 將GA挑擇之特徵與參數代入模型訓練

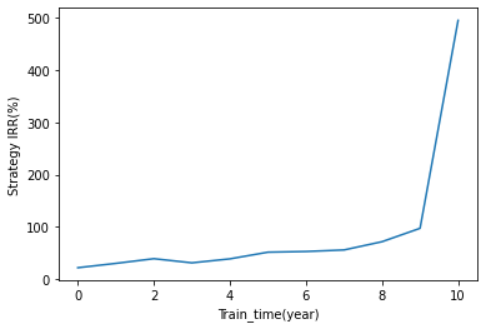
2. 以test資料測試，產生Accuracy與預測結果

3. 將預測結果帶入策略，計算出ROI、IRR來驗證其成效

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TV | Selected Attributes | | | | | | | | | | | | | | | | Best accuracy | IRR |
|  | 市值(百萬元) | 收盤價(元)\_年 | 本益比 | 股價淨值比 | 股價營收比 | M淨值報酬率─稅後 | 資產報酬率ROA | 營業利益率OPM | 利潤邊際NPM | 負債/淨值比 | M流動比率 | M速動比率 | M存貨週轉率 (次) | M應收帳款週轉次 | M營業利益成長率 | M稅後淨利成長率 |  |  |
| 1 |  | **v** | **v** | **v** |  |  | **v** | **v** |  |  |  |  |  | **v** | **v** |  | 0.6233333333 | 21.92691322312641 |
| 2 |  | **v** | **v** | **v** | **v** |  |  | **v** | **v** | **v** |  |  |  |  | **v** | **v** | 0.6322727272 | 30.22581944886726 |
| 3 |  | **v** | **v** | **v** |  |  |  | **v** | **v** | **v** | **v** | **v** | **v** | **v** |  |  | 0.6415 | 39.18513988831804 |
| 4 |  | **v** |  | **v** |  | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** | 0.6344444445 | 31.2587155685228 |
| 5 |  |  | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** | 0.64125 | 38.85367451509338 |
| 6 | **v** | **v** | **v** | **v** |  |  |  | **v** | **v** | **v** | **v** |  | **v** |  |  |  | 0.6578571428 | 51.62737098575700 |
| 7 |  |  | **v** | **v** | **v** | **v** |  | **v** | **v** |  | **v** | **v** | **v** |  | **v** | **v** | 0.6716666666 | 52.95174310639405 |
| 8 |  | **v** | **v** | **v** |  |  |  |  | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** | 0.692 | 55.87401193234977 |
| 9 |  | **v** |  |  | **v** |  | **v** | **v** |  | **v** | **v** |  | **v** | **v** | **v** | **v** | 0.69375 | 71.67321222200962 |
| 10 |  | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** |  | **v** |  | **v** |  | **v** | **v** | 0.7316666667 | 97.27899026505584 |
| 11 | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** |  |  | **v** |  | 0.8275 | 494.7700000000000 |
| 統計 | 2 | 9 | 9 | 10 | 5 | 5 | 6 | 9 | 9 | 8 | 9 | 5 | 8 | 5 | 9 | 7 |  |  |

XGBoost參數

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| TV | learning\_rate | eta | max\_depth | min\_child\_weight | gamma | subsample | colsample\_bytree |
| 1 | 0.36 | 0.08 | 9 | 6 | 2.6 | 0.2 | 0.7 |
| 2 | 0.06 | 0.12 | 9 | 6 | 7.1 | 0.6 | 0.1 |
| 3 | 0.01 | 0.16 | 3 | 7 | 7.6 | 0.7 | 0.2 |
| 4 | 0.01 | 0.07 | 8 | 5 | 7.6 | 0.4 | 0.8 |
| 5 | 0.06 | 0.08 | 3 | 7 | 7.6 | 0.8 | 0.6 |
| 6 | 0.06 | 0.16 | 3 | 7 | 7.6 | 0.6 | 0.3 |
| 7 | 0.01 | 0.16 | 3 | 5 | 7.1 | 0.5 | 0.4 |
| 8 | 0.06 | 0.14 | 5 | 5 | 7.6 | 0.6 | 0.8 |
| 9 | 0.19 | 0.16 | 2 | 3 | 7.6 | 0.7 | 0.2 |
| 10 | 0.06 | 0.15 | 2 | 6 | 7.1 | 0.8 | 0.4 |
| 11 | 0.06 | 0.02 | 3 | 7 | 7.6 | 0.3 | 0.2 |



1. **統計與分析**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | Selected Attributes | | | | | | | | | | | | | | | |
|  | 市值(百萬元) | 收盤價(元)\_年 | 本益比 | 股價淨值比 | 股價營收比 | M淨值報酬率─稅後 | 資產報酬率ROA | 營業利益率OPM | 利潤邊際NPM | 負債/淨值比 | M流動比率 | M速動比率 | M存貨週轉率 (次) | M應收帳款週轉次 | M營業利益成長率 | M稅後淨利成長率 |
| KNN | 1 | 10 | 9 | 11 | 8 | 6 | 7 | 2 | 4 | 6 | 6 | 5 | 7 | 7 | 9 | 7 |
| Decision Tree | 3 | 10 | 8 | 9 | 6 | 8 | 10 | 7 | 8 | 5 | 8 | 6 | 10 | 7 | 10 | 8 |
| Random Forest | 4 | 8 | 7 | 11 | 7 | 5 | 4 | 3 | 9 | 8 | 7 | 6 | 8 | 7 | 7 | 8 |
| XGBoost | 2 | 9 | 9 | 10 | 5 | 5 | 6 | 9 | 9 | 8 | 9 | 5 | 8 | 5 | 9 | 7 |
| 統計 | 10 | 37 | 33 | 41 | 26 | 24 | 27 | 21 | 30 | 27 | 30 | 22 | 33 | 26 | 35 | 30 |

1. **統整與分析**

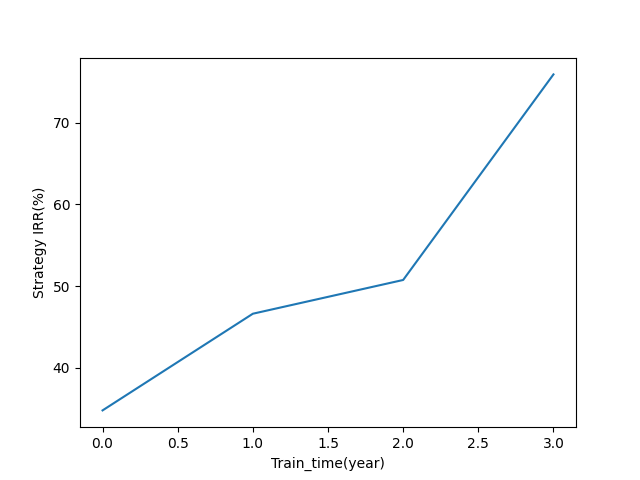
根據圖表統計，我們可以將attribute做簡單的分類：

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

我們認為的可能原因是：

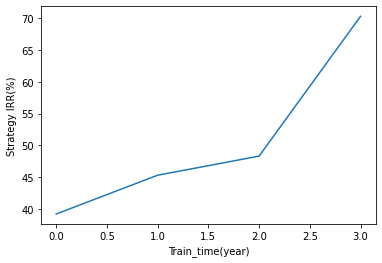
1. ：願意花幾倍的每股淨值去投資這支股票。當股價淨值比＜1：股價較低，潛在報酬率高，可考慮買進；反之(>1)：股價較貴，潛在報酬率低，可考慮賣出。在投資股票時，是一個重要的attribute！
2. 收盤價：可與前一交易日的收盤價格，或下一交易日的開盤價格做比較，作為評估該資產在單一交易日的價格變動之方式。分析股票價格長期變化的關鍵性attribute！
3. 營業利益成長率 ＝（當期的營業利益－去年同期的營業利益）／去年同期的營業利益\* 100%：可得知與去年相比，此公司的營收狀況是進步或是退步。當要買賣股票時，若我們發現有一間公司連續5年營業利益成長率>0，就代表此公司正在成長當中，所以我們可以考慮買入，是一個重要的attribute！
4. 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。
5. 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。
6. **Testing result**
   1. KNN

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TV | K | Selected Attributes | | | | | | | | | | | | | | | | | | | | | Best accuracy | IRR | |
|  |  | 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 |  |  |
| 1 | 16 | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** |  | **v** | **v** |  |  |  | **v** |  | **v** | **v** |  | **v** | **v** | 0.565 | 39.19615296140393 |
| 2 | 6 |  | **v** | **v** | **v** | **v** |  | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** | **v** |  |  | **v** | **v** | **v** |  | 0.58375 | 45.29834197705986 |
| 3 | 8 |  | **v** | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** |  |  | **v** | **v** | **v** |  | **v** | **v** |  | 0.59 | 48.30537143539644 |
| 4 | 6 |  | **v** | **v** | **v** | **v** | **v** | **v** |  |  |  | **v** | **v** | **v** | **v** |  |  | **v** |  | **v** | **v** |  | 0.5825 | 70.29679973505081 |
| 5 | 15 | **v** | **v** |  | **v** | **v** |  | **v** | **v** |  |  |  | **v** | **v** | **v** | **v** |  | **v** | **v** | **v** |  |  | 0.635 |  |
| 統計 |  | 2 | 5 | 4 | 5 | 5 | 2 | 5 | 3 | 1 | 3 | 4 | 4 | 3 | 3 | 4 | 1 | 4 | 3 | 4 | 4 | 1 |  |  |



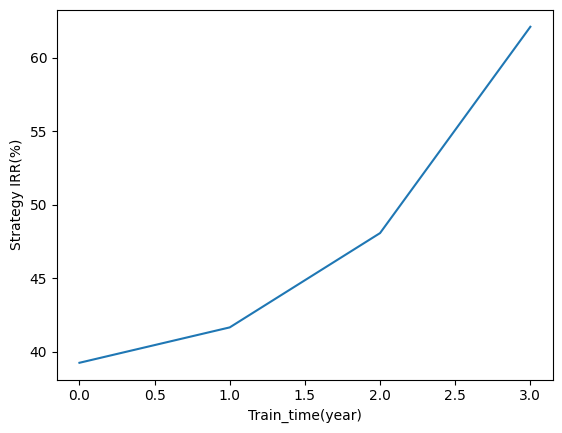
* 1. Decision tree

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TV | Selected Attributes | | | | | | | | | | | | | | | | | | | | | Best accuracy | IRR | |
|  | 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 |  |  |
| 1 |  | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** |  |  | **v** |  | **v** | **v** | **v** | **v** | **v** | 0.565 | 39.19615296140393 |
| 2 | **v** | **v** | **v** | **v** | **v** | **v** |  |  |  |  | **v** |  | **v** |  | **v** | **v** | **v** | **v** | **v** | **v** |  | 0.58375 | 45.29834197705986 |
| 3 | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** |  | **v** | **v** |  | **v** |  | **v** | **v** | 0.59 | 48.30537143539644 |
| 4 |  | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** |  |  | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** |  | 0.5825 | 70.29679973505081 |
| 5 | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** |  | **v** | **v** | **v** |  | **v** | **v** | 0.635 |  |
| 統計 | 3 | 5 | 4 | 4 | 4 | 5 | 3 | 4 | 3 | 4 | 4 | 2 | 4 | 2 | 4 | 4 | 4 | 4 | 3 | 5 | 3 |  |  |



* 1. Random forest

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TV | Selected Attributes | | | | | | | | | | | | | | | | | | | | | Best accuracy | IRR | |
|  | 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 |  |  |
| 1 | **v** | **v** |  | **v** | **v** | **v** |  |  | **v** | **v** |  | **v** |  |  |  | **v** |  |  | **v** | **v** | **v** | 0.599 | 39.25842861935986 |
| 2 | **v** | **v** |  | **v** | **v** | **v** |  | **v** |  |  |  | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** | **v** | **v** | 0.59125 | 41.66467938012597 |
| 3 |  | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** |  |  |  |  |  | **v** | **v** | **v** | **v** |  | **v** | **v** | 0.618333333 | 48.076172135740514 |
| 4 |  | **v** |  |  | **v** | **v** | **v** | **v** |  |  | **v** | **v** |  | **v** | **v** |  | **v** | **v** |  | **v** |  | 0.595 | 62.101819854065795 |
| 5 |  | **v** | **v** | **v** | **v** | **v** | **v** | **v** |  |  |  |  |  |  |  | **v** | **v** | **v** | **v** |  |  | 0.6 |  |
| 統計 | 2 | 5 | 2 | 4 | 5 | 5 | 2 | 4 | 2 | 1 | 1 | 3 | 1 | 1 | 3 | 4 | 4 | 4 | 3 | 4 | 3 |  |  |

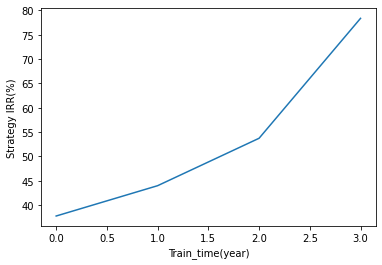


* 1. XGBoost

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TV | Selected Attributes | | | | | | | | | | | | | | | | | | | | | Best accuracy | IRR | |
|  | 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 |  |  |
| 1 | **v** |  | **v** |  |  | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** |  |  | **v** | **v** | **v** | **v** | **v** | **v** | 0.587 | 37.77593068338636 |
| 2 | **v** |  | **v** | **v** | **v** |  |  | **v** |  | **v** | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** | **v** | **v** | **v** | 0.60625 | 44.00320687589756 |
| 3 | **v** | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** |  |  |  |  | **v** | **v** | **v** |  | **v** |  | **v** | **v** | 0.625 | 53.74145633006109 |
| 4 |  | **v** |  | **v** | **v** | **v** | **v** | **v** | **v** |  |  |  |  | **v** | **v** | **v** | **v** | **v** |  | **v** | **v** | 0.6125 | 78.37040113202637 |
| 5 |  |  | **v** | **v** | **v** |  |  | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** | **v** |  | **v** |  |  |  | 0.625 |  |
| 統計 | 3 | 2 | 4 | 4 | 4 | 3 | 2 | 4 | 4 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 3 | 5 | 2 | 4 | 4 |  |  |

XGBoost 參數

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| TV | learning\_rate | eta | max\_depth | min\_child\_weight | gamma | subsample | colsample\_bytree |
| 1 | 0.16 | 0.16 | 7 | 2 | 1.6 | 0.6 | 0.4 |
| 2 | 0.06 | 0.16 | 9 | 6 | 1.9 | 0.8 | 1.6 |
| 3 | 0.16 | 0.14 | 2 | 5 | 1.9 | 0.8 | 0.4 |
| 4 | 0.01 | 0.16 | 9 | 2 | 7.6 | 0.6 | 0.6 |
| 5 | 0.76 | 0.06 | 7 | 1 | 5.6 | 0.8 | 0.5 |



* 1. 分析與統計

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | Selected Attributes | | | | | | | | | | | | | | | | | | | | |
|  | 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 |
| KNN | 2 | 5 | 4 | 5 | 5 | 2 | 5 | 3 | 1 | 3 | 4 | 4 | 3 | 3 | 4 | 1 | 4 | 3 | 4 | 4 | 1 |
| Decision Tree | 3 | 5 | 4 | 4 | 4 | 5 | 3 | 4 | 3 | 4 | 4 | 2 | 4 | 2 | 4 | 4 | 4 | 4 | 3 | 5 | 3 |
| Random Forest | 2 | 5 | 2 | 4 | 5 | 5 | 2 | 4 | 2 | 1 | 1 | 3 | 1 | 1 | 3 | 4 | 4 | 4 | 3 | 4 | 3 |
| XGBoost | 3 | 2 | 4 | 4 | 4 | 3 | 2 | 4 | 4 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 3 | 5 | 2 | 4 | 4 |
| 統計 | 10 | 17 | 14 | 17 | 18 | 15 | 12 | 15 | 10 | 11 | 12 | 12 | 11 | 10 | 15 | 13 | 15 | 16 | 12 | 17 | 11 |

根據圖表統計，我們可以將attribute做簡單的分類：

|  |  |
| --- | --- |
| **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 |

我們認為的可能原因是：

1. ：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. ：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。
5. 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。
6. 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的這重要性較低。
7. 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。
8. **報告問題補充：本益比為負時，該如何看待它？**

每股盈餘為負時，本益比為負數。代表意涵：「評估過了幾年，會把投資的本金全數虧光」。我們在投資時，會注意到的是這間公司除了盈餘外的價值，所以應當使用其他更適合的企業估值方式。

1. **心得**
   1. A1095502 沈品豪：在這堂課之前，我其實對人工智慧不是很了解，也沒有實作過，更沒有想過可以融合到金融的部分，就像我一開始都不知道原來買股票有這麼多的策略可以來選擇，直到上了老師的課，還有透過期中和期末專題的實作，我才對人工智慧和金融的部份有更進一步的了解，我希望我有一天可以再多花一點時間來進一步的研究，這樣才可以賺更多的錢。
   2. A1095517 黃泓茗：不管是期中專題或期末專題，都讓我學習到很多東西，而且這金融與人工智慧相關的技術，在網路上較少相關資源或完整的股票資料可以自己學習或練習，很慶幸有這樣難得的機會可以入門這方面領域，希望未來有一天可以真正運用到這些知識，幫助我賺錢。
   3. A1095521 鄭惠心：其實Decision tree實作的部分都跟CNN、Unet等其他model簡單許多，但更重要的問題是要怎麼提升它的正確率，一開始我只是單純更改它的參數設定，或者乘上每個特徵值的權重再進行訓練但效果都沒有很好，在跟組員一起討論過後我們決定用GA來特徵挑選，果然正確率比之前有提升了!可惜的是我們組的GA本來是想要跟老師的論文裡面的方式，以GA來決定每個特徵值的權重，但最後沒有實做出來。希望以後可以更能善用自己學到的東西!
   4. A1095527 廖習驊：我主要負責的部分是統計與分析，以及部份的程式實作。在執行結束過後，首先得到的是一群資料，並無法直接看出我們執行結果，在經過一陣時間的思考與嘗試，我決定用以上報告內容的表格去統整與歸納，並且將每個TV用GA選到的attribute標記，然後在最後一欄呈現統計內容。但是這樣還是不夠，我進一步的將統計結果繪製圖表，以視覺化的方式就可以很容易的地看出哪個attribute比重要（長條圖較高者），又是哪個attribute比較不重要（長條圖較低者）。

對於我們獲得的比較重要與比較不重要的屬性，我想再進一步的探討，但是我之前並沒有太多金融相關的知識，所以我必須先去研究每個attribute個別代表的意思：

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。
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。
3. 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！
4. 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。
5. 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。
6. 淨值報酬率─稅後=稅後純益/股東權益：評估股東權益投資報酬率的指標，代表公司使用資產淨值產出獲利的能力。
7. 資產報酬率ROA=稅後淨利/平均總資產×100%：公司的整個報酬率，數值越高代表公司資產報酬運用的效率越好，擁有較高的賺錢能力。
8. 營業利益率OPM=(營業收入－營業成本－營業費用)/營業收入：公司每生產一元的營收，所能獲得的利益，是評評估一公司獲利能力的指標。
9. 利潤邊際NPM：評估一間公司賺錢能力的比率，數值越高就表示管控成本的能力越佳。
10. 負債/淨值比：數值比高，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。
11. 流動比率=流動資產/(流動負債 )×100%：公司每一元的流動負債，能夠用幾元的流動資產去還債。
12. 存貨週轉率 (次)=營業成本/平均存貨：評估存貨的周轉速度，判斷存貨的流動性及存資金是否合理。
13. 應收帳款週轉次：呈現公司一年內收回欠款的次數，可評估公司收回帳款或客戶欠款的效率。
14. 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。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。
15. 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，所以可以看出企業經營能力和效率。

在研究完這些金融專有名詞後，我嘗試與我們GA選到的attribute作分析，結果都非常合理，training和testing的結果甚至是有些交集的，表示我們的實作成果具有一定的可信度。

* 1. A1095557 陳柏安：這次期末專題是我第一次嘗試使用GA來決定Machine Learning方法的參數，連帶用GA挑選特徵一起，準確率較原本高出了10幾%，這次的嘗試讓我學了很多。這學期的課程不管是在金融知識抑或是程式方面都獲益良多。
  2. A1095560 歐津萍：在這次期末專題中，我分配到的是決策樹的方面，一開始我其實對實作的方向感到很迷茫，但經過上網搜尋及和同學討論的過程中，我學習到了很多東西，也對金融選股及資料模型多了很多理解，雖然在這次的專題中我的助益不大，但和其他厲害的組員一起分工合作後，我認為我收穫了一次很棒的歷程。
  3. A1095565 張庭瑋：經過這堂課後，讓我對於這個領域有了更不一樣的認識，也藉由期中和期末的專題，有了進一步的了解，但有些東西到現在還不是很理解，自己也沒辦法做出來，所以很謝謝我的組員跟我同一組，沒有他們，我這堂課應該就被當掉了。

1. **分工**

|  |  |
| --- | --- |
| GA\_KNN | 黃泓茗、沈品豪、陳柏安、張庭瑋 |
| GA\_DecisionTree、GA\_RandomForest | 鄭惠心、歐津萍、陳柏安 |
| GA\_XGBoost | 陳柏安、黃泓茗、廖習驊 |
| 策略 | 黃泓茗、陳柏安 |
| 報告 | 鄭惠心、沈品豪、廖習驊 |
| Word、PPT | 廖習樺、陳柏安、黃泓茗、鄭惠心、沈品豪、歐津萍 |