

# 利用機器學習建立最佳投資組合

期末報告: 專案主題、方法、產出及結論

第 N 組

許智評、林姝延、賴宇辰、王奕翔、曾國睿

# Agenda

- |   |                 |             |
|---|-----------------|-------------|
| 1 | 專案摘要            | P.3 - P.4   |
| 2 | 基本面研究流程及方法      | P.5 - P.7   |
| 3 | 基本面投資決策、成效及投資建議 | P.8 - P.11  |
| 4 | 技術面研究流程及方法      | P.12        |
| 5 | 技術面投資決策、成效及投資建議 | P.13 - P.15 |
| 6 | 消息面研究流程及方法      | P.16        |
| 7 | 消息面投資決策、成效及投資建議 | P.17 - P.18 |
| 8 | 研究結果            | P.19        |

## 基於機器學習各模型預測結果建立投資組合，進行各模型跨期投報率比較，形成投資建議

專案主題	利用機器學習來進行股價走勢預測，並以 3 種不同方法進行投資成效衡量比較				
研究目標	由各模型預測股票上漲或下跌幅度，比較各模型準確率，並得出最高報酬率投資組合的模型，以提供投資人決策建議				
研究主軸	Step 1 蒐集資料	Step 2 建立模型，加入測資進行測試		Step 3 建立投資組合	Step 4 進行投資
	選取 5 種產業 各產業取 2 間公司	<div><div><ul style="list-style-type: none"><li>● 建立模型</li><li>● 訓練模型</li><li>● 測試模型</li></ul></div><div><div>基本面模型</div><div>技術面模型</div><div>消息面模型</div></div></div>		進行預測， 並根據預測結果形成投資組合	使用模擬股市的 APP，依照投資組合指示進行投資，以驗證模型投資成效
研究產出	<ul style="list-style-type: none"><li>● 產出各模型的投資組合</li><li>● 各模型平均報酬</li><li>● 各模型投資成效比較</li></ul>				
研究限制	<ul style="list-style-type: none"><li>● 專案時程短</li><li>● 限縮產業、公司數量</li></ul>				

## 透過基本面、技術面及消息面之資料及模型得出各自最佳的投資組合

### 投資標的選項

半導體產業		電子零組件業		金融業		通訊產業		生技醫療產業	
2330 台積電	2303 聯電	2353 宏碁	8046 南電	2880 華南金	2881 富邦金	3045 台灣大	2357 華碩	6547 高端疫苗	1789 神隆

### 蒐集資料

### 訓練模型

### 進行預測

### 最佳投資組合

### 模型

#### 基本面

1. 取得財報數據
2. 進行資料清理與分析建立、訓練、測試模型
3. 建立最佳投資組合

#### 技術面

1. 爬蟲抓取資料，並用套件計算 KD、RSI 指標數據
2. 切割數據，建立、訓練、測試模型
3. 得出最佳投資組合

#### 消息面

1. 學習周遭投資人投資技巧
2. 預估實際投資結果
3. 得出最佳投資組合

### 跨模型投資成效比較

# 基本面研究流程及方法

模型	研究方法	模型應用	
基本面	1. 下載各公司股價及綜合財報資訊	多元回歸模型	1. 數據切分:將數據切分為訓練及與測試集。
	2. 資料清理及合併: - 股價資料處理 保留 15 日均價, 並求 3、6、9、12 月平均股價 - 財報資料處理 進行缺失值處理並保留資料型態為整數或浮點數之欄位 - 資料合併: 將股價資料以及財報資訊按照公司及年月進行合併 - 年月調整: 將財報資訊往後遞延一季		2. 股價預測:通過收集多個公司基本面指標作為特徵變量, 建立多元邏輯回歸模型來預測公司未來的股價走勢。
	3. 熱力圖、相關性變數處理: - 刪除與 y 相關性過低的變數 - 避免多重共線性	XGBoost 模型	1. 數據切分:將數據切分為訓練及與測試集
	4. 模型訓練		2. 參數調整:進行參數調整找出最佳參數組合之模型
	5. 模型預測股價, 得模型準確率	SVR 模型	3. 股價預測:根據公司的基本面財務數據, 預測股票價格的趨勢。通過學習歷史數據, 預測公司未來的表現。
			1. 數據切分:將數據切分為訓練及與測試集
			2. 參數調整:進行參數調整找出最佳參數組合之模型
			3. 股價預測:透過公司基本面財務數據集支持向量回歸, 進行公司股價預測。

# 基本面研究流程及方法

模型	研究方法	模型應用
基本面	1. 下載各公司股價及綜合財報資訊	
	2. 資料清理及合併: - 股價資料處理 保留 15 日均價, 並求 3、6、9、12 月平均股價 - 財報資料處理 進行缺失值處理並保留資料型態為整數或浮點數之欄位 - 資料合併: 將股價資料以及財報資訊按照公司及年月進行合併 - 年月調整: 將財報資訊往後遞延一季	多元回歸模型  1. 數據切分: 將數據切分為訓練及與測試集。 2. 股價預測: 通過收集多個公司基本面指標作為特徵變量, 建立多元邏輯回歸模型來預測公司未來的股價走勢。
	3. 熱力圖、相關性變數處理: - 刪除與 y 相關性過低的變數 - 避免多重共線性	XGBoost 模型  1. 數據切分: 將數據切分為訓練及與測試集 2. 參數調整: 進行參數調整找出最佳參數組合之模型 3. 股價預測: 根據公司的基本面財務數據, 預測股票價格的趨勢。通過學習歷史數據, 預測公司未來的表現。
	4. 模型訓練	SVR 模型  1. 數據切分: 將數據切分為訓練及與測試集 2. 參數調整: 進行參數調整找出最佳參數組合之模型 3. 股價預測: 透過公司基本面財務數據集支持向量回歸, 進行公司股價預測。
	5. 模型預測股價, 得模型準確率	

# 基本面研究流程及方法

模型	研究方法	模型應用	
基本面	1. 下載各公司股價及綜合財報資訊	多元回歸模型	1. 數據切分:將數據切分為訓練及與測試集。
	2. 資料清理及合併: - 股價資料處理 保留 15 日均價, 並求 3、6、9、12 月平均股價 - 財報資料處理 進行缺失值處理並保留資料型態為整數或浮點數之欄位 - 資料合併: 將股價資料以及財報資訊按照公司及年月進行合併 - 年月調整: 將財報資訊往後遞延一季		2. 股價預測:通過收集多個公司基本面指標作為特徵變量, 建立多元邏輯回歸模型來預測公司未來的股價走勢。
	3. 熱力圖、相關性變數處理: - 刪除與 y 相關性過低的變數 - 避免多重共線性	XGBoost 模型	1. 數據切分:將數據切分為訓練及與測試集
	4. 模型訓練		2. 參數調整:進行參數調整找出最佳參數組合之模型
	5. 模型預測股價, 得模型準確率	SVR 模型	3. 股價預測:根據公司的基本面財務數據, 預測股票價格的趨勢。通過學習歷史數據, 預測公司未來的表現。
		1. 數據切分:將數據切分為訓練及與測試集	
		2. 參數調整:進行參數調整找出最佳參數組合之模型	
		3. 股價預測:透過公司基本面財務數據集支持向量回歸, 進行公司股價預測。	

基本面投資決策

Regression 模型										
公司	台積電 (2330)	聯電 (2303)	宏碁 (2353)	南電 (8046)	華南金 (2880)	富邦金 (2881)	台灣大 (3045)	華碩 (2357)	高端疫苗 (6547)	神隆 (1789)
3 月 平均股價	519.04	51.47	26.03	256.15	22.54	58.45	97.74	276.54	59.77	26.08
6 月 預測股價	502.74	31.60	79.79	54.49	5.43	408.88	53.77	206.91	34.11	31.94
3~6 月 股價漲幅	-3.14	-38.60	206.48	-78.73	-75.91	599.58	-44.99	-25.18	-42.94	22.47



公司	3~6 月 股價漲幅	投資權重	2023/03/1 開盤價	2023/06/02 收盤價	買進股數	投入資金	最終金額
富邦金 (2881)	599.58	0.72	59.2	60.6	1216	71,987.2	73,689.6
宏碁 (2353)	206.48	0.25	25.4	33.95	984	24,993.6	33,406.8
神隆 (1789)	22.47	0.03	26.3	26.25	114	2,998.2	2,992.5
總計						99,979	110,088.9
年化報酬率						47.01 %	



基本面投資決策

XGBoost 模型										
公司	台積電 (2330)	聯電 (2303)	宏碁 (2353)	南電 (8046)	華南金 (2880)	富邦金 (2881)	台灣大 (3045)	華碩 (2357)	高端疫苗 (6547)	神隆 (1789)
3 月 平均股價	519.04	51.47	26.03	256.15	22.54	58.45	97.74	276.54	59.77	26.08
6 月 預測股價	100.94	57.40	17.07	24.57	11.73	61.57	58.43	78.09	45.09	42.73
3~6 月 股價漲幅	-80.55	11.53	-34.45	-78.70	-47.95	5.35	-40.22	-71.76	-24.57	63.82



公司	3~6 月 股價漲幅	投資權重	2023/03/1 開盤價	2023/06/02 收盤價	買進股數	投入資金	最終金額
神隆 (1789)	63.82	0.79	26.3	26.25	3004	79,005	78,855
聯電 (2303)	11.53	0.14	49	51.4	286	14,014	14,700.4
富邦金 (2881)	5.35	0.07	0.07	60.6	117	6,926	7,090.2
總計						99,945	100,645.6
年化報酬率						2.83 %	

## 基本投資建議:模型重要變數

	Regression 模型		XGBoost 模型
1	每股稅前淨利 8		稅前純益 / 實收資本
2	稅前純益 / 實收資本 3		當期所得稅資產－流動
3	有息負債利率		普通股股本
4	每股現金流量		淨值/資產
5	ROA(C)稅前息前折舊前 13		採權益法之長期股權投資
6	利息支出率 15		遞延所得稅資產
7	淨值/資產 11		每股稅前淨利
8	當期所得稅資產－流動		ROA(C)稅前息前折舊前
9	遞延所得稅資產		利息支出率
10	普通股股本		應收帳款及票據

## 基本面投資建議:以下重要變數可作為基本面投資者進行決策時的參考依據

變數名稱	敘述
每股稅前淨利 EPS_Earnings Per Share	EPS 反映了公司在每股股份上的盈利能力, 高盈利能力可能意味著更高的股價。當一家公司的 EPS 增長時, 通常會對股價有正面影響。
資產報酬率 ROA_Return on Asset	是一個衡量公司資產運營效率的指標。它表示公司每單位資產所產生的利潤。高ROA表示公司在運營資產方面效率較高, 可能對股價產生正向影響。
淨值/資產 Equity/Assets Ratio	這個比率衡量公司淨值與總資產之間的關係, 反映了公司的財務結構。高淨值/資產比可能表示公司有較低的財務風險, 可能對股價產生正面影響。
稅前純益 /實收資本 PBT_Profit Before Tax / Paid-in Capital	這個比率反映了公司每單位實收資本所產生的稅前利潤。這可以衡量公司有效運用資本的能力。高比率可能表明公司的利潤能力較強, 可能吸引更多投資者, 進而對股價有正向影響。
利息支出率 Interest Expense Ratio	這個比率反映了公司利息支出占總收入的比例。高利息支出率可能表明公司面臨較高的財務負擔, 可能對股價產生負面影響。

## 技術面研究流程及方法

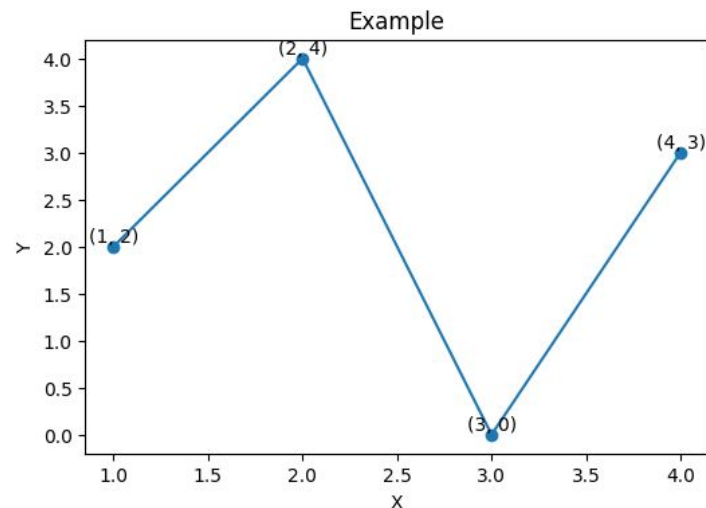
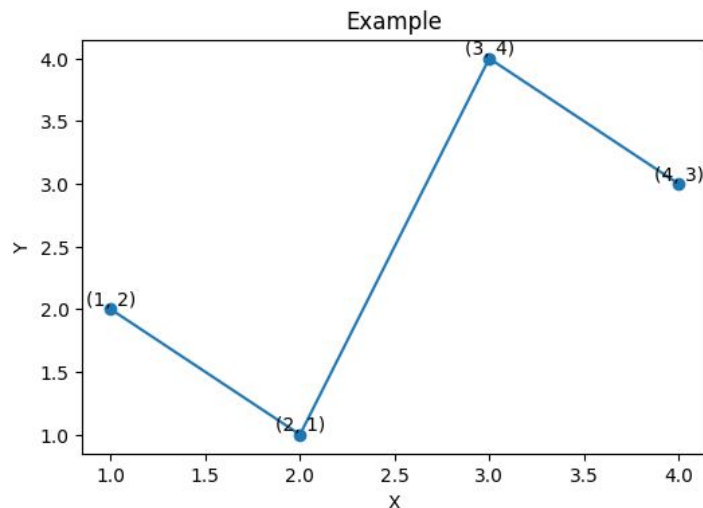
模型	研究方法	模型應用	
技術面	1. 利用爬蟲，下載各公司的開高收低	隨機森林模型	1. 數據切分:將數據切分為訓練及與測試集。
	2. 資料清理及合併 - 切割訓練、測試資料集  - 利用ta套件計算出相關指標(KD, RSI)  - 將所需數據加入新的dataframe		2. 股價預測:通過收集多個公司技術面指標和樹的數量生成多棵決策樹，訓練好的模型對測試集進行預測。
	3. 模型模型	XGBoost模型	1. 數據切分:將數據切分為訓練及與測試集 2. 參數調整:進行參數調整找出最佳參數組合之模型 3. 股價預測:根據公司的技術面指標，預測股票價格的趨勢。通過學習歷史數據，預測公司未來的表現。
	4. 以實際投資期限前60天為模型測試	SVR模型	1. 特徵選擇:利用迴圈來決定需要使用那些特徵。
	5. 模型預測股價，得模型準確率		2. 參數調整:進行參數調整找出最佳參數組合之模型 3. 股價預測:透過公司技術面指標數據集支持向量回歸，進行公司股價預測。

技術面投資:模型預測股價

SVR模型										
公司	台積電 (2330)	聯電 (2303)	宏碁 (2353)	南電 (8046)	華南金 (2880)	富邦金 (2881)	台灣大 (3045)	華碩 (2357)	高端疫苗 (6547)	神隆 (1789)
最高股價	499.84	43.12	24.00	312.45	22.55	56.28	99.03	270.92	123.93	26.29
最低股價	456.00	42.43	21.91	242.09	21.99	55.17	97.30	269.51	118.57	24.46
股價漲幅	9.61	1.63	9.54	29.06	2.95	2.01	1.78	0.52	4.52	7.47
RandomForest模型										
最高股價	539.41	50.43	25.71	263.63	23.48	60.57	97.83	282.91	78.99	27.24
最低股價	443.17	40.82	22.70	217.47	21.69	55.39	92.51	255.60	66.95	23.96
股價漲幅	21.72	23.54	13.26	21.23	8.25	9.35	5.75	10.68	17.98	13.69
XGBoost模型										
最高股價	322.82	29.71	15.26	166.85	14.31	38.04	62.42	178.30	49.09	16.57
最低股價	282.05	25.63	14.50	139.38	14.00	35.15	61.24	161.94	43.84	15.44
股價漲幅	14.45	15.92	5.23	19.71	2.21	8.22	1.93	10.10	11.98	1.07

## 技術面投資策略:不同投資人可能採用不同方式, 造成不同結果

投資人	方式
Tom	根據模型預測結果, 先將最高、最低價之出現先後順序進行第一波篩選, 刪除震盪幅度較大的公司, 再從剩下的公司中, 對比不同模型結果, 選取最佳作為投資標的
Harry	根據模型預測結果, 找出相對最大價差, 對比不同模型結果, 選取最佳作為投資標的



技術面投資:以實際股市來驗證模型結果

投資人	結果					
姓名	公司	預期買進日期 2023/04/14	預期賣出日期 2023/05/19	買進股數	投入資金	最終金額
Tom	台積電(2330)	516.00	532.00	190股	98,040	101,080
報酬率					3.1%	
年化報酬率					12.99%	

姓名	公司	預期買進日期 2023/04/20	預期賣出日期 2023/05/09	買進股數	投入資金	最終金額
Harry	南電(8046)	255.00	287.50	392股	99,960	112,700
總計					12.75%	
年化報酬率					61.58 %	

## 消息面研究流程及方法

模型	研究方法	消息源介紹	
技術面	1.請5個消息源(加自己)投票篩選出四家較不願意進行投資的公司	A	大學生, 投資經歷 2年
	2. 綜合各消息源提供的可獲利(低點)投資點(提供3次投資機會)及最佳投資點(3刺機會的其中1次), 自行消化後, 統整出最好的投資組合, 進行投資。	B	業餘投資者, 年收300萬社會人
		C	Line、Decard、Facebook等股票相關資訊
	3. 根據消息源的建議, 設立做空的時機。	D	職業操盤手
	4. 賣出後, 計算淨利(損)。	E	業餘投資者, 年收約200萬
	5. 與其他兩種層面進行比較。		



## 選出投資時間點 (紅色為最佳投資點)

	消息源	時間點	投資標的	投資策略
A	大學生 投資經歷2年	5/9	宏碁29、華碩290	自己畫趨勢線、考量市場情緒(包含觀察FED)
B	業餘投資者	3/20	台積電510	低點的時候買, 別人恐懼我貪婪
		5/16	華碩302	
C	Line、Dcard、FB 等股票相關資訊	4/24	台積電508、台灣大103	毫無規律, LINE可信度極低, 一旦發現無法詐騙就消失 Dcard、Facebook 內容不可全信以為真
		5/9	聯電49.8	
D	職業投資者	4/24	聯電50、華碩279	合適的選股技巧 找買點 適當停利 寫程式輔助
		5/02	台積電496.5	
		5/16	台積電506	
		5/9	宏碁28.8	
E	業餘投資者	5/16	台積電504	聽內線股居多, 沒有什麼特別投資策略
		5/9	宏碁28.7	

## 投資結果

“ 一共有4票選擇5/9(二)進行投資，為配合只有一次買賣點機會的限制，選擇5/9(二)購買宏碁(28.7)及聯電(49.95)。以9:1分配。於5/24賣出(宏碁31、聯電49.3)

”

“ 賣出的時機點設立於宏碁漲至8%的時刻

”

“ 之所以以9:1的方式形成組合，原因是提議購買宏碁的消息源有3票，而聯電僅1票，且為facebook新聞所推薦，消息來源並非十分可信，因而僅以1成的資金進行投資

”

公司	買進	賣出	淨利 (損)
宏碁 (2353)	3張(86,100元)	3張(93,000元)	6,900元
聯電 (2303)	250股(12,489元)	250股(12,325元)	-164元



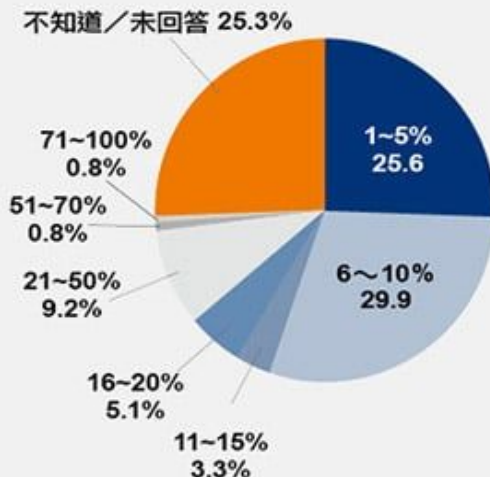
總獲利	年化報酬率
6,736 元	29.79 %

根據年化報酬率比較，可得建構的模型確實可實現幫助投資人提高獲利的目標，驗證本研究對於投資決策的影響力及重要性

### 年化報酬率比較

基本面 最佳模型	47.01 %
技術面 最佳模型	61.58 %
消息面 最佳模型	29.79 %

### 自行選擇投資標的之期望年化報酬率佔比



由上圖可得，在 813 個有效樣本中，僅 10.8% 民眾期望的年化報酬率高於 21%，而本研究中基本面、技術面及消息面年化報酬均高於 25%，因此本研究中的模型投資成效預期可 **滿足市場中超過 90% 投資人的需求**。

### 研究結論

#### ✓ 建立績效好的投資組合

研究結果顯示，模型可幫助投資人選擇較優投資組合以實現投資目標。

#### ✓ 提供投資決策建議

透過模型訓練過程的重要變數，可提供投資人決策時有效參考依據。

**幫助投資人實現高投資報酬**

感謝聆聽！

Q & A

# Appendix

## 相關性變數處理

刪除與 y  
相關性過低  
的變數

避免  
多重共線性

1. 刪除與 y 相關性絕對值  $< 0.2$  之變數
2. 幫助簡化模型、去除不必要特徵
3. 提高模型預測性能及解釋能力

1. 尋找兩特徵相關性  $> 0.7$  的變數 pairs
2. 刪除對目標變數 (y) 的影響較低的不重要變數
3. 提高模型穩定度及解釋能力
4. 幫助簡化模型, 提高模型效率

