

東吳大學巨量資料管理學院  
資料科學系碩士在職專班碩士論文

指導教授：邱嘉豪 博士

傳統量化技術與機器學習在價值投資中的  
投資績效比較分析研究

Comparative Study on the Investment Performance  
of Traditional Quantitative Methods and Machine  
Learning in Value Investing

研究生：梁皓鈞 撰

中華民國 114 年 6 月

# 謝誌



## 中文摘要

本研究旨在比較傳統量化價值投資方法與機器學習模型於台灣股市之選股績效。研究首先以 Piotroski (2000) 所提出之 F-score 為基礎，透過基本面九項指標進行傳統量化選股，建立季交易的選股策略；再進一步利用機器學習模型，採用隨機森林回歸模型，並將 F-score 指標轉換為連續變數、滾動財報數據及上季報酬率等特徵，預測個股下季報酬表現排名並據以選股。回測期間涵蓋 2014 年至 2023 年，資料來源為台灣上市櫃公司，策略調整頻率為每季一次，績效指標包含投資組合年化報酬率、波動度、最大跌幅 (MDD) 及夏普比率等。研究結果顯示，機器學習策略在年化報酬與風險調整後報酬上均優於傳統量化投資策略，顯示結合基本面與機器學習可提升價值型投資的選股效率與報酬表現。



# Abstract

This study aims to compare the stock selection performance of traditional quantitative value investing methods and machine learning models in Taiwan's stock market. The research first adopts the F-score proposed by Piotroski (2000), using nine fundamental indicators to construct a seasonal stock selection strategy based on traditional quantitative methods. It then incorporates a machine learning approach using a Random Forest regression model, which transforms F-score indicators into continuous variables and integrates rolling financial statement data along with previous quarter returns to predict and rank individual stock performance in the following quarter. The backtesting period covers 2014 to 2023, with data sourced from listed companies in Taiwan. The strategy is rebalanced quarterly, and performance metrics include annualized return, volatility, maximum drawdown (MDD), and Sharpe ratio. Results show that the machine learning-based strategy outperforms the traditional quantitative strategy in both annualized return and risk-adjusted performance, indicating that integrating fundamental analysis with machine learning can enhance the efficiency and returns of value investing.

# 目錄

|                            |      |
|----------------------------|------|
| 謝誌.....                    | i    |
| 中文摘要 .....                 | ii   |
| Abstract.....              | iii  |
| 目錄.....                    | iv   |
| 表目錄 .....                  | vi   |
| 圖目錄 .....                  | viii |
| 第一章 緒論.....                | 1    |
| 第一節 背景.....                | 1    |
| 第二節 研究動機 .....             | 2    |
| 第三節 研究目的 .....             | 3    |
| 第四節 研究問題 .....             | 4    |
| 第二章 文獻探討.....              | 6    |
| 第一節 投資策略應用機器學習方法現況說明 ..... | 6    |
| 第二節 傳統量化投資策略.....          | 7    |
| 第三節 F-score 選股策略.....      | 8    |
| 第四節 本研究與過去文獻之主要差異.....     | 11   |
| 第三章 研究方法 .....             | 14   |

|                            |    |
|----------------------------|----|
| 第一節 研究架構 .....             | 14 |
| 第二節 資料蒐集與資料處理 .....        | 14 |
| 第三節 建構投資策略.....            | 17 |
| 第四節 輸入特徵變數及輸出變數 .....      | 21 |
| 第五節 模型選擇與優化.....           | 24 |
| 第六節 投資組合績效的評估方法 .....      | 29 |
| 第四章 研究結果.....              | 32 |
| 第一節 不同模型與變數組合之實驗結果比較 ..... | 32 |
| 第二節 隨機森林模型優化後表現 .....      | 40 |
| 第三節 調整後 F-score 投資策略 ..... | 43 |
| 第四節 機器學習投資策略之績效 .....      | 51 |
| 第五章 結論.....                | 60 |
| 第一節 研究結論 .....             | 60 |
| 第二節 研究建議 .....             | 61 |
| 參考文獻 .....                 | 63 |

## 表目錄

|                                  |    |
|----------------------------------|----|
| 表 1 F-score 指標項目表 .....          | 9  |
| 表 2 本研究與其他文獻之主要差異表 .....         | 12 |
| 表 3 財報會計科目表 .....                | 15 |
| 表 4 財務指標計算方式表.....               | 16 |
| 表 5 調整後 F-score 項目表.....         | 17 |
| 表 6 資料集切分表 .....                 | 22 |
| 表 7 輸入特徵變數設計表.....               | 22 |
| 表 8 輸入變數與輸出變數組合表 .....           | 24 |
| 表 9 模型訓練情境表 .....                | 32 |
| 表 10 XGBoost 驗證集表現評估表 .....      | 34 |
| 表 11 LSTM 驗證集表現評估表 .....         | 36 |
| 表 12 隨機森林驗證集表現評估表 .....          | 38 |
| 表 13 不同模型於最佳變數組合下之驗證集報酬率比較表..... | 39 |
| 表 14 隨機森林參數表 .....               | 40 |
| 表 15 最佳超參數組合表 .....              | 41 |
| 表 16 模型優化後差異表 .....              | 41 |
| 表 17 模型優化前後投組績效差異表 .....         | 42 |

|                                    |    |
|------------------------------------|----|
| 表 18 調整後 F score 投資策略一之績效評估表 ..... | 46 |
| 表 19 調整後 F score 投資策略二之績效評估表 ..... | 50 |
| 表 20 投資組合一之績效評估表 .....             | 54 |
| 表 21 投資組合二之績效評估表 .....             | 55 |
| 表 22 投資組合三之績效評估表 .....             | 57 |
| 表 23 投資組合四之績效評估表 .....             | 58 |



## 圖目錄

|                               |    |
|-------------------------------|----|
| 圖 1 研究架構圖 .....               | 14 |
| 圖 2 調整後 F-score 策略交易時間軸 ..... | 18 |
| 圖 3 F-score 調整後策略運作圖 .....    | 19 |
| 圖 4 機器學習投資策略交易時間軸 .....       | 20 |
| 圖 5 本研究投資策略運作圖 .....          | 21 |
| 圖 6 各模型投組淨值走勢圖 .....          | 40 |
| 圖 7 優化前後投組淨值走勢圖 .....         | 43 |
| 圖 8 投資策略一之選股池上市櫃公司數量圖 .....   | 44 |
| 圖 9 投資策略一之選股池各產業公司數量圖 .....   | 44 |
| 圖 10 投資策略一各季產業分布熱力圖 .....     | 45 |
| 圖 11 投資策略一與 0050 之淨值走勢圖 ..... | 47 |
| 圖 12 投資策略二之選股池上市櫃公司數量圖 .....  | 48 |
| 圖 13 投資策略二之選股池各產業公司數量圖 .....  | 48 |
| 圖 14 投資策略二各季產業分布熱力圖 .....     | 49 |
| 圖 15 投資策略二與 0050 之淨值走勢圖 ..... | 51 |
| 圖 16 機器學習模型選股池上市櫃公司數量圖 .....  | 51 |
| 圖 17 機器學習模型選股池各產業公司數量圖 .....  | 52 |

|                                            |    |
|--------------------------------------------|----|
| 圖 18 隨機森林各季選股分佈熱力圖 .....                   | 52 |
| 圖 19 投資組合一、調整後 F-score 與 0050 之淨值走勢圖 ..... | 54 |
| 圖 20 投資組合一、二之淨值走勢圖 .....                   | 56 |
| 圖 21 投資組合二、三之淨值走勢圖 .....                   | 57 |
| 圖 22 投資組合三、四之淨值走勢圖 .....                   | 59 |



# 第一章 緒論

## 第一節 背景

近年來，台灣股市呈現正成長態勢，尤其在 2024 年 3 月 21 日，台股加權指數由台積電領軍突破 2 萬點，並創下歷史新高。此一里程碑不僅反應台灣股市的發展潛力，亦吸引大量散戶投資者與股市新手投入市場，與此同時，市場上高股息、價值投資型 ETF 不斷推陳出新，進一步推升台股熱潮。然而，台股之成長非僅仰賴少數龍頭企業，市場中仍潛藏許多未被充分發掘、具投資價值之標的值得投資者關注。如何有效篩選具潛力之標的，並將資金配置於具上漲空間之企業，成為當前投資領域之重要課題。

隨著數據科學與人工智慧技術之發展，傳統量化分析結合機器學習之選股策略逐漸受到重視。(Mintarya, L. N., Halim, J. N., Angie, C., Achmad, S., & Kurniawan, A., 2023)表明過去的研究普遍認為相較於傳統量化分析方法，機器學習的應用能提高預測的準確性。(Namdari, A., & Li, Z. S., 2018)運用機器學習結合基本面與技術面來預測股價可提高預測的準確率。(Huang, Y., Capretz, L. F., & Ho, D., 2021)指出機器學習在美股市場可作為提升選股效率和投資績效的工具。(Mintarya, L. N., Halim, J. N., Angie, C., Achmad, S., & Kurniawan, A., 2023)表示過去的研究普遍在比較不同機器學習模型在預測股票的表現，並認為相較於傳統量化分析方法，機器學習的應用能提高預測的準確性。諸多研究證實機器學習的技術用於預測股票市場，不但能提升分析效率，亦有機會改善投資之客觀性與準確率。本研究以台股上市櫃公司為對象，探討傳統量化技術與機器學習技術於投資策略中的應用。

在過去的研究中，多著重於模型的優化與準確率的提升，在實際應用上或許存在限制，尤其在投資標的數量方面，通常未考慮投資人是否便於交易、風險承受度及標的數量的偏好。本研究希望可以建立一套更貼近實際應用的投資策略，使投資人可以根據自

身風險承受度、資金大小，靈活的選擇標的數量，若選擇較少標的會承擔較多風險，但獲得較高的收益；反之則犧牲一些報酬換取風險分散。

除以機器學習評估公司財報表現以預測未來股票報酬排名外，考量歷史財報為落後指標，因此另外加入上一季股價報酬率做為評估指標，以評估歷史財報是否已被反應完畢或股價動能對未來表現的影響。

## 第二節 研究動機

台股近年的高成長趨勢，讓許多股市新手、散戶投資人想投入市場，希望搭上股市成長的熱潮。部分理性的主動投資者，深入研究後才投入股市，部分被動投資人選擇指數型 ETF 以分散風險，然而，面對眾多投資標的，許多投資人缺乏明確方向，亦有投資人認為將資金投入單一龍頭企業就能保障其收益。事實上，台股除了少數龍頭企業外，仍有許多潛在的優質企業值得投資，讓資金更有效率配置於潛力股。

(Fama, 1970)提出效率市場假說，主張效率市場可分為三種型式，分別為弱式效率、半強式效率、強式效率。在弱式效率市場中，股價已反映所有已知的交易資訊，例如價格與成交量，因此技術分析無法獲得超額報酬；在半強式效率市場中，股價不僅已反映已知的交易資訊，甚至已反映已知的財報資訊和新聞消息，因此基本面分析無法獲得超額報酬；而強式效率市場則反映所有資訊，即使有內線消息亦無法獲得超額報酬。然而，市場各種異常現象挑戰了效率市場假說，可推論股票市場並未達到強式效率市場，甚至部份股票屬無效率市場，進而支持利用基本面分析在大多數股票上可為投資人創造收益的觀點。

在眾多分析股票的方法中，基本面分析被視為評估企業價值之核心，揭示企業長期經營績效與財務健康狀況。(Chen, A. S., & Hong, B. S., 2006)研究提到法人投資人的投資決策仰賴多方面的分析，包含基本面分析與股價動能表現，其中又以分析公司財務指

標占大多數，顯示基本面分析的重要性。利用技術分析及籌碼分析可以預見股價短期內的異常反應，股價最終也因企業的內在價值而回到平衡，故技術指標與籌碼指標應僅作為買賣時機的參考，不宜作為長期投資決策的依據。對散戶而言，分析財報需耗費大量時間與專業知識，且面對龐大數據時，如何快速篩選具潛力公司成為一大挑戰。

在投資領域中，股價的預測一直是投資人最關注的事，近年隨著機器學習技術的發展與運算能力的提升，越來越多研究嘗試將機器學習的技術應用於股票市場中。(陳鼎玟, 2018)運用機器學習與深度學習技術預測台股指數、(陳同力, 2019)以 LSTM 加上集成式學習預測上市櫃中 14 檔權值股未來一個月股價的漲跌、(孫永鴻, 2020)使用支援向量機、XGBoost、隨機森林判斷股票買賣時機、(王希佩, 2021)利用機器學習建立股市漲跌預測模型，且皆有不錯的預測能力。由此可知，機器學習方法在投資領域中逐漸成為重要的工具，利用機器學習技術方法有望提升選股效率並優化績效。

研究動機整理如下：

1. 在股市中找出一套績效長期由於大盤表現的方法，並將資金持續投入其中。
2. 希望讓設法讓投資人能從財報中快速辨識出有投資潛力的公司。
3. 比較傳統量化策略及機器學習策略的投資報酬率差異。

### 第三節 研究目的

對於股市新手或散戶投資人而言，逐一分析公司的財報不僅耗時費力，若財報表現未達預期，需重新尋找標，往往錯過最佳進場時機。因此本研究以基本面分析為基礎，利用傳統量化投資策略與機器學習方法，系統化評估台灣上市櫃公司之財報表現，旨在以客觀數據分析取代主觀判斷，協助投資人快速識別具上漲潛力的公司。此外，本研究

不僅利用機器學習提供投資標的建議，亦分析各財報指標對股價影響的程度，使投資人得以用更透明的數據，判斷公司財報的優劣，提升決策品質。

近年人工智慧與機器學習技術蓬勃發展，相關技術已廣泛應用於各領域以解決痛點。本研究希望利用機器學習進行選股，除克服主觀分析的偏見及逐一分析財報的成本外，亦能為投資人打造一套系統化之選股策略。為進一步凸顯機器學習在選股策略中的價值，本研究將深入比較機器學習投資策略與傳統量化投資策略在投資績效與風險程度上的差異，透過回測，比較兩投資策略之投資報酬率、波動程度及夏普比率，以提供投資人更全面的決策參考。

研究目的整理如下：

1. 找出基本面的關鍵指標，提供投資人客觀數據以選擇投資標的。
2. 利用機器學習的方法，更透明的對公司財報表現評價，找出具上漲潛力的公司。
3. 運用機器學習，為自然人投資人打造一個有系統、有彈性又有投資紀律的選股策略。
4. 透過回測比較機器學習投資策略與傳統量化投資策略在投資報酬率、風險程度與夏普比率的差異。

## 第四節 研究問題

本研究及於上述動機與目的，提出以下研究問題，以探討傳統量化技術與機器學習於價值投資中之應用效果。

1. 驗證傳統量化的投資策略建構的投資組合績效，是否能贏過加權指數。

本研究利用傳統量化的投資策略，透過財務指標篩選具基本面優勢之公司組成投資組合，並驗證其報酬率與風險程度，再與加權指數表現進行比較，由於加權指數無法投資，故以元大台灣 50 ETF 為比較對象。

2. 將機器學習應用於基本面分析，對選股策略績效之影響。

本研究探討機器學習選股的準確性與績效表現，並進一步比較傳統量化投資策略的投資報酬率與風險程度之差異。

3. 優化機器學習的模型後，是否能減少投資組合的標的數量，同時提升投資績效。

本研究希望透過優化機器學習模型，精簡投資組合的規模，以減少交易及管理成本，提升報酬率與穩定性，並實現長期超越加權指數的目標。



## 第二章 文獻探討

### 第一節 投資策略應用機器學習方法現況說明

隨著機器學習技術的發展，此技術在投資領域的應用受到許多投資人的關注。因此愈來愈多研究在探討如何利用機器學習提升投資策略的準確性、報酬率及風險管理能力。(Mintarya, L. N., Halim, J. N., Angie, C., Achmad, S., & Kurniawan, A., 2023)回顧了2012~2020 將機器學習應用於股票市場預測的研究，分析了30 篇相關文獻，發現神經網路最常被用來預測股票市場，其次為長短記憶模型 (LSTM)，但這並不意味其他模型在預測股價的表現不佳。(Gu, S., Kelly, B., & Xiu, D., 2020)指出機器學習方法，如隨機森林、神經網路、XGBoost，在捕捉非線性模式與市場異常的方面較傳統線性模型的能力更勝一籌，利用大量股票特徵數據來建構投資策略，並採用機器學習模型能穩定產生超額報酬，但同時亦需注意模型的過擬合風險。(Rasekhschaffe, K. C., & Jones, R. C., 2019)使用了194 個特徵變數預測股票未來1 個月的風險調整後報酬，並且測試了四種機器學習模型，結果顯示以神經網路模型表現最佳，研究也指出，若將各模型組合，將預測結果加權平均能有效提升整體預測能力。

在台股市場中，(卓越, 2019)利用支援向量回歸 (SVR) 建構選股模型，並使用主成分分析對170 個因子進行特徵提取與降維，預測股票下一季的收盤價以計算預期報酬率，並選取前數名組成投資組合，研究結果指出模型對台股市場具有一定程度的預測能力，回測績效亦優於大盤。(謝曜宏、許資沄、朱涵宇、林祝興, 2019)採用隨機森林模型，結合財務指標與技術指標預測股票2 個月後漲跌，並根據特徵重要度重新計算股票漲跌的機率，經過回測顯示，選股策略在2014~2019 年化報酬率可達12%。(田鉉元, 2022)以台股市值前300 大公司為對象，透過隨機森林預測下一季報酬後，依特徵重要度篩選股票，最後再透過葛拉漢和巴菲特的價值投資原則進行篩選建立投資策略。(郭峻

宇, 2023)利用 (Piotroski, 2000)提出的 F-score 財務指標加入貨幣供給量變化作為特徵變數，採用 XGBoost 預測 0050 成分股的漲跌，最後利用特徵重要度與特徵變數的乘積來對公司評分，並買進得分較高的股票以建構投資組合。

不論是國內外股票市場，利用機器學習來預測股價表現的研究已相當普遍，且多數研究證實了機器學習對於股價表現具有一定的預測能力，並著重於提升模型的準確率，未提出實際的投資策略。而在台股市場，大多數研究多以股價的漲跌作為預測對象，再利用特徵重要度設計選股策略。相較之下，直接預測股票報酬率或股票價格的研究較少，主要原因可能是股票價格或報酬率隨機性較高，較難以精準預測所致。

## 第二節 傳統量化投資策略

(Lakonishok, J., Shleifer, A., & Vishny, R. W., 1994)研究指出，買入低本益比、低本淨比的股票，長期而言表現優於市場平均表現。(Chan, L. K., Jegadeesh, N., & Lakonishok, J., 1996)研究指出基本面指標對股票未來表現有顯著預測能力。(Graham, B., & McGowan, B., 2003)是著名的價值派投資人，其於《The intelligent investor》一書中強調，長期投資應根據公司的內在價值分析，利用財報分析，找出具備穩定收益、良好的財務槓桿與成長潛力的公司，而非著重於短期市場波動，並建議投資人應分散持股，避免將資金集中於單一股票或行業。

鑒於國外著名價值派投資者長年以來投資績效表現優異，有多項研究參考著名投資人的選股邏輯，透過設定選股條件方式建構投資組合。(李雅璇, 2013)依照巴菲特的選股邏輯，略為調整以適應台股市場特性。(張祺峻, 2023)則結合巴菲特、葛拉漢、彼得林區之選股邏輯，再加入其他基本面指標選股。(韓家韋, 2024)設定基本面、技術面、籌碼面之篩選條件選股，研究證明純基本面選股之績效優於加入技術面與籌碼面選股之績效。多數研究設定數個條件，如過去數年股東權益報酬率大於一定值、過去本益比小

於一定值、營業活動現金流量大於零等，篩選出符合條件之股票以建構投資組合，長期來看皆有超越大盤的績效表現。然而，篩選條件若過於嚴苛，可能導致無符合之標的，反之則無參考價值。

另一種量化投資策略則係利用各項指標設法分辨公司的優劣，選出排名前數名標的建構投資組合。(葉千綺, 2015)注重基本面分析，探討財報發布時，從財報資訊中找出具有影響力的指標，透過指標間的關係建構投資組合，具體而言先利用獨立樣本 T 檢定，檢驗哪些指標對於個股報酬能區分出差異性，再利用主成分分析賦予各指標權重，建立多因子模型對股票評分，最終設定每季投資家數，挑選分數高的股票買進，回測結果顯示於 2001~2014 年間最高年平均報酬率可達 25%。(許珮琪, 2019)以 2008~2019 年台灣上市櫃公司為對象，驗證股價淨值比、股價營收比、本益比等 7 項財報指標個別與股票報酬率的關係，透過單因子模型建構投資組合，找出績效表現最佳的兩個財務指標，進而建構雙因子的選股策略。透過研究發現在選股策略上，以股價營收比和盈餘產出率建構的選股策略優於單一指標建構的選股策略。(Piotroski, 2000)提出 F-score 來對公司財報評價，利用 9 項基本面指標為公司財報評分，並根據得分來建構投資組合，最終發現以此策略所建構的投資組合可打敗大盤。

## 第三節 F-score 選股策略

### 一、F-score 選股策略介紹

(Piotroski, 2000)主要研究如何針對價值股利用簡單的財務表現，辨識出具投資價值的公司，因而提出了「F-score」，利用歷史的財報區分哪些公司在股價表現上是贏家和輸家。

F-score 的核心在於針對高帳面市值比的價值型公司，利用 9 項容易實施的基本面指標來衡量公司的財務表現，9 項指標涵蓋了三大面向，分別為獲利性、安全

性及成長性。獲利性的財務指標為：ROA（當期資產報酬率為正）、 $\Delta$ ROA（當期資產報酬率大於前期資產報酬率）、CFO（當期營業活動現金流為正）、ACCRUAL（當期營業活動現金流大於稅後淨利）；安全性的財務指標為： $\Delta$ LEVER（當期長期負債小於前期長期負債）、 $\Delta$ LIQUID（當期流動比率大於前期流動比率）、EQ\_OFFER（前一年沒有發行新股）；成長性的財務指標為： $\Delta$ MARGIN（當期毛利率大於前期毛利率）、 $\Delta$ TURN（當期總資產週轉率大於前期總資產週轉率）。若符合 F-score 中任一項指標得 1 分，否則 0 分，總分最高可獲得 9 分，各項指標整理如表 1。

表 1 F-score 指標項目表

| 指標面項 | 指標名稱            | 指標意義                    |
|------|-----------------|-------------------------|
| 獲利性  | ROA             | 今年度資產報酬率 $>0$           |
|      | $\Delta$ ROA    | 今年度資產報酬率 $>$ 去年度資產報酬率   |
|      | CFO             | 今年度營業活動現金流為正            |
|      | ACCRUAL         | 今年度營業活動現金流 $>$ 稅後淨利     |
| 安全性  | $\Delta$ LEVER  | 今年度長期負債 $<$ 去年度長期負債     |
|      | $\Delta$ LIQUID | 今年度流動比率 $>$ 去年度流動比率     |
|      | EQ_OFFER        | 前一年沒有發行新股               |
| 成長性  | $\Delta$ MARGIN | 今年度毛利率 $>$ 去年度毛利率       |
|      | $\Delta$ TURN   | 今年度總資產週轉率 $>$ 去年度總資產週轉率 |

研究的樣本以 1976~1996 年美國股市為主，為確保投資人在建構投資組合時可以獲得年度的財務報表，因此投資組合需在會計年度結束後的第五個月初，建構投資組合後再衡量投資組合持有一年和兩年的報酬率。經過研究顯示 F-score 高的企業表示財務表現優異，持有高分企業一年平均可獲得 13.4% 的報酬率，相較於持有 F-score 低的企業，一年平均報酬率為 -9.6%。

研究最終透過建構多空的投資策略並回測策略的績效，投資組合的建構先篩選出價值股，研究將價值股定義為帳面市值比最高的前 20% 企業，再從這些企業中找出高 F-score（指總分 8、9 分）的族群與低 F-score（指總分 0、1 分）的族群。由

於高分族群表示財務表現強勁，因此投資組合將買入 F-score 高的企業並做空 F-score 低的企業，經過回測顯示在 1976~1996 年間，該策略平均每年可獲得 23% 的報酬率。

## 二、F-score 的應用與調整

由於 F-score 是為價值股量身設計，且樣本是 1976~1996 年間的美國企業，為了瞭解其在不同市場或不同財務屬性的適用性，國內外有諸多研究將 F-score 調整或加入不同的指標以適應不同的市場環境與公司屬性。

(Mohanram, 2005)希望從成長股中區分好公司與壞公司，因此結合傳統的基本面指標，加入專為低帳面市值比（成長股）公司設計的指標，提出 G-score。最後與 F-score 比較發現，G-score 在成長股中的績效表現更佳。(林佳靜, 2019)未調整 F-score 的得分及權重，但增加了殖利率、自由現金流量、董監持股的條件來篩選公司，依 12 個指標的得分，搭配本益比高低建構兩個投資組合，並比較兩投資組合與台灣加權指數的績效。研究結果發現加入其他指標後的投資組合表現皆優於大盤，其中又以低本益比的投資組合表現尤佳。(張嘉芸, 2024)將 F-score 應用於台灣上市公司，除 F-score 原有的指標外，另加入了公司規模、現金股利率及公司治理評鑑分數，並以 2015~2022 的台股上市公司為對象選股，透過實證認為 F-score 應用於台灣上市公司可獲得超越大盤的投資表現。

(王銘駿、陳宜伶、吳昭億、黃巧雯, 2015)參考 F-score，考量台灣會計制度與台灣股票市場特有的現金增資結果，調整後提出更適合台灣股票市場的 A-score。A-score 與 F-score 的差異之處在於考量會計制度對非常損益的限制，計算稅後淨利時，不將非常損益項從中扣除；且研究認為現金增資對台灣企業具正面影響，故得分邏輯與 F-score 相反。研究結果也顯示在台灣股市中，A-score 的報酬在大多數期間顯著優於 F-score，因此 A-score 成為台股值得參考的投資策略。(夏明義, 2022)根據

公司在 F-score 各項指標的達成率調整分數，提出了「Revised F-score」，並透過各種交易策略比較 F-score 修正前後的績效，結果顯示無論是在何種交易策略下，修正後的 F-score 皆可提升年化報酬率。

## 第四節 本研究與過去文獻之主要差異

既有運用機器學習技術進行台股選股之文獻與本研究之主要差異如表 2。本研究從 F-score 九項指標出發，將原本 F-score 之類別型變數轉為連續型變數，並且將近三年之指標作為滾動特徵，不僅保留原有的基本面選股邏輯，亦使模型更能辨識歷年各指標對報酬率的影響。同時，為補足財報為落後指標的缺陷，本研究加入上季股價報酬率作為動能特徵變數，以捕捉市場對財報的即時反應，進一步提高預測模型的實務準確性。在模型選擇方面，本研究採用隨機森林回歸模型，以預測持有一季後的報酬率，並依預測值排序作為選股依據。與多數文獻採用分類模型不同，本研究重視模型區辨報酬率表現之差異，有助於實際投資決策中挑選高潛力標的。在投資組合建構方面，過去多數研究往往需買進所有符合篩選條件之股票，導致每次調整時所需納入的投資標的數量過多，實務執行上較為困難，本研究允許投資人根據資金與風險承受度，彈性選擇標的數量，更符合實務資產配置需求。在回測方法與績效評估上，本研究採用總報酬率、波動度、最大回撤（MDD）及夏普比率等指標進行全面性檢驗，而非僅以報酬率或準確率為績效指標，能更全面衡量策略風險調整後之報酬能力。。

整體而言，本研究在延續基本面選股精神下，導入機器學習技術進行系統性改造，從特徵轉換、模型設計、調整頻率到績效指標均貼近實務操作環境，並強調投資組合管理之彈性與風險控制。因此，相較於以往文獻著重理論模型準確率或靜態條件式選股之侷限，本研究更重視「選股策略在真實投資環境下的可操作性」，有助於縮短研究與實務之落差，為自然人投資人提供具體可行之策略。

表 2 本研究與其他文獻之主要差異表

| 研究者                                            | 資料範圍                                | 研究貢獻                                                                                                                                                                                                                                                                                                                                                                            |
|------------------------------------------------|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 本研究                                            | 2014~2023<br>台灣上市櫃<br>公司排除金<br>融類股  | <p><b>研究設計：</b></p> <ol style="list-style-type: none"> <li><b>選股模型：</b>參考 9 項 F-score 基本面指標轉換為連續型變數轉換，加入過去滾動財報資訊及上季已發生之個股報酬率作為特徵變數。利用隨機森林回歸模型預測 1 季後之報酬率，再依預測值排名選股。</li> <li><b>策略執行頻率：</b>每季一次。</li> <li><b>回測方法：</b>未考慮手續費、交易稅及滑價，每季買進排名前數名之標的，次季財報公布後進行換股，並進行再平衡。</li> <li><b>績效評估指標：</b>投組淨值變化、總報酬率、波動度、MDD、夏普比率。</li> </ol> <p><b>研究貢獻：</b><br/>結合傳統與機器學習，優化標的數量與彈性選股。</p> |
| (謝曜<br>宏、許<br>資宏、<br>朱涵<br>宇、林<br>祝興，<br>2019) | 2014~2019<br>台股上市櫃<br>公司            | <p><b>研究設計：</b></p> <ol style="list-style-type: none"> <li><b>選股模型：</b>利用多項基本面指標及技術面指標作為特徵變數。利用隨機森林預測 2 各月後之漲跌，再依特徵重要度選股。</li> <li><b>策略執行頻率：</b>每 2 個月一次。</li> <li><b>回測方法：</b>無手續費、交易稅及滑價，持有 2 個月後調整投組並再平衡。</li> <li><b>績效評估指標：</b>投組報酬率</li> </ol> <p><b>研究貢獻：</b><br/>透過機器學習找出有效的基本面及技術面指標。</p>                                                                             |
| (田鉉元，<br>2022)                                 | 2013~2021<br>年市值前<br>300 大之上<br>市公司 | <p><b>研究設計：</b></p> <ol style="list-style-type: none"> <li><b>選股模型：</b>利用 25 個基本面指標作為特徵變數。利用隨機森林回歸模型預測下一季收盤價後，再依特徵重要度及該研究自行定義之價值投資選股原則篩選標的。</li> <li><b>策略執行頻率：</b>每季一次</li> <li><b>回測方法：</b>未考慮手續費、交易稅及滑價，季持有回測。</li> <li><b>績效評估指標：</b>年化報酬率、年化波動度、夏普比率</li> </ol> <p><b>研究貢獻：</b><br/>將隨機森林結合價值投資理念用於台股。</p>                                                                 |

| 研究者         | 資料範圍                              | 研究貢獻                                                                                                                                                                                                                                                                                                                                                     |
|-------------|-----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (郭峻宇, 2023) | 2013~2023<br>第一季之<br>0050 成分<br>股 | <p><b>研究設計：</b></p> <ol style="list-style-type: none"> <li><b>選股模型：</b>利用 9 項原始 F-score 指標，加入貨幣供給變化轉化為類別型變數作為特徵變數。利用 XGBoost 分類模型預測次季漲跌，再依特徵重要加權分數選股。</li> <li><b>策略執行頻率：</b>每季一次。</li> <li><b>回測方法：</b>未考慮手續費、交易稅及滑價，比較持有 1 季、2 季及 1 年後賣出之績效。</li> <li><b>績效評估指標：</b>投組正報酬勝率、投組報酬率</li> </ol> <p><b>研究貢獻：</b></p> <p>利用機器學習加入總體經濟指標提升 F-score 之準確性。</p> |
| (張嘉芸, 2024) | 2015~2022<br>台灣上市公<br>司           | <p><b>研究設計：</b></p> <ol style="list-style-type: none"> <li><b>選股模型：</b>利用原始 9 項 F-score 指標加入公司規模、帳面市值比、現金股利率、公司治理評鑑分數篩選股票。</li> <li><b>策略執行頻率：</b>每年一次</li> <li><b>回測方法：</b>比較以年初之收盤價買賣及年末之收盤價買賣之差異，未考慮手續費、交易稅及滑價</li> <li><b>績效評估指標：</b>投資報酬率</li> </ol> <p><b>研究貢獻：</b></p> <p>透過 F-score 結合其他基本面資訊，驗證 F-score 應用於台股之可行性。</p>                          |

## 第三章 研究方法

### 第一節 研究架構

本研究之整體架構如圖 1 所示。本研究首先進行資料蒐集與處理，並將所得資料分別應用於傳統量化投資策略及機器學習投資策略的分析。研究流程分為兩大主軸，分別是「傳統量化投資策略的績效評估」、「機器學習投資策略的績效評估」。在建構機器學習模型前，需先進行輸入與輸出變數的設計；而在模型選擇階段，則測試隨機森林(Random Forest)、XGBoost 及長短期記憶網路(LSTM)的表現，並選取績效最佳的模型進行優化。透過績效回測，探討傳統量化投資策略與機器學習投資策略之間的差異。

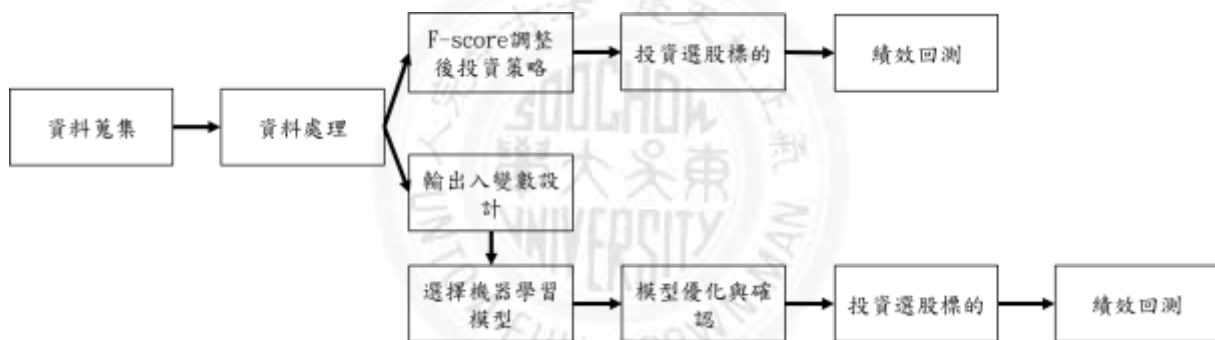


圖 1 研究架構圖

### 第二節 資料蒐集與資料處理

#### 一、資料蒐集

本研究以台灣上市櫃公司為研究對象，但排除金融類股（如金控業、銀行業、證券業及保險業）以及部分業務模式特殊的企業（如新纖、中纖、和泰車等）。此類公司因財務報表結構與會計模式異於一般企業，且多不具毛利率或涉及大量金融性

資產與負債，故不納入分析範圍。為避免資料欄位過多，財務報表蒐集僅涵蓋研究所需項目，各報表選取的會計科目詳見表 3。

表 3 財報會計科目表

| 報表名稱  | 會計科目           |
|-------|----------------|
| 綜合損益表 | 公司代號           |
|       | 公司名稱           |
|       | 營業收入           |
|       | 營業成本           |
|       | 營業毛利（毛損）       |
|       | 營業利益（損失）       |
|       | 本期淨利（淨損）       |
|       | 基本每股盈餘（元）      |
| 資產負債表 | 公司代號           |
|       | 公司名稱           |
|       | 流動資產           |
|       | 非流動資產          |
|       | 資產總計           |
|       | 流動負債           |
|       | 非流動負債          |
|       | 負債總計           |
|       | 權益總計           |
|       | 每股參考淨值         |
| 現金流量表 | 公司代號           |
|       | 公司名稱           |
|       | 營業活動之淨現金流入（流出） |
|       | 投資活動之淨現金流入（流出） |

資料來源係利用 Python 程式語言，自公開資訊觀測站之彙總報表中，蒐集除上述排除產業與公司外的一般企業於 2013 年第一季至 2023 年第四季的综合損益表、資產負債表及現金流量表。此外，因發行新股大致上可分為公司直接籌措資金發行新股，及其他特殊發行新股，如盈餘增資、資本公積轉增資…。前者屬於現金增資，係公司為改良財務結構而發行新股，會導致在外流通股數增加，影響原始股東權益；

後者特殊性發行新股通常不為改善財務結構所用，對原始股東影響較小。考量「近一年度有無發行新股」屬「安全性」面向，若有發行新股則無法在此項目得分，故自公開資訊觀測站公告查詢，蒐集各家公司在各季財報公布日前，近一年度有無「現金增資」的資訊，作為判斷有無發行新股的依據。另外自台灣證券交易所、台灣櫃檯買賣中心蒐集了一般公司於上述區間每季財報公布月份最後一個交易日的股價淨值比。依此方式蒐集了 11 年度共 44 季的財務報表及淨值比。

## 二、資料處理

將蒐集之財報資料依公司代號及季別合併後，共計 71,168 筆資料。財報中之遺漏值主要源於公開資訊觀測站對於數值為零之會計科目以「--」表示。為充分利用資料，本研究不剔除含遺漏值之數據，改以補零處理。此外，部分公司於回測期間因暫停交易，缺乏股價淨值比及收盤價數據，無法計算淨值比或進行投資，故予以刪除。資料處理後，剩餘 67,678 筆資料，涵蓋 2013 年至 2023 年每季財報。

合併資料並處理遺漏值後，部分財務指標如資產報酬率（ROA）、流動比率、毛利率及總資產週轉率需依財報會計科目自行計算。因此，本研究先行計算並建置相關財務指標以供後續分析，其計算方式見表 4。

表 4 財務指標計算方式表

| 名稱         | 計算方式          |
|------------|---------------|
| 資產報酬率(ROA) | 本期淨利÷資產總計     |
| 流動比率       | 流動資產÷流動負債     |
| 毛利率        | 營業毛利（毛損）÷營業收入 |
| 資產週轉率      | 營業收入÷資產總計     |

## 第三節 建構投資策略

考量基本面選股需參考歷史財報表現，股價反映財報資訊需一定時間，持股期間不宜過短。本研究旨在設計每季調整的投資組合，且為了確保投資人及模型可以利用每季財務指標選股，故投資組合的調整需配合台灣上櫃公司公布財報時間。台灣上市櫃公司財報公布時點如下：第一季（Q1）於 5 月 31 日前、第二季（Q2）於 8 月 31 日前、第三季（Q3）於 11 月 31 日前、第四季（Q4）於次年 3 月 31 日前。有鑑於此，本研究之投資策略於每季財報公布後首個交易日將資金平均投入至具成長潛力的公司。

### 一、F-score 調整後之投資策略

原始 F-score 選股策略以年度財務指標評估公司，並每年調整投資組合。為適應本研究每季再平衡之設計，調整 F-score 為以每季財務指標評估公司表現，下稱「調整後 F-score」。各指標得分方式詳見表 5，每項指標符合條件得 1 分，總分範圍為 0 至 9 分。惟因現金增資為重大決策且影響原始股東權益的期間較長，「近一年是否發行新股」仍以過去一年資訊判斷得分。

表 5 調整後 F-score 項目表

| 指標面項 | 指標項目                 | 得分    |
|------|----------------------|-------|
| 獲利性  | 本季資產報酬率(ROA) > 0     | 0 或 1 |
|      | 本季資產報酬率(ROA) > 去年同季  | 0 或 1 |
|      | 本季營業現金流(CF) > 0      | 0 或 1 |
|      | 本季營業現金流(CF) > 本季稅後淨利 | 0 或 1 |
| 安全性  | 本季長期負債 < 去年同季        | 0 或 1 |
|      | 本季流動比率 > 去年同季        | 1 或 0 |
|      | 近一年有無發行新股（現金增資）      | 0 或 1 |
| 成長性  | 本季毛利率 > 去年同季         | 0 或 1 |
|      | 本季資產週轉率 > 去年同季       | 0 或 1 |

本研究首先進行選股策略之績效回測。因投資組合持有期至少一季，交易頻率較低，故回測時暫不計入手續費及證券交易稅影響。部分指標需與前一年同期數據比較，因 2013 年無前期數據可供對照，該年資料予以排除。回測起點為 2014 年第一季財報公布後（即 2014 年 6 月 1 日），終點為 2024 年第一季財報公布前（即 2024 年 5 月 31 日），回測期間共 10 年，各季投資日期如圖 2



圖 2 調整後 F-score 策略交易時間軸

原始 F Score 策略屬價值投資派，會先篩選每期股價淨值比排名前 25% 的公司後計算得分。為全面評估傳統量化投資策略績效，本研究回測兩種投資組合：投組一為買進所有每季得分達 8 分以上的公司；投組二為買進所有得分達 8 分以上且股價淨值比數值較小的 25% 的公司。投資組合建立日 2014 年 6 月 1 日，首先挑選出所有調整後 F-score 得分大於 7 分之標的，以等比例的方式買進所有標的，次季財報公布後，再依調整後 F-score 得分汰弱留強，將得分低於 8 分之個股剔除出投資組合，同時重新配置權重，使各標的在投資組合中維持相同比例之部位，即進行再平衡操作，每年不斷循環直至回測期間終止，投資策略流程如圖 3。

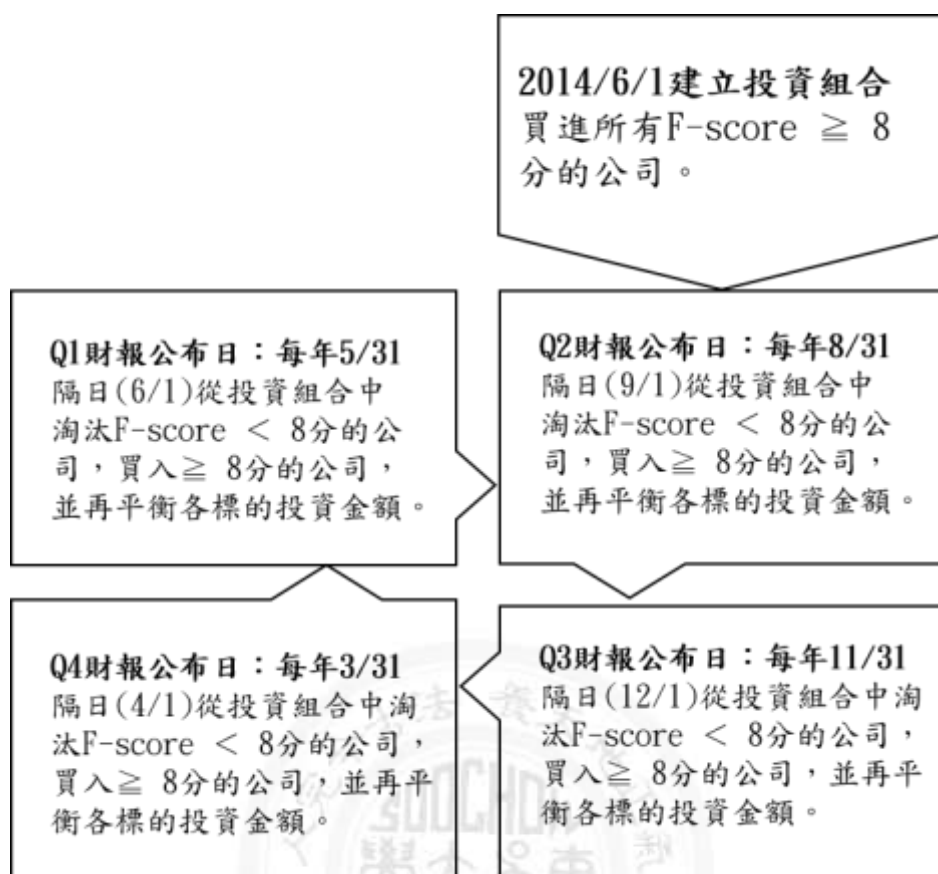


圖 3 F-score 調整後策略運作圖

## 二、以機器學習建構之投資策略

由於 F-score 篩選出的標的數量較多，一般投資者難以逐一人工下單。雖然結合本淨比的篩選可減少標的數量，但此方法仍無法讓投資者根據自身的風險承受度靈活調整標的數量。因此，本研究的最終目標是使投資者能夠依據個人風險偏好，參考模型預測結果，自由選擇排名靠前的若干標的來構建投資組合。

與 F-score 的交易方式類似，本研究所提出的模型亦需利用完整的財務指標來預測股票未來表現。因此，基於機器學習構建的投資策略，其交易日期與 F-score 相同，皆為每季財報公布後交易，首次建構投組日期為 2021 年 6 月 1 日，利用模型

預測之排名，依投資人風險承受度選擇前數名標的建構投資組合，並於每季進行再平衡，如圖 4 所示。

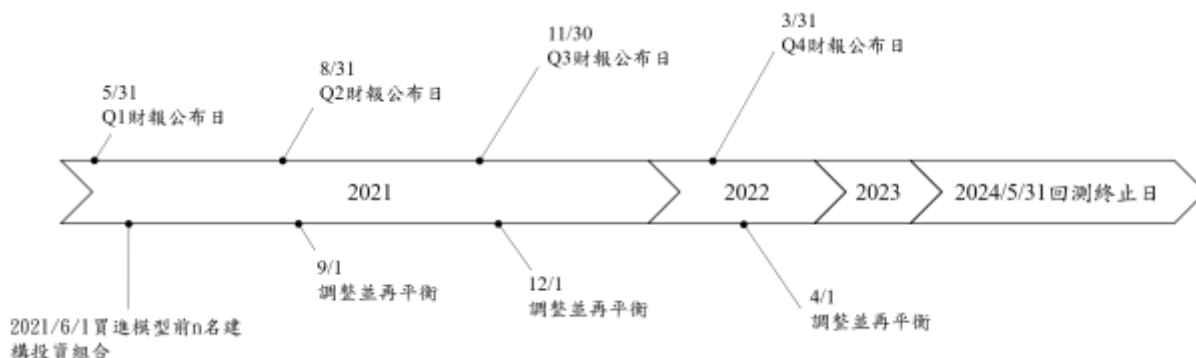


圖 4 機器學習投資策略交易時間軸

本研究最終透過機器學習對每季各股票的表現進行排名，並依據排名結果設計投資策略，運作方式如圖 5。與 F-score 的主要差異在於，投資者可根據自身的風險承受度選擇標的數量。具體而言，投資組合的建立日期定為 2021 年 6 月 1 日，即 2021 年第一季財報發布後的第一個交易日，選取排名前 n 的標的進行買入以構建投資組合，次季財報公布後，再依模型預測結果汰弱留強，剔除投資組合中未進入自訂排名的標的，並買入符合自訂排名的標的，同時亦須重新配置權重，使各標的在投資組合中維持相同比例之部位。舉例而言，2021 年 6 月 1 日分別買入前三名之公司 A、B、C 各 10,000 元之部位，2021 年 8 月 31 財報公布後，投組淨值成長至 33,000 元，公司 A、B、C 的淨值分別為 12,000、10,000、11,000 元。根據最新排名，原有的公司 C 已跌出前段排名，而公司 D 新進入前三名，需依據最新排名進行投組調整。投資人將以公司 D 取代公司 C，並將原本公司 C 的 11,000 元部位賣出後重新配置，首先，從公司 A 中賣出 1,000 元部位，並將該金額加至公司 B，使公司 B 之持股從 10,000 元提升至 11,000 元，接著以原本公司 C 之部位金額 11,000 元投入至新納入的公司 D，使調整後的公司 B 與公司 D 持股均為 11,000 元，完成再平衡作業。

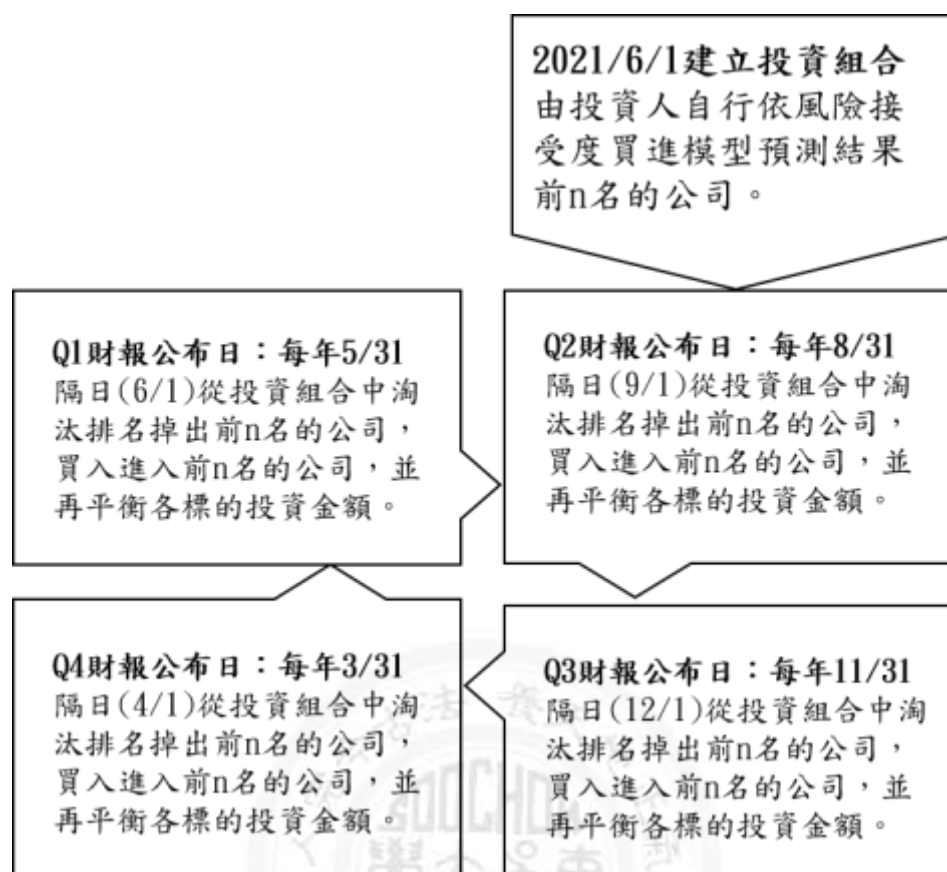


圖 5 本研究投資策略運作圖

## 第四節 輸入特徵變數及輸出變數

為評估模型表現並優化參數，本研究將 2014 至 2023 年資料依 7:1:2 比例劃分為訓練集、驗證集與測試集，如表 6。訓練集用於模型訓練，驗證集用於評估模型在不同情境下的表現以選取最佳模型及參數，測試集則用於最終績效評估。其中，訓練集涵蓋 2014 年第一季至 2020 年第四季（7 年，37,918 筆）；驗證集涵蓋 2021 年第一季至 2021 年第四季（1 年，6,087 筆）；測試集涵蓋 2022 年第一季至 2023 年第四季（2 年，12,591 筆）。

表 6 資料集切分表

| 資料集名稱 | 涵蓋區間          | 公司家數  | 資料筆數   |
|-------|---------------|-------|--------|
| 訓練集   | 2014Q1~2020Q4 | 1,554 | 37,918 |
| 驗證集   | 2021Q1~2021Q4 | 1,554 | 6,087  |
| 測試集   | 2022Q1~2023Q4 | 1,641 | 12,591 |

## 一、輸入特徵變數的計算方式

在設計輸入特徵變數時，本研究調整 F Score 組成指標作為輸入特徵變數。除「近一年是否發行新股（現金增資）」維持 0 和 1 計分外，另考量本季營業現金流過於離散，故以 1 和 -1 表示正面影響和負面影響，其餘指標改為連續型變數，以提升模型對指標差異的辨識能力，計算方式詳見表 7，此組特徵稱為「最近一季財務指標」。

表 7 輸入特徵變數設計表

| 原得分計算方式              | 調整後計算方式                                                  |
|----------------------|----------------------------------------------------------|
| 本季資產報酬率(ROA) > 0     | 本季資產報酬率(ROA)數值                                           |
| 本季資產報酬率(ROA) > 去年同季  | $(\text{本季資產報酬率(ROA)} \div \text{去年同季} - 1) \times 100$  |
| 本季營業現金流(CF) > 0      | 調整為 1 與 -1 表示                                            |
| 本季營業現金流(CF) > 本季稅後淨利 | $(\text{本季營業現金流(CF)} \div \text{本季稅後淨利} - 1) \times 100$ |
| 本季長期負債 < 去年同季        | $(\text{本季長期負債} \div \text{去年同季} - 1) \times 100$        |
| 本季流動比率 > 去年同季        | $(\text{本季流動比率} \div \text{去年同季} - 1) \times 100$        |
| 近一年有無發行新股（現金增資）      | 維持 0 與 1，不調整                                             |
| 本季毛利率 > 去年同季         | $(\text{本季毛利率} \div \text{去年同季} - 1) \times 100$         |
| 本季資產週轉率 > 去年同季       | $(\text{本季資產周轉率} \div \text{去年同季} - 1) \times 100$       |

基本面選股除參考最近一季財報外，亦需考量過去數年表現。為使模型綜合歷史財報資訊，本研究建構 1 年、2 年、3 年及 5 年之滾動特徵，測試不同年限的影響。以 1 年滾動特徵為例，假設當前為 2014 年 6 月 1 日，除 2014 年第一季財務指標外，另納入 2013 年第一季至第四季的指標作為滾動特徵。依此建立 4 組（共 36

項) 滾動特徵，若因 2013 年數據缺失，則以向後填補法補值，即以 2014 年第一季數據填入 2013 年各季之中。

此外，根據效率市場假說，在效率較高的市場中，股價對基本面資訊反應迅速，甚至於財報公布時已反映完畢。為探討上一季股價漲跌對本季報酬率的影響，本研究測試是否納入「上一季報酬率」作為特徵變數，並比較模型表現差異。

## 二、輸出變數的計算方式

本研究旨在讓模型識別每季表現優異的公司，故以持有該公司一季後的投資報酬率作為輸出變數，讓模型從歷史財報預測持有 1 季報酬率。報酬率計算採用 yfinance 套件，提取每季財報公布後首個交易日的調整後收盤價 ( $Adj\ Close_i$ ) 與下一季財報公布前最後一個交易日的調整後收盤價 ( $Adj\ Close_{i+1}$ )。

$$\text{持有 1 季報酬率} = Adj\ Close_{t+1} \div Adj\ Close_t - 1$$

考量報酬率離散值可能過大影響模型效能，本研究另將報酬率轉換為 0 至 1 的連續變數，命名為「持有 1 季報酬 PR」，並比較以「持有 1 季報酬率」或「持有 1 季報酬 PR」作為輸出變數的模型表現。

輸入特徵變數測試組合包括：最近一季財務指標搭配 1、2、3、5 年滾動特徵，以及是否加入上一季報酬率；輸出變數則測試兩種定義。總計形成 16 種變數組合，如表 8 所示。

表 8 輸入變數與輸出變數組合表

| 變數類別  | 輸入特徵變數          |                                 |                                 |                                 |                                 |            | 輸出變數             |                          |
|-------|-----------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|------------|------------------|--------------------------|
| 變數名稱  | 近一季<br>財務指<br>標 | 近一季<br>之前，<br>近 1 年<br>財務指<br>標 | 近一季<br>之前，<br>近 2 年<br>財務指<br>標 | 近一季<br>之前，<br>近 3 年<br>財務指<br>標 | 近一季<br>之前，<br>近 5 年<br>財務指<br>標 | 上一季<br>報酬率 | 持有 1<br>季報酬<br>率 | 持有 1<br>季報酬<br>率<br>PR 值 |
| 組合 1  | ●               |                                 |                                 |                                 | ●                               |            | ●                |                          |
| 組合 2  | ●               |                                 |                                 |                                 | ●                               |            |                  | ●                        |
| 組合 3  | ●               |                                 |                                 |                                 | ●                               | ●          | ●                |                          |
| 組合 4  | ●               |                                 |                                 |                                 | ●                               | ●          |                  | ●                        |
| 組合 5  | ●               |                                 |                                 | ●                               |                                 |            | ●                |                          |
| 組合 6  | ●               |                                 |                                 | ●                               |                                 |            |                  | ●                        |
| 組合 7  | ●               |                                 |                                 | ●                               |                                 | ●          | ●                |                          |
| 組合 8  | ●               |                                 |                                 | ●                               |                                 | ●          |                  | ●                        |
| 組合 9  | ●               |                                 | ●                               |                                 |                                 |            | ●                |                          |
| 組合 10 | ●               |                                 | ●                               |                                 |                                 |            |                  | ●                        |
| 組合 11 | ●               |                                 | ●                               |                                 |                                 | ●          | ●                |                          |
| 組合 12 | ●               |                                 | ●                               |                                 |                                 | ●          |                  | ●                        |
| 組合 13 | ●               | ●                               |                                 |                                 |                                 |            | ●                |                          |
| 組合 14 | ●               | ●                               |                                 |                                 |                                 |            |                  | ●                        |
| 組合 15 | ●               | ●                               |                                 |                                 |                                 | ●          | ●                |                          |
| 組合 16 | ●               | ●                               |                                 |                                 |                                 | ●          |                  | ●                        |

## 第五節 模型選擇與優化

為探究不同模型利用財務指標預測股價報酬的表現，本研究先透過模型評估指標測試隨機森林、XGBoost 及 LSTM 在上述 16 種變數組合下的效果，並選取最佳模型與變數組合後，進一步調整參數優化模型。

## 一、回歸模型評估指標

回歸模型常見評估指標包括均方誤差 (MSE)、平均絕對誤差 (MAE) 及決定係數 ( $R^2$ )，本研究透過驗證集在此三個指標的表現評估模型效能。

### 1. 均方誤差 (Mean Squared Error, MSE)

MSE 為預測值與實際值平方差的平均，對極端值敏感，適合評估模型預測極端值的能力。考慮股市可能出現劇烈漲跌，希望模型準確預測此類報酬率，故選用 MSE 作為主要指標，期望值越小越佳。

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

### 2. 平均絕對誤差 (Mean Absolute Error, MAE)

MAE 為預測值與實際值絕對差的平均，對異常值較不敏感，用於評估模型穩定性。股市中，除非財報表現超出預期導致波動劇烈，多數公司報酬率較為平穩，故以 MAE 與 MSE 交叉驗證模型預測能力，期望值越小越佳。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

### 3. 決定係數 (Coefficient of Determination, $R^2$ )

決定係數表示模型預測變異占總變異的比例，用於衡量模型擬合度，值介於 0 至 1，接近 1 表示預測能力佳，小於 0 則表示不如平均值預測。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

## 二、模型優化

在前節選出最佳模型與變數組合後，本研究進一步透過調整模型超參數，提升模型表現。本節首先說明隨機森林、XGBoost、LSTM 的重要超參數及其對模型的影響，再闡述超參數的調整方法。

### 1. 各模型參數意義

#### (1) 隨機森林(Random Forest)

隨機森林為集成式訓練，以隨機抽樣的方式生成數棵決策樹，每棵決策樹獨立訓練且互不相關，在處理回歸問題時，最終會透過所有樹的平來決定預測結果。其優勢在於穩定性高且不易過擬合，適合處理多維度特徵。

A. `n_estimators`：表示隨機森林中決策樹的數量，預設值為 100。增加決策樹的數量，能夠使模型越穩定，但過多的決策樹可能會增加計算成本且邊際效益遞減，調整範圍為 100~1000。

B. `max_depth`：表示每棵樹的最大深度，預設值為無限制。提升樹的最大深度可以捕捉特徵間更複雜的關係，但過度擬合的風險也會增加。調整範圍為 10~30。

C. `min_samples_split`、`min_samples_leaf`：分前者表示內部節點分裂所需的最小樣本數；後者表示葉節點所需的最小樣本數，預設值分別為 2 與 1。兩參數會影響樹的深淺，調升會使樹的深度變淺，防止模型過度擬合，但也可能導致欠擬合；反之則可能導致模型過度擬合。調整範圍分別為 2~10 與 1~5。

#### (2) XGBoost

XGBoost 與隨機森林同為基於決策樹的學習方法，但 XGBoost 係採用梯度提升法，每顆樹依序學習前一棵樹的殘差，擅長處理非線性關係，相較於隨機森林更具預測能力，同時也較隨機森林更容易過擬合。

- A. `n_estimators`：表示樹的數量，預設值為 100。提高樹的數量可以提升模型複雜度及預測能力，但也可能導致模型過擬合，調整範圍為 100~500。
- B. `learning_rate`：表示每棵樹的學習率，預設值為 0.3。可以透過提升參數值來加速收斂，減少所需樹的數量，但提升參數值可能導致模型欠擬合，調整範圍為 0.1~0.3。
- C. `max_depth`：表示樹的最大深度，預設值為 6。提高樹的最大深度亦能學習較多特徵變數之間的組合關係，來提高模型的預測能力，但也可能導致模型過度擬合，調整範圍為 6~10。
- D. `min_child_weight`：表示子節點所需的最小樣本權重和，預設值為 1。當子節點的樣本權重和小於設定的值時，節點不再分裂，調高參數值可使模型更保守，但可能導致欠擬合，調整範圍為 1~10。

### (3) LSTM

LSTM 為遞迴神經網路之一，常被用來處理有關時間序列的問題，且能夠捕捉時間序列中的長期依賴關係，保留重要的訊息並丟棄無關的細節，適合股市之時序特性。

- A. `units`：表示隱藏單元數，無預設值。單元數量決定模型及易容量，若將值設定過小，可能無法捕捉長期趨勢，反之可能導致過擬合。本研究初始設為 64，調整範圍為 32~256。

- B. dropout：表示隨機丟棄輸入單元的比例，預設值為 0。隨機丟棄一定比例之單元數是為了防止過擬合，調整範圍為 0.2~0.5。
- C. learning\_rate：表示學習率，為優化器的步長，使用 Adam 優化器，預設值為 0.001。係用來控制權重的更新速度和收斂訓練，過高的學習率可能會導致模型跳過最佳解，反之則會導致收斂過慢，調整範圍為 0.0001~0.01。
- D. batch\_size：表示批次大小，無預設值，通常設為 32。為每次訓練迭代中使用的樣本數，調升數值可使訓練加速，但可能使準確度下降，調整範圍為 16~128。
- E. epochs：表示模型遍歷整個數據集的次數，無預設值，通常設為 50。提高訓練的次數可使模型更充分訓練，亦可能導致模型過擬合，可搭配早停機制來結束訓練，調整範圍為 50~200。

## 2. 超參數的調整

超參數為訓練模型之前，需手動設置的參數，不同的參數組合，對模型亦有不同的影響，然而超參數並不會在訓練過程中自動學習優化，因此需要透過其他方式來找到超參數的最佳組合，常見的方法如下，找出最佳參數確定最終模型後，再將模型用於預測未來股票表現，並以測試集回測，評估選股模型的績效表現。

### (1) 網格搜尋(Grid Search)

網格搜尋為每個超參數設定一組數值，窮舉可能組合，訓練並評估每組超參數，確保找到局部最佳解。該雖然簡單直接，但當搜尋範圍大於，將耗費過多運算時間及成本，較適用於參數範圍較小的情境。

## (2) 隨機搜尋(Random Search)

隨機搜尋為每個超參數設定一個範圍，並透過隨機的方式，在範圍內選擇  $n$  組超參數，適合初步探索參數空間，雖然未必保證能找到全域最佳解，但能快速縮小範圍，且耗費的時間和成本較網格搜尋小。

## (3) 貝葉斯優化(Bayesian Optimization)

貝葉斯優化與傳統網格搜尋及隨機搜尋不同，是一種用於優化黑盒函數的方法，首先會隨機選擇幾個參數點，並計算目標值作為初始數據，透過概率模型（通常為高斯過程）根據已知的點擬合一個猜測模型，並使用採集函數來決定下一個測試的參數點，挑選高潛力的點，並不斷的更新與迭代直到到達預定試驗次數或目標值收斂，適合計算資源有限時進行調參。

# 第六節 投資組合績效的評估方法

機器學習模型建立後，將用於預測各家企業未來股價表現並建構投資策略，故先利用測試集進行回測。投資策略為將模型預測報酬率排名，選取每季前  $n$  名股票建構投資組合，並於每季再平衡。若次季預測報酬率跌出前  $n$  名，則剔除；若仍在前  $n$  名，則繼續持有。

回測方法與傳統量化投資策略一致，採用 `yfinance` 套件下載調整後收盤價計算報酬率，而投資組合績效除報酬率外，亦須考量風險，常用指標包括總報酬率、年化報酬率、標準差、年化波動率、最大回撤及夏普比率。本研究以模型於驗證集與測試集之預測結果進行回測。

### 1. 總報酬率

總報酬率即為投資組合在回測期間內之總報酬，反映了投資入資金的成長幅度，可以快速知道執行此投資策略共為投資人帶來多少盈虧，總報酬率的數值越大，投資組合的績效越佳。

$$\text{總報酬率} = \text{投資組合現值} \div \text{投資成本} - 1$$

## 2. 年化報酬率

由於總報酬率並不考慮時間價值，當投資時間長度不同時，難以判斷策略的優劣，因此加入年化報酬率來判斷投資人每年平均的盈虧，此外，年化報酬率與平均報酬率之差異在於，平均報酬率並無考慮複利的影響，計算出的報酬率通常較高，未避免投資人高估報酬率，並讓投資人可以自行判斷不同時間長度的投資策略，故使用之，年化報酬率的數值越大，投資組合的績效越佳。

$$\text{年化報酬率} = \sqrt[n]{(1 + \text{報酬率})} - 1$$

## 3. 標準差

若只關注報酬率而未了解其投資風險，會使投資人過度樂觀導致決策錯誤，所以除了揭露投資組合的投資報酬率外，亦須評估其風險程度，而標準差是常被用來衡量投資組合波動度的指標，數值越小，表示投資組合的風險越小。

$$\text{標準差} = \frac{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}{n}$$

## 4. 年化波動度

由於標準差為衡量每次投資時的波動程度，若不同策略的投資頻率或時間長短不同，則難以比較不同策略的優劣，因此將標準差年化以便比較。

$$\text{年化波動度} = \text{標準差} \times \sqrt{\text{一年交易次數}}$$

## 5. 最大回檔(Max Drawdown)

在投資期間，回檔是無法避免的，平常的偶爾小回檔大多數人皆可以接受，倘若較大的回檔，可能會造成投資人恐慌而做出錯誤的決定，因此最大回檔要讓投資人了解在投資期間可能會發生的最大虧損，以評估自身對風險的承受能力，最大回檔指的是在回測期間內，投資組合的淨值從最高點回落至最低點的幅度。

$$MDD = \max_{t \in T} \left( \frac{\max_{0 \leq s \leq t} (P_s) - P_t}{\max_{0 \leq s \leq t} (P_s)} \right)$$

## 6. 夏普比率(Sharpe Ratio)

在評估投資策略的優劣時，不但要評估報酬率，又要評估風險程度，因為夏普比率是在衡量每承受一單位的風險可以獲得多少報酬，此時，利用夏普比率係一個簡單快速的方式來判斷不同策略的優劣，數值越大，表示承受一單位的風險可以獲得較高的報酬。

$$\text{Sharpe Ratio} = \frac{\text{年化報酬率}}{\text{年化波動度}}$$

## 第四章 研究結果

本研究在第一節先闡述各模型在驗證集的表現，並說明將選擇何種模型做訓練集優化。第二節說明最終選擇的模型集優化的結果。第三節將參考 (Piotroski, 2000) 所提出之 F-score 以建構季調整之投資策略，下稱為「調整後 F-score」，說明此投資策略的回測績效。第四節則說明如何利用本研究所訓練之模型建構投資策略，並與「調整後 F-score」比較。最終再回測結合「調整後 F-score」與本研究所訓練之模型，是否有更優秀的績效。

### 第一節 不同模型與變數組合之實驗結果比較

本研究分別利用 XGBoost、LSTM 及隨機森林，並測試第三章第四節中表 8 的 16 種輸入與輸出變數組合。為便於說明，以下將這 16 種變數組合依輸入特徵變數與輸出變數之設計劃分為四種情境，如表 9 所示。

表 9 模型訓練情境表

|     | 輸出入變數組合說明           | 對應至表 7 組合   |
|-----|---------------------|-------------|
| 情境一 | 輸出變數為持有 1 季報酬率      | 組合 1、組合 5   |
|     | 特徵變數無加入上季報酬率        | 組合 9、組合 13  |
| 情境二 | 輸出變數為持有 1 季報酬率      | 組合 2、組合 6   |
|     | 特徵變數加入上季報酬率         | 組合 10、組合 14 |
| 情境三 | 輸出變數為持有 1 季報酬率 PR 值 | 組合 3、組合 7   |
|     | 特徵變數無加入上季報酬率        | 組合 11、組合 15 |
| 情境四 | 輸出變數為持有 1 季報酬率 PR 值 | 組合 4、組合 8   |
|     | 特徵變數加入上季報酬率         | 組合 12、組合 16 |

#### 一、XGBoost

XGBoost 模型在驗證集的表現如表 10 所示。從結果觀察，在均方誤差 (MSE) 及平均絕對誤差 (MAE) 的表現上，情境一、情境二的表現皆優於情境三、情境四。

MSE 在情境一、二與情境三、四的值分別為 0.0392~0.0396 與 0.0819~0.0832；MAE 則分別為 0.1216~0.1228 與 0.2466~0.2493。最小的輸入特徵組合為組合 13 與 14，也就是將輸出變數設定為持有 1 季報酬率，並以過去 1 年的財務資料做為輸入特徵變數（情境一與二）。其中，MSE 與 MAE 分別為 0.0392 與 0.1216，顯示模型在數值預測上的表現相對穩定。然而，該組合的決定係數 ( $R^2$ ) 卻為負值，顯示模型在解釋報酬率變異方面的效果不佳。

另一方面，決定係數 ( $R^2$ ) 則以情境三、四的表現優於情境一、二。最佳的特徵組合為組合 16，即以持有 1 季報酬率的 PR 值作為輸出變數，並以過去 1 年的財務資料結合上一季報酬率作為輸入特徵（情境四）。此組合的  $R^2$  為 0.0173，顯示模型在解釋報酬率變異方面表現較佳。雖然其數值預測的穩定性略遜於情境一和情境二，但在相同情境下，其 MSE 和 MAE 值同樣為最低。

整體而言，XGBoost 模型採用過去 1 年的財務資料作為特徵變數進行預測時，驗證集表現優於使用過去 2 年、3 年或 5 年的財務資料，表明短期財務資料可能更能捕捉影響報酬率的關鍵變異。

表 10 XGBoost 驗證集表現評估表

| 利用過去 5 年資料預測(組合 1、2、3、4)     |         |         |        |         |
|------------------------------|---------|---------|--------|---------|
| 評估指標                         | 情境一     | 情境二     | 情境三    | 情境四     |
| 驗證集 MSE                      | 0.0393  | 0.0394  | 0.0832 | 0.0825  |
| 驗證集 MAE                      | 0.1225  | 0.1224  | 0.2493 | 0.2485  |
| 驗證集 R2                       | -0.0083 | -0.0108 | 0.0021 | 0.0095  |
| 利用過去 3 年資料預測(組合 5、6、7、8)     |         |         |        |         |
| 評估指標                         | 情境一     | 情境二     | 情境三    | 情境四     |
| 驗證集 MSE                      | 0.0394  | 0.0396  | 0.0830 | 0.0824  |
| 驗證集 MAE                      | 0.1228  | 0.1223  | 0.2486 | 0.2473  |
| 驗證集 R2                       | -0.0109 | -0.0145 | 0.0037 | 0.0114  |
| 利用過去 2 年資料預測(組合 9、10、11、12)  |         |         |        |         |
| 評估指標                         | 情境一     | 情境二     | 情境三    | 情境四     |
| 驗證集 MSE                      | 0.0393  | 0.0396  | 0.0828 | 0.0822  |
| 驗證集 MAE                      | 0.1223  | 0.1223  | 0.2485 | 0.2476  |
| 驗證集 R2                       | -0.0065 | -0.0153 | 0.0061 | 0.0138  |
| 利用過去 1 年資料預測(組合 13、14、15、16) |         |         |        |         |
| 評估指標                         | 情境一     | 情境二     | 情境三    | 情境四     |
| 驗證集 MSE                      | 0.0392* | 0.0392* | 0.0828 | 0.0819  |
| 驗證集 MAE                      | 0.1222  | 0.1216* | 0.2483 | 0.2466  |
| 驗證集 R2                       | -0.0042 | -0.0034 | 0.0063 | 0.0173* |

## 二、LSTM

LSTM 模型在驗證集的表現如表 11 所示。同 XGBoost 的實驗結果，在均方誤差 (MSE) 及平均絕對誤差 (MAE) 的表現上，情境一、情境二的結果亦優於情境三、情境四。MSE 在情境一與二的範圍為 0.0390~0.0391，情境三與四則為 0.0833~0.0834；MAE 分別為 0.1188~0.1218 與 0.2498~0.2500。最佳的特徵組合為組合 6，即將輸出設定為持有一季報酬率，並使用過去三年財務資料加上上一季報酬率為特徵 (情境二)，該組合的 MAE 為 0.1188，為全體組合中最低。

另一方面，決定係數 ( $R^2$ ) 則全體偏低，多數組合接近 0 或呈現微幅負值，顯示模型在預測值與實際值的變異趨勢上仍難以準確掌握。表現最佳的  $R^2$  出現在情境三的組合 12，為 0.0006，雖優於其他組合，但整體解釋力仍有限。

整體而言，LSTM 模型在 MSE 與 MAE 表現上相對穩定，尤其在加入上一季報酬率後，部分組合略有改善，顯示 LSTM 對於時間序列特徵具一定敏感度。然而其  $R^2$  整體表現不如 XGBoost，表示對報酬率變異的掌握力仍有限。相較之下，XGBoost 不僅在預測誤差指標上稍具優勢，其於情境四所對應之組合 16 亦能達到最高的  $R^2$  (0.0173)，顯示其整體預測能力與資料解釋力皆優於 LSTM。

表 11 LSTM 驗證集表現評估表

| 利用過去 5 年資料預測(組合 1、2、3、4)     |         |         |         |         |
|------------------------------|---------|---------|---------|---------|
| 評估指標                         | 情境一     | 情境二     | 情境三     | 情境四     |
| 驗證集 MSE                      | 0.0390  | 0.0390  | 0.0834  | 0.0833  |
| 驗證集 MAE                      | 0.1200  | 0.1199  | 0.2500  | 0.2500  |
| 驗證集 R2                       | -0.0002 | -0.0003 | -0.0004 | -0.0001 |
| 利用過去 3 年資料預測(組合 5、6、7、8)     |         |         |         |         |
| 評估指標                         | 情境一     | 情境二     | 情境三     | 情境四     |
| 驗證集 MSE                      | 0.0390  | 0.0390  | 0.0834  | 0.0833  |
| 驗證集 MAE                      | 0.1218  | 0.1188* | 0.2499  | 0.2499  |
| 驗證集 R2                       | -0.0004 | 0.0001  | -0.0003 | 0.0003  |
| 利用過去 2 年資料預測(組合 9、10、11、12)  |         |         |         |         |
| 評估指標                         | 情境一     | 情境二     | 情境三     | 情境四     |
| 驗證集 MSE                      | 0.0391  | 0.0391  | 0.0833  | 0.0833  |
| 驗證集 MAE                      | 0.1205  | 0.1213  | 0.2499  | 0.2498  |
| 驗證集 R2                       | -0.0010 | -0.0011 | 0.0006* | 0.0003  |
| 利用過去 1 年資料預測(組合 13、14、15、16) |         |         |         |         |
| 評估指標                         | 情境一     | 情境二     | 情境三     | 情境四     |
| 驗證集 MSE                      | 0.0390  | 0.0390  | 0.0833  | 0.0833  |
| 驗證集 MAE                      | 0.1205  | 0.1197  | 0.2499  | 0.2499  |
| 驗證集 R2                       | 0.0003  | 0.0003  | 0.0003  | 0.0004  |

### 三、隨機森林

隨機森林模型在驗證集的表現如表 12 所示。就誤差指標而言，情境一與二之 MSE 為 0.0421~0.0447，明顯低於情境三與四的 0.0821~0.0833；MAE 部分情境一與二介於 0.1288~0.1411，優於情境三與四的 0.2467~0.2489。數值預測表現最佳的組合為組合 14，亦即使用過去一年財務資料與上一季報酬率預測持有一季報酬率（情境二），其 MSE 為 0.0421，MAE 為 0.1288。

在解釋力方面，情境一、二之  $R^2$  整體為負值，但在將輸出變數調整為持有 1 季報酬率 PR 值(情境三、四)後，有效提升  $R^2$ ，由其以組合 8 表現最佳， $R^2$  達 0.0147(情境四)，即以持有一季報酬率 PR 值為輸出，並加入三年財務資料與上一季報酬率為輸入特徵。儘管此數值不高，但已優於隨機森林其他組合，顯示其對於報酬率排序具備一定的辨識能力。

整體而言，隨機森林模型在相同情境下驗證集的均方誤差（MSE）與平均絕對誤差（MAE）雖略高於 XGBoost 和 LSTM 模型，但差距並不顯著。在情境三與情境四的變數組合中，隨機森林對報酬率變異的解釋能力優於 LSTM 模型，尤其在情境四的組合 8，其  $R^2$ （0.0147）僅略低於 XGBoost 的最佳表現（0.0173）。

表 12 隨機森林驗證集表現評估表

| 利用過去 5 年資料預測(組合 1、2、3、4)     |         |         |        |         |
|------------------------------|---------|---------|--------|---------|
| 評估指標                         | 情境一     | 情境二     | 情境三    | 情境四     |
| 驗證集 MSE                      | 0.0447  | 0.0440  | 0.0831 | 0.0826  |
| 驗證集 MAE                      | 0.1411  | 0.1359  | 0.2489 | 0.2473  |
| 驗證集 R2                       | -0.1466 | -0.1274 | 0.0023 | 0.0093  |
| 利用過去 3 年資料預測(組合 5、6、7、8)     |         |         |        |         |
| 評估指標                         | 情境一     | 情境二     | 情境三    | 情境四     |
| 驗證集 MSE                      | 0.0430  | 0.0429  | 0.0830 | 0.0821  |
| 驗證集 MAE                      | 0.1359  | 0.1324  | 0.2485 | 0.2468  |
| 驗證集 R2                       | -0.1024 | -0.0983 | 0.0035 | 0.0147* |
| 利用過去 2 年資料預測(組合 9、10、11、12)  |         |         |        |         |
| 評估指標                         | 情境一     | 情境二     | 情境三    | 情境四     |
| 驗證集 MSE                      | 0.0429  | 0.0423  | 0.0825 | 0.0823  |
| 驗證集 MAE                      | 0.1349  | 0.1314  | 0.2474 | 0.2467  |
| 驗證集 R2                       | -0.0994 | -0.0850 | 0.0104 | 0.0125  |
| 利用過去 1 年資料預測(組合 13、14、15、16) |         |         |        |         |
| 評估指標                         | 情境一     | 情境二     | 情境三    | 情境四     |
| 驗證集 MSE                      | 0.0423  | 0.0421* | 0.0833 | 0.0825  |
| 驗證集 MAE                      | 0.1325  | 0.1288* | 0.2479 | 0.2467  |
| 驗證集 R2                       | -0.0831 | -0.0787 | 0.0007 | 0.0101  |

透過實驗發現，將「持有 1 季報酬率 PR 值」設定為輸出變數時，三種模型對於報酬率的變異皆有較好的解釋能力，其中 XGBoost 模型利用過去一年的財務指標，加上上季報酬率來預測報酬率 PR 值時，可得到最高的  $R^2$  值為 0.0173；其次是隨機森林模型利用過去三年的財務指標並加入上季報酬率預測， $R^2$  值為 0.0147；最後是 LSTM 模型利用過去兩年財務指標來預測， $R^2$  值為 0.0006。

考量其餘變數組合  $R^2$  表現不佳，顯示模型對於報酬率的變異解釋能力不佳，因預測結果不如預期，回測結果亦不具參考價值。另為瞭解此三個模型在  $R^2$  表現最佳的變數組合之投資效果，進一步利用三個模型在驗證集的預測結果進行績效回測，報酬率表現如表 13。透過報酬率可得知以變數組合 8 並採用隨機森林模型預測，除了可以有較良好的預測能力，在驗證集的績效亦為三個模型中最佳。

表 13 不同模型於最佳變數組合下之驗證集報酬率比較表

|        | 隨機森林模型<br>變數組合 8 | XGBoost 模型<br>變數組合 16 | LSTM 模型<br>變數組合 11 |
|--------|------------------|-----------------------|--------------------|
| 驗證集報酬率 | 29%              | 19%                   | 12%                |

為了與元大台灣 50 ETF（下稱 0050）比較，本研究將模型預測結果由大至小排序，並選出前 50 名建構投資組合，投資組合淨值走勢如圖 6。從走勢圖可看出，透過三個模型所建構的投資組合，在驗證集的累積淨值皆優於 0050 淨值，其中又以利用隨機森林建構而成的投資組合最終累積淨值最高，且抗跌能力亦優於 LSTM 模型及 0050；其次為 XGBoost 模型，雖然在 2022 年 3 月前淨值成長小於 LSTM 模型，但也可看出在整體股市下修時，比 LSTM 模型及 0050 更有抗跌的能力。綜上分析，本研究最終選擇隨機森林模型來做優化。

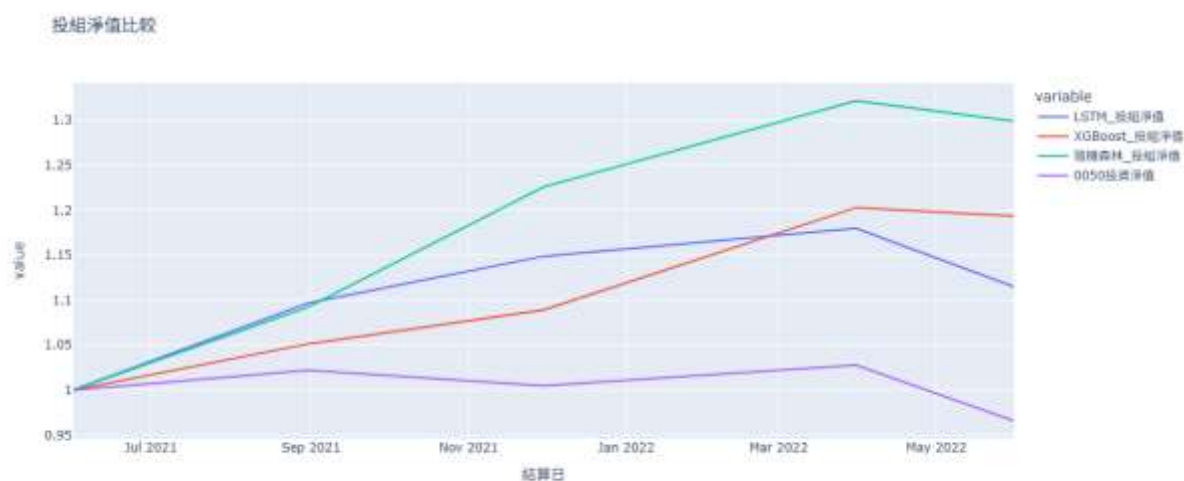


圖 6 各模型投組淨值走勢圖

## 第二節 隨機森林模型優化後表現

透過各模型測試各種變數組合，可發現提升模型決定係數 ( $R^2$ ) 除了能使提升模型解釋力，亦能提高投資組合績效表現，因此在優化模型時，以提升  $R^2$  為目標。本研究希望透過調整模型超參數提升預測能力，調整的超參數如表 14。為了以較有效率的方式找出超參數組合，本研究採用貝葉斯優化(Bayesian Optimization)來找出最佳超參數組合。

表 14 隨機森林參數表

| 超參數名稱             | 說明                    | 調整範圍     |
|-------------------|-----------------------|----------|
| n_estimators      | 決策樹的數量，預設值為 100       | 100~1000 |
| max_depth         | 每棵樹的最大深度，預設值為 None    | 10~30    |
| min_samples_split | 內部節點分裂所需的最小樣本數，預設值為 1 | 2~50     |
| min_samples_leaf  | 葉節點所需的最小樣本數，預設值為 2    | 1~10     |

本研究以  $R^2$  最大化為目標進行貝葉斯優化，鑒於 Google Colab 執行 Python 的環境限制，首先隨機測試 2 組超參數組合後再進行優化，經過 20 組測試後，得到較佳的超參數組合如表 15，最後利用此組超參數組合訓練模型，並觀察模型在測試集的表現。

表 15 最佳超參數組合表

| target  | max_depth | min_samples_leaf | min_samples_split | n_estimators |
|---------|-----------|------------------|-------------------|--------------|
| 0.01887 | 10        | 9                | 49                | 547          |

模型經過優化後，成功提升了驗證集的表現，差異如表 16。驗證集的 MSE、MAE 皆有些微的下降，並提升了  $R^2$ ，進一步觀察優化後模型在測試集的表現，3 項評估指標與驗證集並無明顯落差，可見模型在預測誤差上與驗證集一致，說明模型具有良好的泛化能力，雖然  $R^2$  對比驗證集略有下降，但差距不大且亦較優化前模型表現良好，顯示模型保持相對穩定的擬合效果。考量貝葉斯優化經過 20 次迭代出現最佳的超參數組合後，已逾 10 次迭代未出現更好的超參數組合，因此暫時停止模型的優化，並先以表 15 之超參數組合建構模型，評估前後績效差異。

表 16 模型優化後差異表

| 評估指標      | 優化後    | 優化前    | 差異      |
|-----------|--------|--------|---------|
| 驗證集 MSE   | 0.0816 | 0.0821 | -0.0005 |
| 驗證集 MAE   | 0.2467 | 0.2468 | -0.0001 |
| 驗證集 $R^2$ | 0.0210 | 0.0147 | 0.0063  |
| 測試集 MSE   | 0.0818 |        |         |
| 測試集 MAE   | 0.2468 |        |         |
| 測試集 $R^2$ | 0.0182 |        |         |

利用驗證集與測試集的預測結果進行績效回測之結果如表 17。建構投資組合的方式為取模型預測值前 50 名、30 名、20 名或 10 名作為投資標的，並於每季依照模型預測值重新調整標的及平均分配投資比例。從表中可看出優化後模型的表現皆較優化前有更高的報酬率，雖然在波動度的部分較高，以至於投資 50 檔標的及投資 30 檔標的之夏

普值不如模型優化前，但優化後模型在降低每季投資標的數量後，卻能獲得更高的報酬率。

表 17 模型優化前後投組績效差異表

| <div> <div>投組名稱</div> <div>評估指標</div> </div> | 元大台灣 50 ETF           | 優化前<br>投組<br>_50 檔 | 優化後<br>投組<br>_50 檔 | 優化前<br>投組<br>_30 檔 | 優化後<br>投組<br>_30 檔 | 優化前<br>投組<br>_20 檔 | 優化後<br>投組<br>_20 檔 | 優化前<br>投組<br>_10 檔 | 優化後<br>投組<br>_10 檔 |
|----------------------------------------------|-----------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 回測期間                                         | 2021-06-01~2024-05-31 |                    |                    |                    |                    |                    |                    |                    |                    |
| 期間(年)                                        | 3                     |                    |                    |                    |                    |                    |                    |                    |                    |
| 每季平均報酬率(%)                                   | 2.57                  | 6.92               | 7.04               | 6.97               | 8.04               | 6.26               | 7.88               | 6.53               | 8.91               |
| 年化報酬率(%)                                     | 9.53                  | 30.29              | 30.82              | 30.47              | 35.45              | 27.05              | 34.53              | 28.19              | 38.39              |
| 累積報酬(%)                                      | 131.39                | 221.19             | 223.88             | 222.09             | 248.51             | 205.08             | 243.48             | 210.64             | 265.02             |
| 標準差(%)                                       | 7.62                  | 4.48               | 4.83               | 4.95               | 6.32               | 4.68               | 6.62               | 5.54               | 10.19              |
| 年化波動率(%)                                     | 15.25                 | 8.96               | 9.67               | 9.89               | 12.64              | 9.35               | 13.25              | 11.08              | 20.39              |
| Max Drawdown                                 | -15.59                | -1.69              | -0.91              | -2.41              | -2.37              | -2.13              | -3.40              | -4.25              | -8.29              |
| Sharpe Ratio                                 | 0.62                  | 3.38               | 3.19               | 3.08               | 2.80               | 2.89               | 2.61               | 2.54               | 1.88               |

從圖 7 亦可看出利用優化後的模型所建構的投資組合（綠線）與較優化前（紅線）之淨值變化，在回測期間優化後模型之投組皆較優化前有更高的成長率，因此本研究最終將利用優化後的模型來建構長期的投資策略。

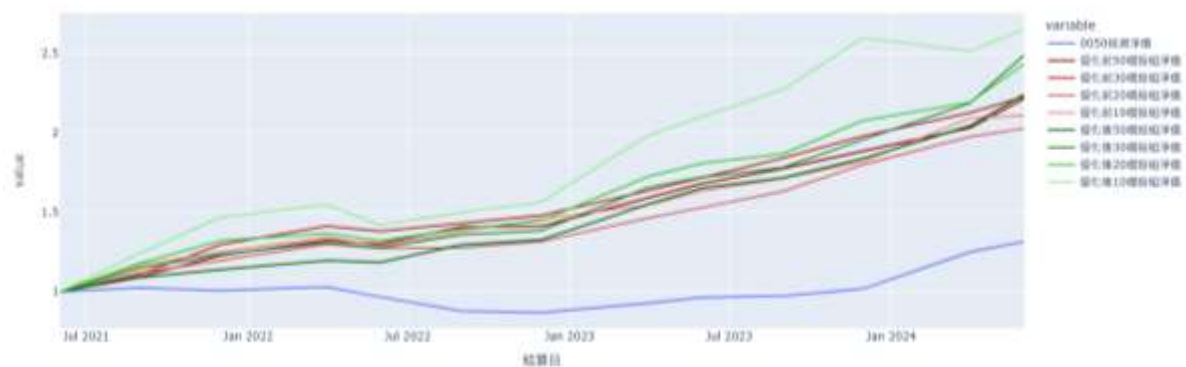


圖 7 優化前後投組淨值走勢圖

### 第三節 調整後 F-score 投資策略

本研究調整 F-score 投資策略，建構季調整投資組合配置的策略，另因原始 F-score 屬價值投資策略，故將當季本淨比為最低的前 25% 為選股條件之一，考量若將「當季本淨比為最低的前 25%」列為選股條件之一，可能會造成許多權值股被篩除。因此本研究將分別實驗有無篩選本淨比最低的前 25% 執行績效回測。調整後 F-score 投資策略一為篩選 F-score 總分大於等於 8 分的企業；調整後 F-score 投資策略二為篩選 F-score 總分大於等於 8 分的企業，並剔除當季本淨比排名超過前 25% 的標的以建構投資組合建構投資組合。

#### 一、投資策略一（未篩選本淨比較低的公司）

此投資策略首先於 2014 年第一季財報公布 (2014/5/31) 後，選出調整後 F-score 分數 8、9 分的公司組成投資組合，並於隔日買進所有標的，再於下一季財報公布 (2014/8/31) 後調整及再平衡投資組合，重複調整投資組合至回測區間結束 (2024/5/31)，即為 2024 第一季財報公布前，以下稱為投資策略一。在回測期間，符合投資策略一選股條件之標的共有 1,297 家上市櫃公司，上市櫃公司數量如圖 8 以上市公司居多，數量為 757 家，而各產業數量分布如圖 9，以電子零組件業最多，

公司數接近 180 家，其次為半導體產業數量接近 120 家。投資策略一每季從 1,297 家上市櫃公司中選出當季 F-score 分數大於 7 分之標的以建構投資組合。

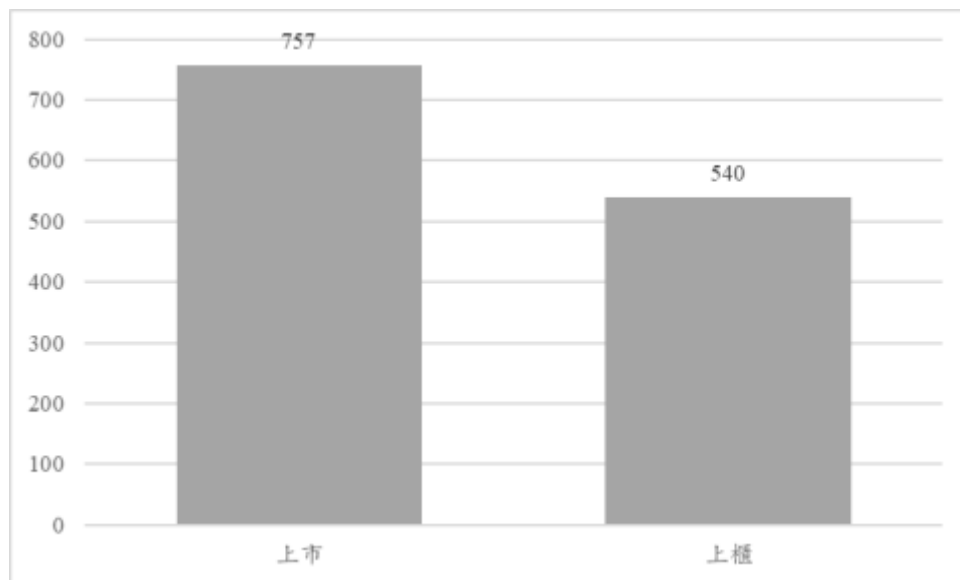


圖 8 投資策略一之選股池上市櫃公司數量圖

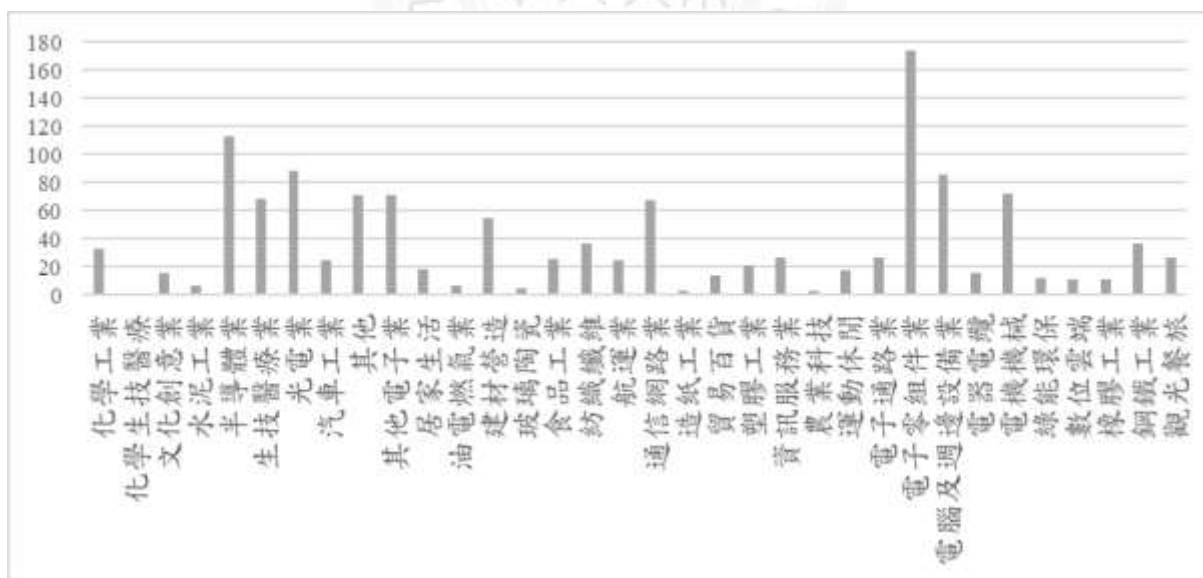


圖 9 投資策略一之選股池各產業公司數量圖

圖 10 為投資策略一各季度入選股票所屬產業分佈狀況，顏色越深表示該產業於該季入選的家數越多，從圖中可看出電子零組件業、半導體業、光電業為入選頻率及家數

最多之產業，其中又以電子零組件業佔比最多，應為該類產業之公司財報較常符合投資策略一之評分標準所致。

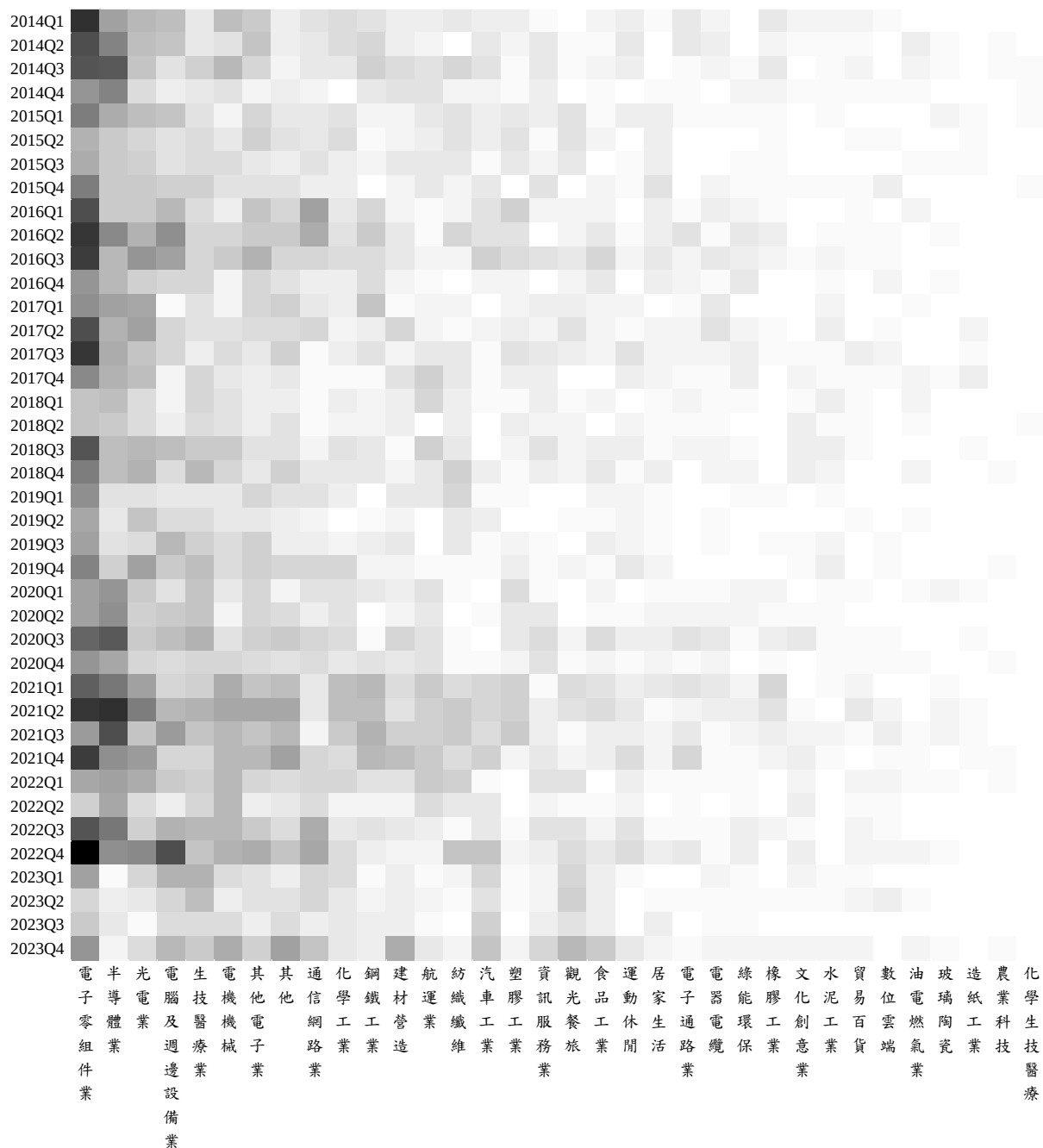


圖 10 投資策略一各季產業分布熱力圖

投資策略一之投資績效如表 18，勝率表示投資組合每季報酬率為正的機率，投資策略一在回測區間內，每季有 72%的機率可獲得正報酬率。每季平均報酬率約

為 4.67%，年化報酬率為 18.57%，皆優於 0050 之獲利，最終投組淨值成長至 549%。但在標準差、年化波動度及 Max Drawdown 皆較 0050 差，即投資策略一必須承受較多的風險。但整體而言，投資策略一的 Sharpe Ratio 約為 1.14，較 0050 來的高，表示平均每承擔一單位的風險投資策略一卻能有較多的回報。

表 18 調整後 F score 投資策略一之績效評估表

| <div> <div>投組名稱</div> <div>評估指標</div> </div> | 元大台灣 50 ETF           | 調整後 F score 未篩低本淨比    |
|----------------------------------------------|-----------------------|-----------------------|
| 回測期間                                         | 2014-06-01~2024-05-31 | 2014-06-01~2024-05-31 |
| 期間(年)                                        | 10                    | 10                    |
| 每季平均投資家數                                     | 50                    | 140                   |
| 勝率                                           | 67.50%                | 72.50%                |
| 每季平均報酬率                                      | 3.58%                 | 4.67%                 |
| 年化報酬率                                        | 13.71%                | 18.57%                |
| 累積報酬                                         | 361.46%               | 549.37%               |
| 標準差                                          | 8.13%                 | 8.18%                 |
| 年化波動率                                        | 16.26%                | 16.35%                |
| Max Drawdown                                 | -15.59%               | -19.37%               |
| Sharpe Ratio                                 | 0.84                  | 1.14                  |

投資策略一在回測期間內的淨值變化如圖 11，從圖中可看出，雖然投資策略一之整體績效表現較 0050 佳，但在 2021 年 3 月 31 日（即 2020Q4 財報公布日）前僅略優於 0050，走勢幾乎與其相同，甚至在 2017 年 11 月 30 日前之績效略輸於 0050。而投資策略一每季平均必須投資高達 140 家公司才能達到此績效，實務上一般投資人較難以操作。



圖 11 投資策略一與 0050 之淨值走勢圖

## 二、投資策略二（篩選本淨比較低的公司）

此投資策略首先於 2014 年第一季財報公布 (2014/5/31) 後，選出調整後 F-score 分數 8、9 分的公司後，再挑選當季本淨比最低的前 25% 之公司組成投資組合，並於隔日買進所有標的，再於下一季財報公布 (2014/8/31) 後調整及再平衡投資組合，重複調整投資組合至回測區間結束 (2024/5/31)，即為 2024 第一季財報公布前，以下稱為投資策略二。在回測期間，符合投資策略二選股條件之標的共有 453 家上市櫃公司，上市櫃公司數量如圖 12 以上市公司居多，數量為 312 家，而各產業數量分布如圖 13，以電子零組件業最多，公司數接近 80 家，其次為光電業數量接近 60 家。投資策略二每季從 453 家上市櫃公司中選出當季 F-score 分數大於 7 分之標的以建構投資組合。

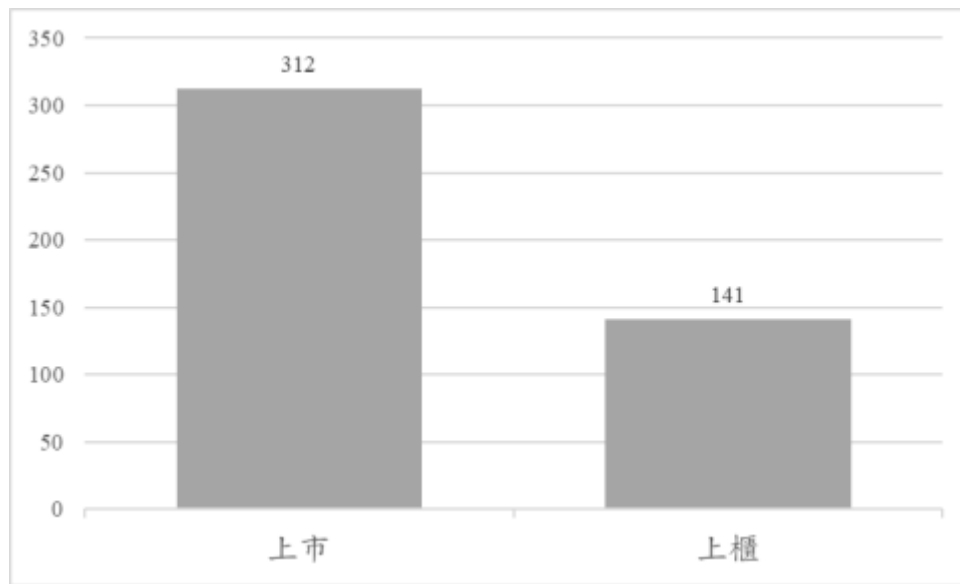


圖 12 投資策略二之選股池上市櫃公司數量圖

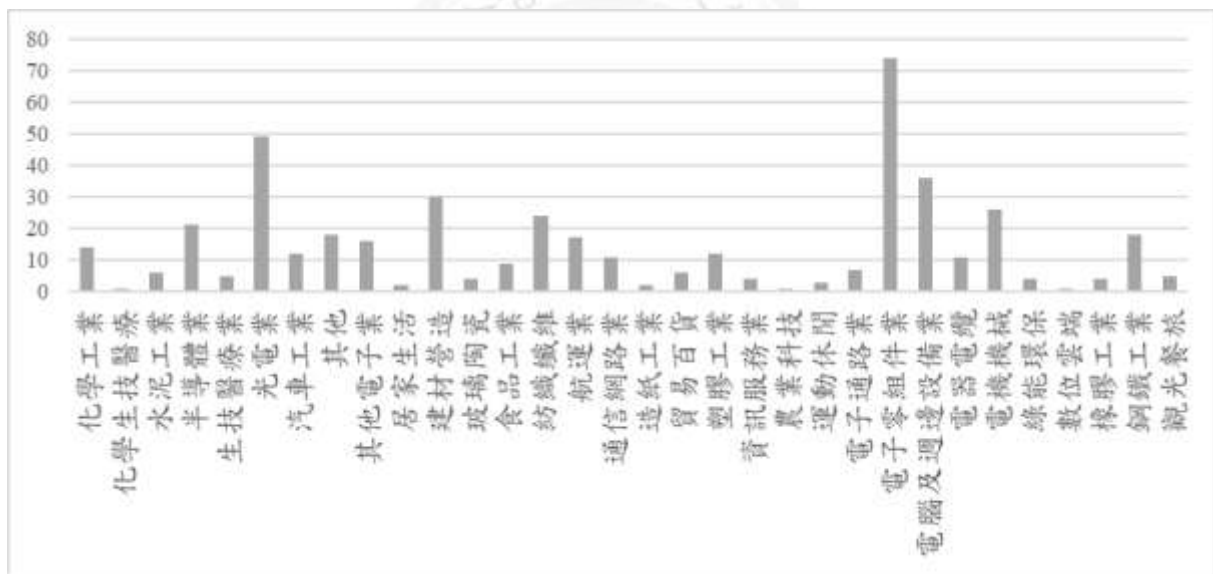


圖 13 投資策略二之選股池各產業公司數量圖

圖 14 為投資策略二各季度入選股票所屬產業分佈狀況，顏色越深表示該產業於該季入選的家數越多，從圖中可看出分佈狀況與投資策略一類似，電子零組件業、光電業亦為入選頻率及家數最多之產業，其中又以電子零組件業佔比最多，應為該類產業之公

司財報較常符合投資策略一之評分標準，另外部分產業如航運業等公司可能具較低的本淨比，符合投資策略二之選股邏輯，因此入選頻率提高。

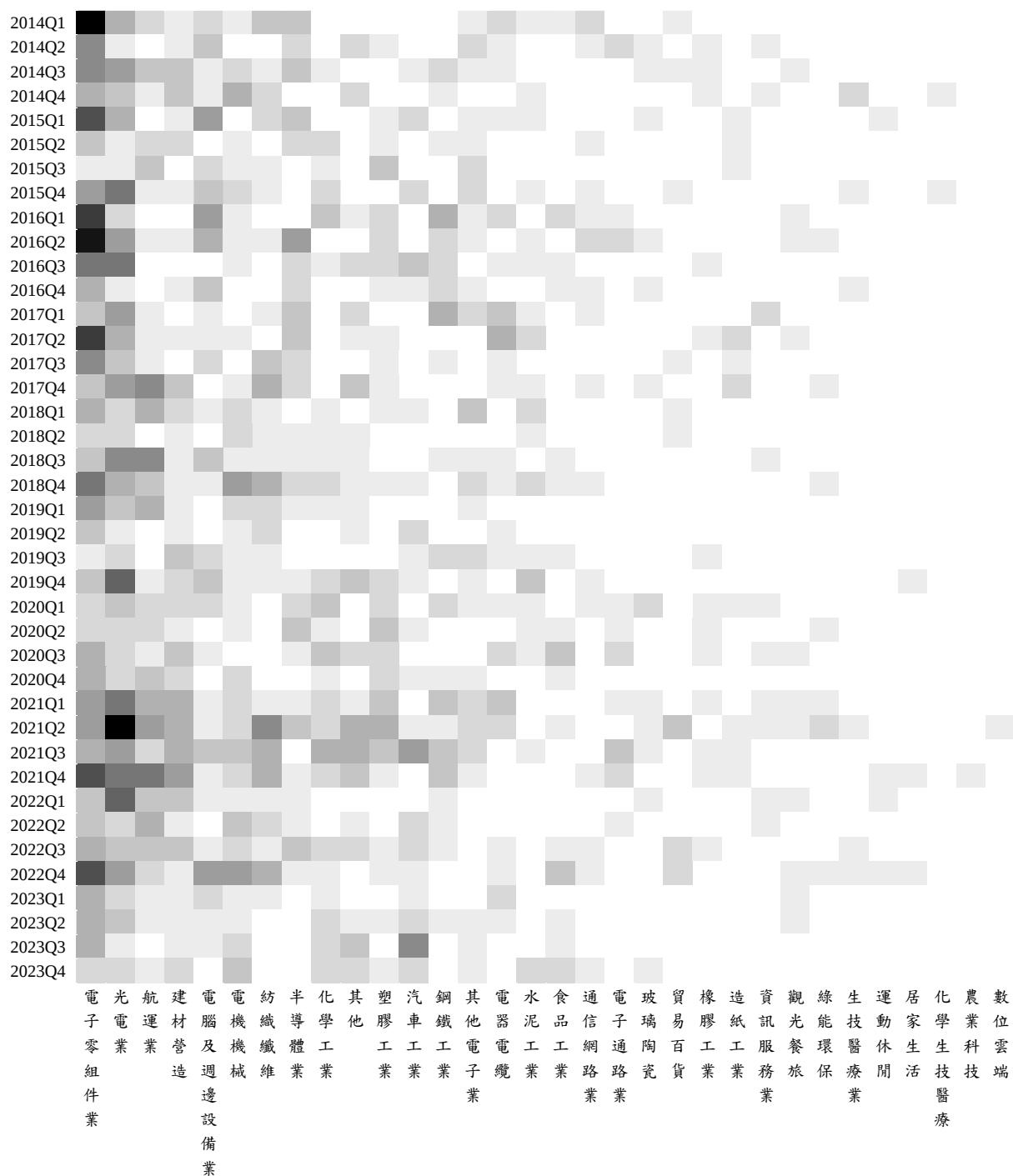


圖 14 投資策略二各季產業分布熱力圖

投資策略二之投資績效如表 19，投資策略二在回測區間內與投資策略一相同，每季有 72% 的機率可獲得正報酬率。投資策略二每季平均報酬率約為 4.87%，年化報酬率為 18.91%，除優於 0050 之獲利能力外，亦略勝於投資策略一，最終投組淨值成長至 565%。但在標準差、年化波動度及 Max Drawdown 皆較 0050 及投資策略一差，即投資策略二必須承受較多的風險。但整體而言，投資策略二的 Sharpe Ratio 約為 0.99，雖然高於 0050 的 0.84，但低於投資策略一的 1.14，平均每承擔一單位的風險，三種投資策略的報酬高低為投資策略一 > 投資策略二 > 0050。

表 19 調整後 F score 投資策略二之績效評估表

| 投組名稱<br>評估指標 | 元大台灣 50 ETF           | 調整後 F score 挑選低本淨比    |
|--------------|-----------------------|-----------------------|
| 回測期間         | 2014-06-01~2024-05-31 | 2014-06-01~2024-05-31 |
| 期間(年)        | 10                    | 10                    |
| 每季平均投資家數     | 50                    | 30                    |
| 勝率           | 67.50%                | 72.50%                |
| 每季平均報酬率      | 3.58%                 | 4.87%                 |
| 年化報酬率        | 13.71%                | 18.91%                |
| 累積報酬         | 361.46%               | 565.37%               |
| 標準差          | 8.13%                 | 9.52%                 |
| 年化波動率        | 16.26%                | 19.05%                |
| Max Drawdown | -15.59%               | -20.10%               |
| Sharpe Ratio | 0.84                  | 0.99                  |

投資策略二在回測期間內的淨值變化如圖 15，從圖中可看出，雖然投資策略二之整體績效表現較 0050 佳，但在 2021 年 8 月 31 日（即 2021Q2 財報公布日）前，走勢幾乎與 0050 相同，且績效表現略輸 0050 之表現，可見投資策略二在特定區間績效表現較 0050 弱。而投資策略而每季平均必須投資標的數雖然降低至 30 家公司，但標的數量的選擇不具彈性，無法供投資人依風險承受度或資金量體選擇要投入的標的數量。



圖 15 投資策略二與 0050 之淨值走勢圖

## 第四節 機器學習投資策略之績效

本節將採用本研究最終選定之模型，預測 2021 年至 2023 年間台灣上市櫃公司每季的報酬率表現，回測期間模型共挑出 343 家企業，上市櫃公司數量如圖 16，分別為 225 家、118 家企業，產業分布如圖 17，在所有產業中以電子零組件業占多數。

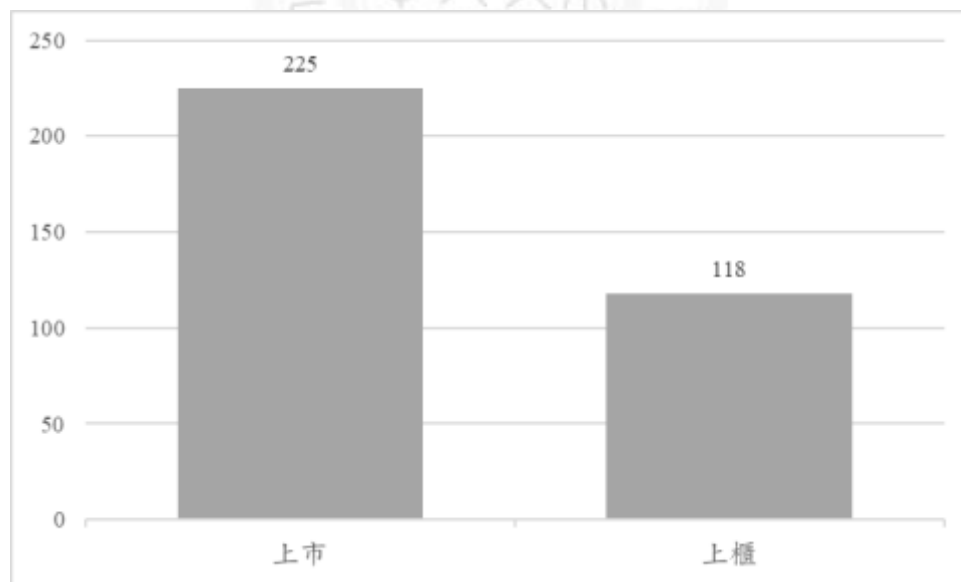


圖 16 機器學習模型選股池上市櫃公司數量圖

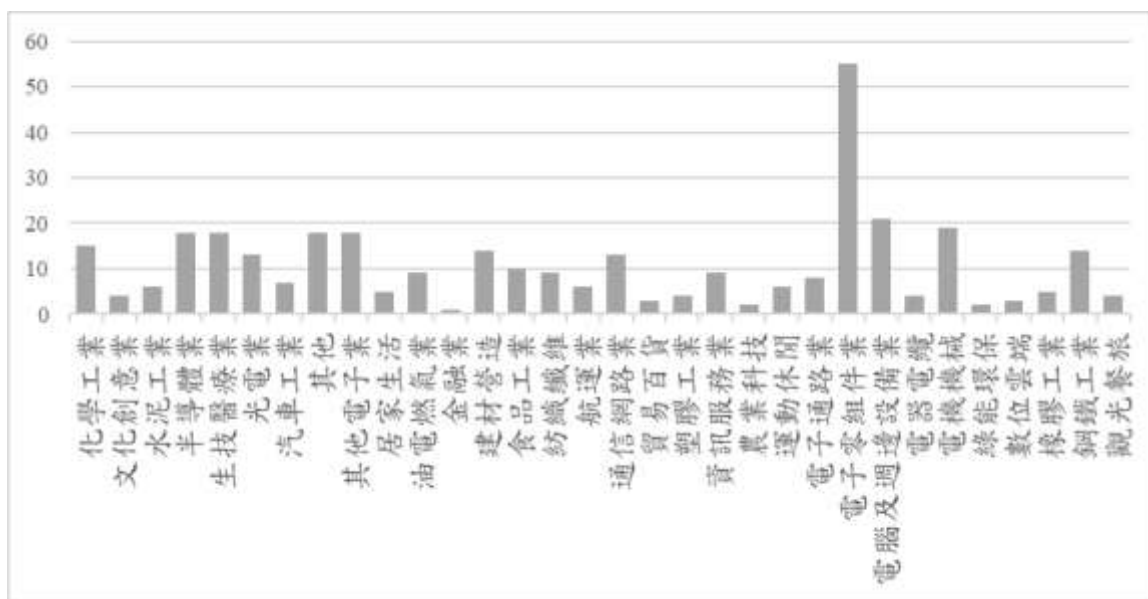


圖 17 機器學習模型選股池各產業公司數量圖

從圖 18 中可看出電子零組件業仍為入選頻率最高之產業，此現象可能來自於該產業公司之財報表現較符合模型選股邏輯，此外，在台灣股市結構中以電子零組件業、半導體業與電腦及周邊設備業佔比較高，且此類產業亦為市場交易較活絡之族群，而本研究特徵變數中納入上季報酬率，亦可能使具有較高動能特性的產業更容易入選。

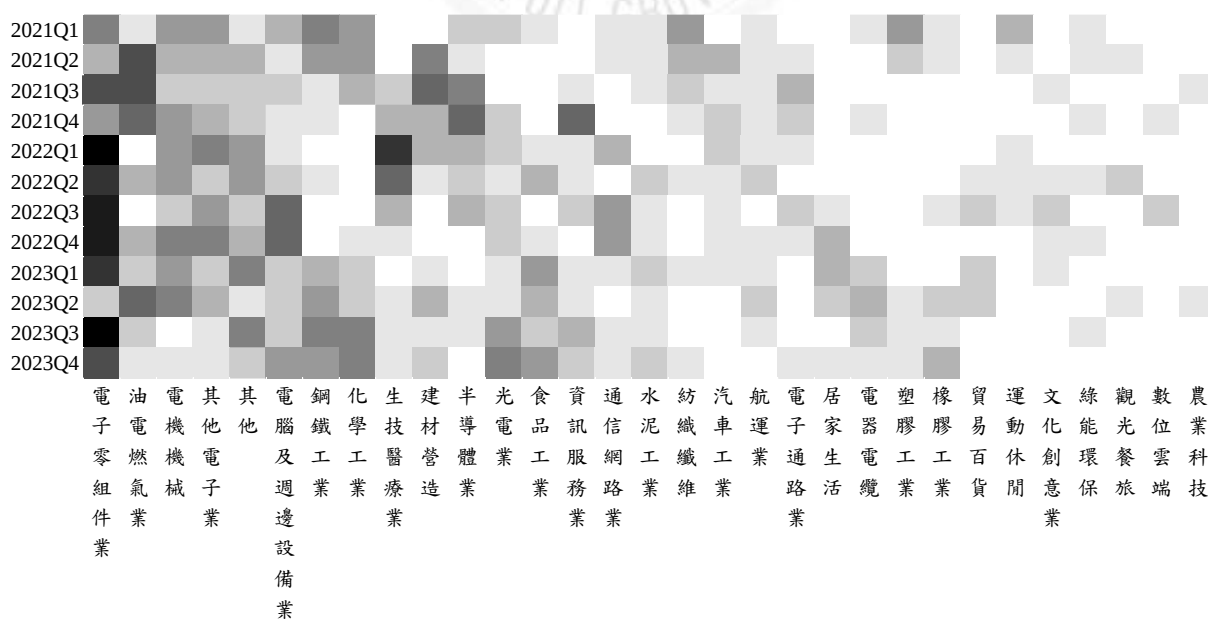


圖 18 隨機森林各季選股分佈熱力圖

本研究在每季財報公布後，自上述 343 家公司分別挑選預測報酬排名前 50、30、20 與 10 名之個股建構投資組合進行回測。所建構之投資組合將與傳統量化投資策略（即調整後 F-score）及指數型 ETF（元大 0050 ETF）比較，以驗證本研究利用機器學習所建構模型之績效，並提供投資人作為實務參考。

## 一、以模型所建構之投資組合一：投資前 50 名

本投資組合首先於 2021 年第一季財報公布（2021/5/31）後，利用模型預測未來一季報酬率之排名，再挑選前 50 名建構投資組合，並於下一季財報公布（2021/8/31）後調整及再平衡投資組合，重複調整投資組合至回測區間結束（2024/5/31），即為 2024 第一季財報公布前，以下稱為投資組合一。

投資組合一之績效如表 20，在回測區間內，投資組合一每季約有 83% 的機率可獲得正報酬率。每季平均報酬率約為 7.04%，年化報酬率為 30.82%，不但優於 0050 之獲利，更勝於調整後 F-score，最終投組淨值成長至 223.88%。而在標準差、年化波動度皆較 0050 差及調整後 F-score 投資策略差，即投資組合一必須承受較多的風險，但在 Max Drawdown 的表現上較 0050 及調整後 F-score 佳，表示投資組合一在股市循環中保有一定的抗跌能力。整體而言，投資組合一的 Sharpe Ratio 約為 3.19，較 0050 及調整後 F-score 來的高，表示平均每承擔一單位的風險投資組合一卻能有較多的回報。

表 20 投資組合一之績效評估表

| 評估指標 \ 投組名稱  | 元大台灣 50 ETF           | 調整後 F-score<br>篩選低 P/B | 模型選股前 50 名 |
|--------------|-----------------------|------------------------|------------|
| 回測期間         | 2021-06-01~2024-05-31 |                        |            |
| 期間(年)        | 3                     |                        |            |
| 每季平均投資家數     | 50                    | 51                     | 50         |
| 勝率           | 58.33%                | 75.00%                 | 83.33%     |
| 每季平均報酬率      | 2.57%                 | 6.39%                  | 7.04%      |
| 年化報酬率        | 9.53%                 | 27.27%                 | 30.82%     |
| 累積報酬         | 131.39%               | 206.17%                | 223.88%    |
| 標準差          | 7.62%                 | 6.26%                  | 4.83%      |
| 年化波動率        | 15.25%                | 12.51%                 | 9.67%      |
| Max Drawdown | -15.59%               | -5.24%                 | -0.91%     |
| Sharpe Ratio | 0.62                  | 2.18                   | 3.19       |

從圖 19 中可瞭解投資組合一、調整後 F-score 與 0050 自 2021 年以來的淨值表現。可觀察到在三種投資組合中，投資組合一之淨值成長幅度最大，而投資組合一與調整後 F-score 在 2022 年市場循環時，皆有優於大盤的表現。至 2024 年 5 月止，累積報酬已大幅拉開與 0050 的差距，顯示本研究模型具備良好的長期選股能力與市場超額報酬潛力。

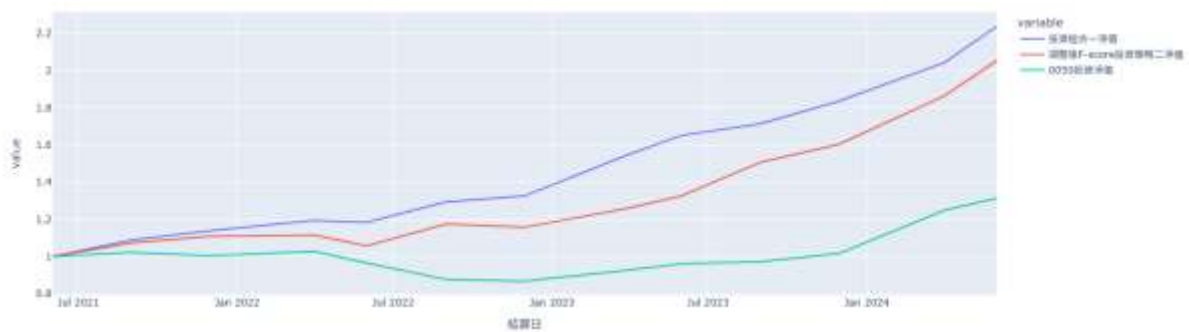


圖 19 投資組合一、調整後 F-score 與 0050 之淨值走勢圖

## 二、以模型所建構之投資組合二：投資前 30 名

本投資組合之投資方式、區間、調整與再平衡時間點皆與投資組合一相同，而投資標的則為每季挑選前模型預測值前 30 名以建構投資組合，以下稱為投資組合二。

投資組合二之績效如表 21，在回測區間內，投資組合二展現出每季獲得正報酬率的機率達 83.33%。其每季平均報酬率約為 8.04%，年化報酬率為 35.45%，最終投組淨值成長至 248.51%。相較於投資組合一，投資組合二除風險表現較弱之外，其餘投資績效指標均優於前者。整體而言，投資組合二的 Sharpe Ratio 約為 2.80，顯示減少投資組合之標的數量確實會使其報酬表現優於投資組合一，同時也因為風險集中使波動較大。

表 21 投資組合二之績效評估表

| 投組名稱<br>評估指標 | 模型選股前 50 名            | 模型選股前 30 名 |
|--------------|-----------------------|------------|
| 回測期間         | 2021-06-01~2024-05-31 |            |
| 期間(年)        | 3                     |            |
| 每季平均投資家數     | 50                    | 30         |
| 勝率           | 83.33%                | 83.33%     |
| 每季平均報酬率      | 7.04%                 | 8.04%      |
| 年化報酬率        | 30.82%                | 35.45%     |
| 累積報酬         | 223.88%               | 248.51%    |
| 標準差          | 4.83%                 | 6.32%      |
| 年化波動率        | 9.67%                 | 12.64%     |
| Max Drawdown | -0.91%                | -2.37%     |
| Sharpe Ratio | 3.19                  | 2.80       |

從圖 20 可瞭解投組淨值\_50（投資組合一）與投組淨值\_30（投資組合二）自 2021 年以來的淨值表現。可觀察到，兩組投資組合在回測期間表現相近。惟投資組

合二支淨值始終大於投資組合一。至 2024 年 5 月止，投資組合二的淨值成長至約 2.48 倍，顯示更集中的選股策略具備更高的報酬潛力。

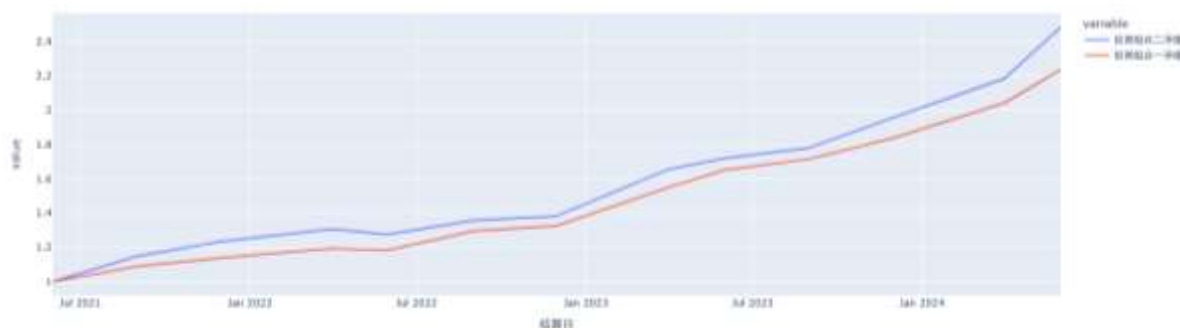


圖 20 投資組合一、二之淨值走勢圖

### 三、以模型所建構之投資組合三：投資前 20 名

本投資組合之投資方式、區間、調整與再平衡時間點皆與投資組合一、二相同，而投資標的則為每季挑選前模型預測值前 20 名以建構投資組合，以下稱為投資組合三。

投資組合三之績效如表 22，在回測區間內，投資組合三的每季勝率同為 83.33%。其每季平均報酬率約為 7.86%，年化報酬率為 34.45%，最終投組淨值成長至 243.03%，在報酬表現上均略輸於投資組合二。相較於投資組合二，投資組合三的風險表現則稍顯劣勢。整體而言，投資組合三的 Sharpe Ratio 約為 2.60，與投資組合二的表現差距不大，顯示兩者在風險調整後的報酬能力相近。

表 22 投資組合三之績效評估表

| <div>投組名稱</div> <div>評估指標</div> | 模型選股前 30 名            | 模型選股前 20 名 |
|---------------------------------|-----------------------|------------|
| 回測期間                            | 2021-06-01~2024-05-31 |            |
| 期間(年)                           | 3                     |            |
| 每季平均投資家數                        | 30                    | 20         |
| 勝率                              | 83.33%                | 83.33%     |
| 每季平均報酬率                         | 8.04%                 | 7.86%      |
| 年化報酬率                           | 35.45%                | 34.45%     |
| 累積報酬                            | 248.51%               | 243.03%    |
| 標準差                             | 6.32%                 | 6.62%      |
| 年化波動率                           | 12.64%                | 13.25%     |
| Max Drawdown                    | -2.37%                | -3.40%     |
| Sharpe Ratio                    | 2.80                  | 2.60       |

從圖 21 可瞭解投組淨值\_30（投資組合二）與投組淨值\_20（投資組合三）自 2021 年以來的淨值表現。透過觀察淨值變化圖可知道投資組合三自 2021 年起，投資績效變優於投資組合二，惟因投資組合三標的較投資組合二集中，故承受更多風險，使最終淨值輸於投資組合二。

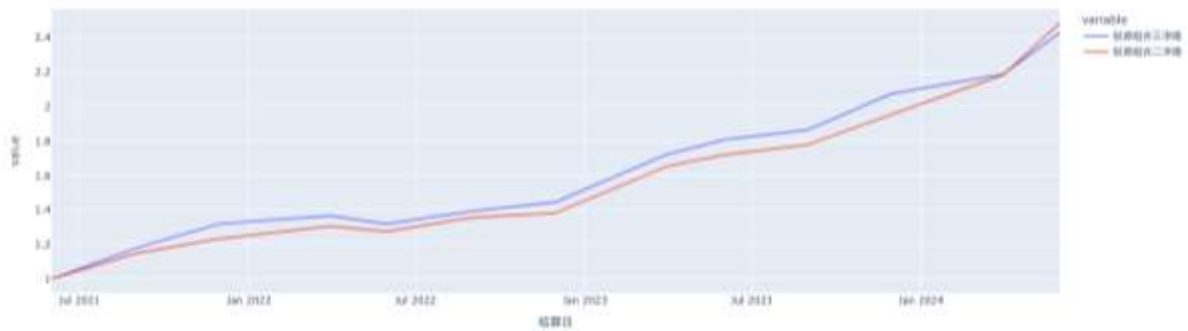


圖 21 投資組合二、三之淨值走勢圖

#### 四、以模型所建構之投資組合四：投資前 10 名

本投資組合之投資方式、區間、調整與再平衡時間點皆與投資組合一、二、三相同，而投資標的則為每季挑選前模型預測值前 10 名以建構投資組合，以下稱為投資組合四。

投資組合四之績效如表 23，在回測區間內，投資組合四的每季勝率下降至 75%。其每季平均報酬率約為 8.91%，年化報酬率為 38.39%，最終投組淨值成長至 265.02%，在報酬表現上均優於投資組合三。相較於投資組合一、二、三，雖然投資組合四擁有最高的報酬表現，但同時投資組合四亦須承擔更大的風險，因此其 Sharpe Ratio 約為 1.88。

表 23 投資組合四之績效評估表

| 評估指標 \ 投組名稱  | 模型選股前 20 名            | 模型選股前 10 名 |
|--------------|-----------------------|------------|
|              |                       |            |
| 回測期間         | 2021-06-01~2024-05-31 |            |
| 期間(年)        | 3                     |            |
| 每季平均投資家數     | 20                    | 10         |
| 勝率           | 83.33%                | 75.00%     |
| 每季平均報酬率      | 7.86%                 | 8.91%      |
| 年化報酬率        | 34.45%                | 38.39%     |
| 累積報酬         | 243.03%               | 265.02%    |
| 標準差          | 6.62%                 | 10.19%     |
| 年化波動率        | 13.25%                | 20.39%     |
| Max Drawdown | -3.40%                | -8.29%     |
| Sharpe Ratio | 2.0                   | 1.88       |

從圖 22 可瞭解投組淨值\_20（投資組合三）與投組淨值\_10（投資組合四）自 2021 年以來的淨值表現。投資組合四自 2021 年起即有較投資組合三優異之報酬率，尤其在 2023 年過後更擴大與投資組合三的差異，使投資組合四最終獲得最高的淨

值，但期間波動度卻是四個投資組合中最高的，透過趨勢圖可知道降低投資標的後雖然可以獲得超額報酬，惟投資人需要承擔更大的風險。

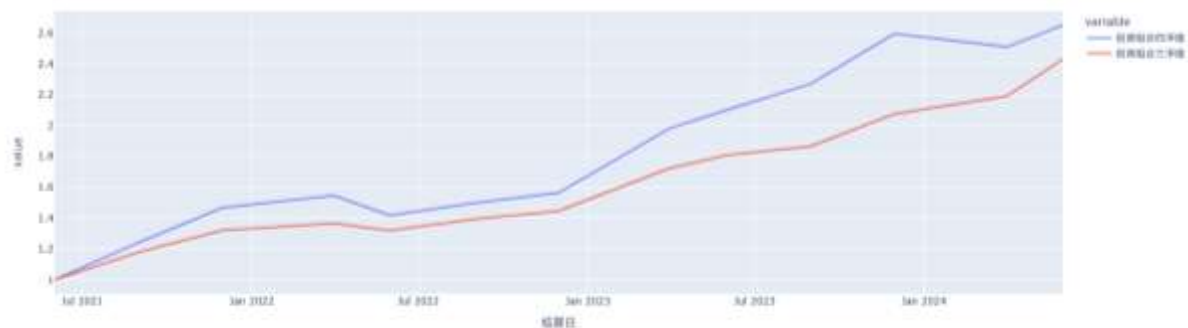


圖 22 投資組合三、四之淨值走勢圖



## 第五章 結論

### 第一節 研究結論

本研究旨在比較基本面指標與機器學習技術應用於台灣股市選股策略中的績效差異，並藉由回測績效驗證兩者之投資報酬率、風險程度與夏普比率等指標的差異。根據研究設計與實證結果，歸納出以下三項主要研究結論：

#### 1. 調整後 F-score 之投資策略確實有優於大盤之績效

本研究以調整後的 F-score 投資策略建構投資組合，實證顯示該策略在長期持有下的報酬率表現明顯優於大盤，亦具備相對穩定的風險控制能力。因此參考 (Piotroski, 2000) 所提出之基本面財務指標來選股並調整為際操作的策略，長期績效優於投資大盤，投資人可依據 F-score 之指標找出基本面良好之標的，建立長期穩健的投資組合。

#### 2. 機器學習模型能提升選股準確性與績效潛力

本研究將 F-score 九項指標轉為連續型變數，並納入 3 年之滾動財報資訊及上一季報酬率，透過隨機森林模型預測未來一季報酬率排名，實證結果顯示機器學習模型較傳統量化投資策略更能有效識別出具潛力之個股。回測結果亦顯示，在多數情境下，機器學習策略的報酬率與夏普比率皆優於調整後 F-score 策略，顯示本研究所建構之模型能提升投資績效，若投資人願意承擔較多的風險以換取更多的報酬率，可依據本研究所建構之模型選股。

#### 3. 選股模型具備操作彈性，有助自然人投資人依風險偏好調整策略

傳統量化策略常需買進所有符合篩選標準之股票，導致投資組合標的數量過多，執行上較為困難。相較之下，本研究所建構之選股模型可依預測報酬率進行排名，

讓投資人依據自身風險承受度與資金配置需求，彈性調整投資標的數量，若投資人風險承受度較高或資金量較低，可以僅挑選前 10、20 名組成投資組合，以換取更高的報酬率；若投資人風險承受度較低，則可以挑選更多檔股票以分散風險，以追求較穩定之投資組合。此一特性不僅提升實務應用性，亦有助於控制交易成本，對自然人投資者而言更具參考價值。

## 第二節 研究建議

本研究透過比較傳統量化技術與機器學習之投資績效，證實了機器學習具備能從基本面指標中辨識出具成長潛力之公司，並利用模型建構具彈性之選股策略，透過實證顯示可有效提升投資組合之績效。然而，考量資料限制與研究設計之侷限性，仍有進一步精進與擴充之空間，提出以下四項建議，作為後續研究參考：

### 1. 將金融類股納入分析，以擴大模型適用範圍

本研究因金融業財務報表結構與一般產業不同，故在樣本挑選時將其排除。然而，金融類股在台灣股市中佔有相當權重，亦為多數投資人重要配置標的。建議未來研究可針對金融產業特性進行指標調整，將其納入模型中，以擴大選股策略之適用範圍，並提供投資人更多元之選擇。

### 2. 擴充特徵變數來源，提升模型預測力

本研究主要以 F-score 九項指標為核心，輔以財報歷史資料與前一期報酬率作為特徵變數。惟股價變動受到多種因素影響，建議後續研究可嘗試引入其他基本面、技術面及籌碼面指標，以強化模型對市場短期與中期反應的捕捉能力。亦可蒐集 2013 年以前的資料，使模型有更多的資料量可以學習，或許可以提升模型的預測能力。

### 3. 提升策略反應頻率，納入月營收資訊以即時調整

本研究策略執行頻率為每季一次，主要配合財報公告時程。然而，台灣上市櫃公司每月皆須公告營收數據，營收變化常為市場重要參考依據，若能即時反映於投資決策中，或可進一步提升選股策略之即時性與靈敏度。建議未來研究可將月營收變化納入特徵變數，並嘗試提升至每月調整一次之策略執行頻率，以提升模型在實務操作上的時效性與反應能力。

### 4. 納入交易成本與滑價因素，使回測結果更貼近實務

本研究回測尚未納入交易成本與滑價之考量，然在實際操作中，投資人往往面臨手續費、證交稅與因個股流動性不佳所造成之滑價問題，尤其在標的數量較多時更為明顯。為使回測結果更貼近真實市場情境，建議後續研究可於模擬交易中加入合理的成本假設，並針對不同流動性之股票設計滑價參數，以建立更具實務參考價值之選股與交易模型。

## 參考文獻

- Chan, L. K., Jegadeesh, N., & Lakonishok, J. (1996). Momentum strategies. *The Journal of Finance*, 51(5), 1681-1713.
- Chen, A. S., & Hong, B. S. (2006). Institutional ownership changes and returns around analysts' earnings forecast release events: Evidence from Taiwan. *Journal of Banking & Finance*, 9(30), 2471-2488.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 2(25), 383-417.
- Graham, B., & McGowan, B. (2003). *The intelligent investor*. New York: HarperBusiness Essentials.
- Gu, S., Kelly, B., & Xiu, D. (2020, May). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273.
- Huang, Y., Capretz, L. F., & Ho, D. (2021). Machine learning for stock prediction based on fundamental analysis. *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1-10. Orlando, FL, USA.
- Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance*, 5(49), 1541-1578.
- Mintarya, L. N., Halim, J. N., Angie, C., Achmad, S., & Kurniawan, A. (2023). Machine learning approaches in stock market prediction: A systematic literature review. *Procedia Computer Science*(216), 96-102.
- Mohanram, P. S. (2005). Separating winners from losers among lowbook-to-market stocks

using financial statement analysis. *Review of Accounting Studies*(10), 133-170.

Namdari, A., & Li, Z. S. (2018). Integrating fundamental and technical analysis of stock market through multi-layer perceptron. *2018 IEEE Technology and Engineering Management Conference (TEMSCON)*, 1-6. Evanston, IL, USA.

Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*(38), 1-41.

Rasekhschaffe, K. C., & Jones, R. C. (2019). Machine learning for stock selection. *Financial Analysts Journal*, 3(75), 70-88.

王希佩 (2021)。以機器學習建構股價預測模型：以台灣股市為例。(碩士論文，國立中央大學)。臺灣博碩士論文知識加值系統。

王銘駿、陳宜伶、吳昭億、黃巧雯 (2015)。台灣政治循環下股票市場投資組合績效之探討。 *商管科技季刊*，16 (2)，145-179。

田鉉元 (2022)。以隨機森林法建構投資組合績效—以台灣股票市場為例。(碩士論文，國立清華大學)。臺灣博碩士論文知識加值系統。

李雅璇 (2013)。以巴菲特選股模式建構基本面投資組合：在台股之實證研究。(碩士論文，國立中央大學)。臺灣博碩士論文知識加值系統。

卓越 (2019)。基於支持向量回歸的選股模型實證研究—以台股市場為例。(碩士論文，國立政治大學)。臺灣博碩士論文知識加值系統。

林佳靜 (2019)。以改良 F-score 指標建構投資組合績效之探討。 *管理資訊計算* (8)，1-23。

夏明義 (2022)。基於 F-score 以及修正式 F-score 的交易策略。(碩士論文，國立政治大

學)。臺灣博碩士論文知識加值系統。

孫永鴻 (2020)。機器學習演算法在股票交易策略之應用。(碩士論文，國立臺北大學)。

臺灣博碩士論文知識加值系統。

張祺峻 (2023)。台股基本面量化選股策略優化分析。(碩士論文，國立中山大學)。臺灣

博碩士論文知識加值系統。

張嘉芸 (2024)。以 F-Score 為投資策略之實證研究：以台灣上市公司為例。(碩士論文，

國立成功大學)。臺灣博碩士論文知識加值系統。

許珮琪 (2019)。基本面之選股模型的策略與優化-以台灣股市為例。(碩士論文，國立中

山大學)。臺灣博碩士論文知識加值系統。

郭峻宇 (2023)。利用機器學習提升 F-score 的投資績效。(碩士論文，國立成功大學)。

臺灣博碩士論文知識加值系統。

陳同力 (2019)。運用機器學習預測台股股價走勢。(碩士論文，國立臺灣大學)。臺灣博

碩士論文知識加值系統。

陳鼎玟 (2018)。應用機器學習與深度學習預測台股指數。(碩士論文，朝陽科技大學)。

臺灣博碩士論文知識加值系統。

葉千綺 (2015)。量化選股策略--以台灣市場為例。(碩士論文，國立中山大學)。臺灣博

碩士論文知識加值系統。

謝曜宏、許資法、朱涵宇、林祝興 (2019)。使用機器學習的智慧選股策略之研究。載於

NCS 2019 全國計算機會議 (467-472)。國立金門大學。

韓家韋 (2024)。應用基本面、籌碼面及技術面指標建構投資組合實證分析。(碩士論文，

國立中山大學)。臺灣博碩士論文知識加值系統。