



國立中山大學財務管理學系

碩士論文

Department of Finance

National Sun Yat-sen University

Master's Thesis

運用機器學習於台股價格突破策略之實證分析

Empirical Analysis of the Application of Machine Learning in

Price Breakout Strategies for Taiwan Stock Market

研究生：吳泰緯

Tai-Wei Wu

指導教授：王昭文 博士

Dr. Chou-Wen Wang

中華民國 112 年 7 月

July 2023

論文審定書

國立中山大學研究生學位論文審定書

本校財務管理學系碩士在職專班

研究生吳泰緯（學號：N104030008）所提論文

運用機器學習於台股價格突破策略之實證分析

Empirical Analysis of the Application of Machine Learning in Price Breakout
Strategies for Taiwan Stock Market

於中華民國 112 年 6 月 27 日經本委員會審查並舉行口試，符合碩士學位論文標準。

學位考試委員簽章：

召集人 吳明政 吳明政 委員 王昭文 王昭文

委員 吳錦文 吳錦文 委員 王銘駿 王銘駿

委員 洪志興 洪志興 委員 _____

指導教授(王昭文) 王昭文 (簽名)

致謝

兩年的時間真的很快，一晃眼就過去了，還記得當初還經歷了疫情的遠端上課，轉眼間就已經畢業了，感謝在這修課兩年的期間，所認識的朋友同學們，在每一堂的分組課程都能給予最大的支持及合作，讓我又拓展了一些人脈，及每一堂課的老師增加我的學理知識，也感謝昭文老師，讓我有機會可以更深入的去學習運用程式來輔助投資交易，以及給予論文上的方向。也感謝當初在運錫時，幫忙推薦的顏董和顏太，讓我有機會能夠進來中山大學學習。

也感謝爸爸、媽媽從以前就不斷地支持我所想要做的事情，也感謝我的女朋友一路伴我至今，從最早的面試到畢業典禮都沒缺席，以及各位一路上幫助我的朋友和貴人們，有你們大家才有現在的我。

雖然畢業了，但人生的道路上還有很多值得學習的事物，和未知的旅程，將繼續向前行，並也期許能在往後的道路上，能有了這兩年的經驗基礎，讓我向上一層樓，也期望我未來能繼續保有我的初衷，眾善奉行，能夠有能力的去回饋這社會。

摘要

本研究使用了 2007 年 1 月至 2022 年 12 月之台股上市櫃股價資料，應用價格突破法則，並考慮突破日前 20 日的高低振幅 20%作為策略的買入依據。我們針對短期（5 日、10 日）、中期（20 日、60 日）、長期（120 日）的持有報酬率實施回測。在引入機器學習技術後，我們利用極限梯度提升模型（XGBOOST）、輕量梯度提升模型（LGBM）、多元羅吉斯迴歸模型來進一步優化該投資策略。原始策略的實施結果表明，儘管能夠捕捉到大漲股的市場趨勢，卻也購入了許多大幅下跌的股票。僅以價格突破為基礎的投資策略顯然有其局限性，且整體策略的平均報酬率都無法超越同期大盤的表現。然而，我們發現中長期的投資勝率和報酬率明顯優於短期，這也證明了股票的長期持有策略較具優勢。在應用機器學習技術後，三種模型的報酬率和勝率均有所提升，特別在短期投資上，多元羅吉斯迴歸模型的勝率和報酬率提升最為明顯，然而隨著時間的拉長，這些指標卻持續下滑。而在中長期投資上，極限梯度提升模型(XGBOOST)表現最為出色。總之，本研究主要探討了價格突破後的買入策略是否具有較高的勝率，而除了價格因素外，我們認為仍需考慮到股價的位階、營收、籌碼等其他因素的影響。

關鍵字:價格突破、極限梯度提升模型、輕量梯度提升模型、多元羅吉斯模型、台灣股市

Abstract

This study utilizes the Taiwan Stock Exchange data from January 2007 to December 2022, implementing the price breakout rule, and considering a 20% fluctuation in the 20 days prior to the breakout day as the basis for our buy-in strategy. We performed backtesting for short-term (5 days, 10 days), mid-term (20 days, 60 days), and long-term (120 days) holding return rates. After incorporating machine learning techniques, we employed the Extreme Gradient Boosting Model (XGBOOST), Light Gradient Boosting Model (LGBM), and Multivariate Logistic Regression Model to further optimize this investment strategy. Results from the initial strategy implementation indicate that despite capturing trends of sharply rising stocks, a considerable number of sharply falling stocks were also purchased. This shows the limitations of investment strategies based solely on price breakout, as the overall strategy's average return rate could not surpass the market's performance over the same period. However, we discovered that the win rate and return rate for mid and long-term investments were significantly superior to short-term investments, demonstrating the advantages of long-term holding strategies for stocks. Upon the application of machine learning techniques, all three models showed an increase in return and win rates,

particularly the Multivariate Logistic Regression Model, which showed the most substantial improvements for short-term investments. However, these metrics gradually declined over time. In contrast, the Extreme Gradient Boosting Model demonstrated the best performance for mid and long-term investments. In conclusion, this research primarily explores whether buy-in strategies following a price breakout have a higher win rate. Beyond the price factor, we believe it is also crucial to consider other factors such as the rank of stock price, revenue, and chips in their impacts.

Keywords: Price Breakthrough 、 XGBOOST 、 LGBM 、 Multinomial Logistic Regression Model 、 Taiwan stock

目錄

論文審定書	i
致謝	ii
摘要	iii
Abstract	iv
第一章 緒論	1
1.1 研究背景與動機.....	1
1.2 研究目的.....	2
1.3 研究架構.....	4
1.4 研究流程圖.....	5
第二章 文獻探討	6
2.1 價格趨勢面之相關研究.....	6
2.2 籌碼面之相關研究.....	7
第三章 研究方法	10
3.1 資料來源.....	10
3.2 策略設計.....	10
3.3 機器學習模型.....	12
3.4 極限梯度提升(XGBOOST, eXtreme Gradient Boosting).....	12
3.5 輕量梯度提升(LightGBM, Light Gradient Boosting Machine).....	13
3.6 多元羅吉斯迴歸.....	14
3.7 機器學習特徵選用.....	15
3.8 機器學習時間設計.....	17
第四章 實證結果與分析	19
4.1 總樣本敘述性統計.....	19

4.2	大盤季線上揚時的策略表現.....	21
4.3	突破時成交量明顯放大.....	22
4.4	均線糾結時.....	23
4.5	設定停損條件出場.....	24
4.6	策略加入濾網 持有天數比較.....	25
4.7	加入機器學習.....	28
4.7.1	原始策略使用機器學習	28
4.7.2	大盤季線上揚加入機器學習	37
4.7.3	成交量放大	44
4.7.4	均線糾結	56
4.8	小結.....	63
第五章 結論		64
5.1	研究結論.....	64
5.2	研究限制與建議.....	65
文獻參考		66
1.	中文文獻	66
2.	英文文獻	67

圖目錄

圖 3-1 箱型區間示意圖	11
圖 4-1 閾值 0.5 報酬率對照	34
圖 4-2 閾值 0.5 勝率對照	34
圖 4-3 閾值 0.55 報酬率對照	35
圖 4-4 閾值 0.55 勝率對照	35
圖 4-5 閾值 0.6 報酬率對照	36
圖 4-6 閾值 0.6 勝率對照	36
圖 4-7 閾值 0.5 季線上揚報酬率對照	41
圖 4-8 閾值 0.5 季線上揚勝率對照	41
圖 4-9 閾值 0.55 季線上揚報酬率對照	42
圖 4-10 閾值 0.55 季線上揚勝率對照	42
圖 4-11 閾值 0.6 季線上揚報酬率對照	43
圖 4-12 閾值 0.6 季線上揚勝率對照	43
圖 4-13 閾值 0.5 兩倍成交量報酬率對照	50
圖 4-14 閾值 0.5 兩倍成交量勝率對照	50
圖 4-15 閾值 0.5 三倍成交量報酬率對照	51
圖 4-16 閾值 0.5 三倍成交量勝率對照	51
圖 4-17 閾值 0.55 兩倍成交量報酬率對照	52
圖 4-18 閾值 0.55 兩倍成交量勝率對照	52
圖 4-19 閾值 0.55 三倍成交量報酬率對照	53
圖 4-20 閾值 0.55 三倍成交量勝率對照	53
圖 4-21 閾值 0.6 二倍成交量報酬率對照	54
圖 4-22 閾值 0.6 二倍成交量勝率對照	54
圖 4-23 閾值 0.6 三倍成交量報酬率對照	55
圖 4-24 閾值 0.6 三倍成交量勝率對照	55
圖 4-25 閾值 0.5 均線糾結報酬率對照	60
圖 4-26 閾值 0.5 均線糾結勝率對照	60
圖 4-27 閾值 0.55 均線糾結報酬率對照	61
圖 4-28 閾值 0.55 均線糾結勝率對照	61
圖 4-29 閾值 0.6 均線糾結報酬率對照	62
圖 4-30 閾值 0.6 均線糾結勝率對照	62

表目錄

表 3-1 特徵選用彙整	16
表 3-2 時間窗格法表格	18
表 4-1 原始樣本敘述統計表	20
表 4-2 季線上揚敘述統計表	21
表 4-3 成交量至少兩倍敘述統計表	22
表 4-4 成交量至少三倍敘述統計表	23
表 4-5 均線糾結時敘述統計表	24
表 4-6 設定停損條件敘述統計表	25
表 4-7 持有 5 天敘述統計表	26
表 4-8 持有 10 天敘述統計表	26
表 4-9 持有 20 天敘述統計表	27
表 4-10 持有 60 天敘述統計表	27
表 4-11 持有 120 天敘述統計表	28
表 4-12 原始策略時間窗格法資料筆數	29
表 4-13 原始策略閾值 0.5	32
表 4-14 原始策略閾值 0.55	32
表 4-15 原始策略閾值 0.6	33
表 4-16 季線上揚時間窗格法資料筆數	38
表 4-17 季線上揚閾值 0.5	39
表 4-18 季線上揚閾值 0.55	39
表 4-19 季線上揚閾值 0.6	40
表 4-20 兩倍成交量時間窗格法資料	45
表 4-21 三倍成交量時間窗格法資料	46
表 4-22 成交量兩倍閾值 0.5	47
表 4-23 成交量兩倍閾值 0.55	47
表 4-24 成交量兩倍閾值 0.6	48
表 4-25 成交量三倍閾值 0.5	48
表 4-26 成交量三倍閾值 0.55	49
表 4-27 成交量三倍閾值 0.6	49
表 4-28 均線糾結時間窗格法資料	57
表 4-29 均線糾結閾值 0.5	58
表 4-30 均線糾結閾值 0.55	58

表 4-31 均線糾結閾值 0.6.....	59
------------------------	----



第一章緒論

1.1 研究背景與動機

2020 年 COVID-19 疫情對全球經濟及股市的衝擊無疑是劇烈且深遠。然而，這場全球的風暴並非全然負面，事實上，它創造了一個前所未有的經濟環境，產生了許多難得的投資機會。Fed 的寬鬆政策以及其它各國央行的量化寬鬆政策，在市場上引導了龐大的資金流向，這也引發了一波波的股市上揚。

此外，疫情導致的全球封鎖重創了各產業的供應鏈，大幅增加了需求訂單和運價，甚至出現了超額預訂的情況。這些變化為股市帶來劇烈波動，也使得許多航海王、少年股神橫空出世。這也讓價格突破交易受到矚目，然而，這種交易方式是否真的能夠長期穩定地賺錢，尚未有明確的答案。因此，本研究的動機在於，我們希望能夠深入探討並評估，是否可以僅依靠技術線型的突破追漲來獲取利益。希望透過對過去市場資料的深入分析，能夠找出有系統的方法，來評估並應對這種交易方式。

我們也注意到台股每年都有特定的題材股，如 2018 年的被動元件:國巨，2020 年初防疫類股，口罩:恆大;耳溫槍:泰博，2021 年航運類股，因此想知道若不熟基本面，財務等狀況，且沒有領先的市場資訊時，是否可以只透過技術面的價格突破去做買入持有，來獲取最佳的報酬率。

人工智能（AI）與機器學習的快速發展，已使其成為數據驅動決策的關鍵工具。AI 不僅在日常生活的許多領域中被廣泛應用，更在金融領域顯示出巨大的潛力。它們被越來越多的研究證明可以有效預測未來市場走勢，甚至提供投資策略決策。在 2022 年底，以 ChatGPT 為代表的 AI 模型的出現，不僅提升了

AI 的知名度，更推動了 AI 技術的前沿發展，將其應用性和準確性推向了新的高度。

因此，本研究的目標之一是探討如何運用機器學習的方法來提高我們的投資收益。我們計劃使用不同的機器學習模型來篩選個股，並試圖找出 AI 是否能夠有效地尋找到最佳的投資目標。這涉及到對各種模型的評估，包括其預測能力、適應市場變化的能力、以及處理大規模數據的能力等。

進行這種研究的同時，我們還將對過去的市場數據進行深入分析，以期能找出一套有系統的方法來評估和應用機器學習模型。我們相信，透過這種深度研究的方法，我們可以更深入地理解如何在一個複雜且變化多端的市場環境中，使用技術分析和機器學習技術來做出最好的投資決策。

這項研究充滿挑戰，並將在金融投資領域中尋求突破。我們期待通過這樣的研究，不僅能對 AI 在股市投資中的應用有更深入的理解，也能對金融投資領域產生重大的貢獻。從而將機器學習與投資策略相結合，達到最大化投資回報的目標，並將 AI 在此領域的潛力推向新的高度。

1.2 研究目的

本研究的目的是透過探討股價的技術面，深入理解和評估市場的效率。尤金·法馬 (Fama, 1970) 提出了效率市場假說，這假說的核心觀點：如果證券市場中，價格完全反映了所有可以獲得的訊息，那麼就稱這樣的市場為有效市場。然而，在實際的市場中，價格是否真的能夠反映所有的信息，這問題仍然值得探究。特別是在遭遇多頭和空頭市場的循環，以及股價的震盪盤整之後，價格的向上或向下突破是否能夠引發新的市場波動，這一問題尤其值得我們深入探討。考慮到市場經濟的通膨以及科技的迅猛發展，股市在長期的時間上表

現出明顯的上升趨勢，這也使得價格的向上突破成為一種具有吸引力的投資策略。因此，本研究將重點探討在股價經過一段時間的盤整之後，向上突破的作多策略在短線、中線、長線持有報酬的表現。除了價格突破，我們將加入一些篩選機制，以篩選出具有較高勝率的個股。我們期望透過這種方式，能夠更精準地捕捉市場的動能，並將其轉化為穩定的投資回報。

(Barber, 2001)指出大多數的投資人有過度自信的傾向，導致頻繁交易，而只有獲得低報酬，因此本研究會比較著重於探討持有長期的報酬率與勝率。因此本研究想要了解

1. 價格突破盤整區間後持有的短、中、長期報酬。
2. 納入成交量及大盤的多空頭背景及均線條件，是否更有效篩選出較高報酬率之個股。
3. 運用機器學習三個模型 XGBOOST、Light GBM、多元羅吉斯迴歸來預測是否可以得到更好的選股報酬。

1.3 研究架構

本研究架構分為五大章節，如下

第一章 緒論

說明本研究動機之目的，並論述架構

第二章 文獻探討

文獻探討相關投資理論的應用，以及技術分析指標與價量關

係之相關論文

第三章 研究方法

策略之設計與濾網之選用，以及機器學習的模型相關介紹

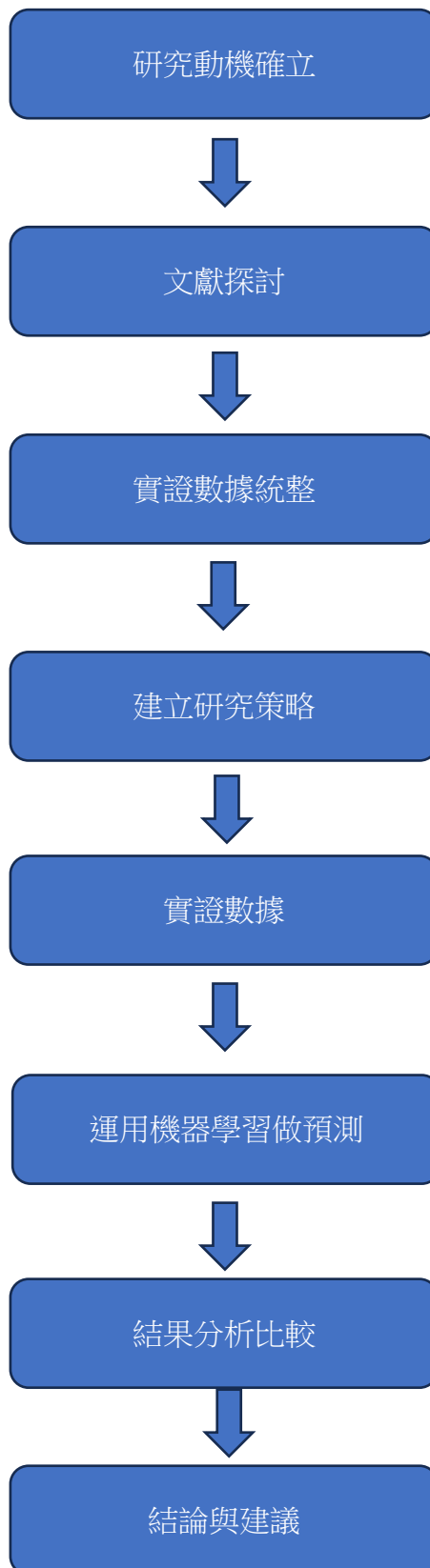
第四章 實證結果與分析

將本研究之實證結果提出說明與機器學習所得結果做比較

第五章 結論

總結本研究之結果

1.4 研究流程圖



第二章文獻探討

2.1 價格趨勢面之相關研究

樓禎祺等(2003)，以台灣來說短期最常被用的 MA 為 5 日、10 日，短中期為 20 日、中期 60 日，中長期 120 日，並探討作為個股股價趨勢與投資的方向，他們研究發現以半年的周期，投資個股來說 20 日(20MA)、30 日(30MA)的效果報酬是最好的，再來是 60 日(60MA)，最差的則是 5 日(5MA)、10 日(10MA)。再來則是運用在指數上的效果並不能獲得比較好的績效。

許溪南等(2005)，發現市場的每日交易量影響著股市的報酬，量能確實有助漲助跌的效果，以及融資的量也都會影響股市的價格波動，融資的金額多半會認為是散戶的籌碼，說明這些籌碼的變動主要也和情緒有關，以台股來說散戶相較於法人就是容易受到情緒的影響，所以使得短線的籌碼變動較大，可以從融資的變動量觀察出來，

陳淑玲等(2011)，研究發現移動平均線以 60 日為出發點的投報效果最佳，不論在多空頭市場，以及與 9 日 K D 的技術指標組合的效果最好，且當大盤處於多空轉折處時，60 日的移動平均線穩定性及參考性較高，較能反映出趨勢的變化，半年線（120 日）及年線（240 日）的反應會較慢。

林睦融(2012)研究中發現，K 線圖所組成的個股趨勢，可以當作操作的依據，並加入實體 K 棒強度，以及成交量，還有個股的位階，可以更有效地來運用進出場，當個股在低位階，出現爆量長紅的突破訊號所帶來的績效可以擊敗

大盤，相對的在高檔出現賣出訊號，例如爆大量或是長黑棒，都是可視為賣出訊號，所放空的報酬還是優於大盤表現。

Mark(2015)，多次贏得美國投資冠軍，在他的書籍超級績效 1~3，採訪了多位美國的期貨、股市贏家，以及他自己的個人致富心得，並提出了 S E P A（特定進場點分析）：分別是 1.趨勢 2.基本面 3.催化事件 4.進場點 5.出場點，可以發現這些贏家的共通點都是，在價格下跌時不買進或是攤平，這是非常危險的動作，反而在價格即將突破高點時，或是突破高點時買入，並也提到價漲量增、價跌量縮的多頭上漲型態的成交量會出現的現象。

王震頤(2020)，用 K 線組合與均線做策略基礎，發現突破買進的策略為交易次數較多的策略，但隨著位階高低，影響突破策略的勝率，因為突破又分為真突破與假突破，因此勝率會較其所測試的拉回買進還差，股價在低檔時的突破勝率及報酬率高於高檔突破進場。

2.2 籌碼面之相關研究

蘇鵲翎(2000)，週轉率的高低可以當作股票流動性高低以及當前投機的氣氛，並發現散戶有追高殺低做短線的傾向，外資則較不會，研究也發現外資的持有較為中長期，且有穩定籌碼的效益。

王月玲(2003)研究發現外資在台股市場扮演領先者的角色，因為其研究機構的規模及收集資訊具有指標性，因此也會讓國內的法人（自營商、投信）有跟隨的跡象，外資的買賣超對股票的報酬率有正向效果，加入市場多空頭的背景時，當市場空頭且蕭條時，外資的買賣超對於個股的影響更大，最後外資的加入，讓台股的整體波動率流動性大增不少。

楊淑惠(2007)外資的連續買超行為，對個股具有一定的影響力，長天期的買超對於個股有一定的上漲助力，在多頭時期這種現象會特別明顯，但是連續賣超對股價也有一定的壓抑作用，在空頭時，外資連續賣超的情形會比較多。也可以發現外資主要都以順勢操作為主，因此跟著外資籌碼操作，中長線上會更有優勢。

徐中琦等(2008)依據 Lakonishok et al.(1992)所提出衡量投信法人從眾行為的概念，後發現不論是國內自營商或是外資，在買賣時都有從眾效應，並發現自營商不論在金融、電子、傳產從眾的程度又比外資明顯，特別是傳產類股，因為地緣的關係，自營商的著墨也較深，相較於外資機構法人較著重於在電子和金融的研究，因此當自營商大買傳產股時，會吸引更多的投資人及法人跟進，在股市多空的背景下，空頭時的從眾效果更明顯，外資在賣出時，會有連續賣出的情況，對股價更壓抑，也可以看出外資對風險的趨避，但也可以了解外資、自營收等法人的籌碼對於股價的助漲助跌。

于慎為(2011)傳統價值投資中，一起加入法人的籌碼當作考量，研究價值型投資找到被低估的個股買入持有，報酬率長線來看都優於大盤，然後再將法人的買入籌碼加入後的績效，優於同期大盤的表現，也比原本的投資組合，未納入法人籌碼時報酬率來的還要好。林軒白(2016)研究發現投信與外資的買超行為，一般來說外資具有持續性，甚至季買超的行為也是有連續的現象，但是投信沒有，也發現如果小型股有外資和投信的買超痕跡，那麼下一季賣掉的機會很大，在小型股的操作上都是偏向短線，在外資的買超行為中，可以發現外資更喜歡中大型股，投信則是比較喜歡小型。外資和投信的買超前十名中，有

顯著的正超額報酬，研究也發現投信的投組中，第四季的績效都較第一季來得差，這也可能是投信為了績效作帳的關係，都偏向於年底作帳。

劉百耀(2020)發現法人的籌碼中，投信的籌碼持股、買賣數量較高的，夏普比率和累計的報酬率都是優於大盤的表現，借券賣出的張數下降或是較低的個股，可以視為法人放空較少，可當作消極看多的籌碼，累積的報酬率還是優於大盤，反而在融資使用率和股票報酬呈現反向關係，一般也將融資視為散戶的籌碼。

總結來說，本研究的貢獻可以歸納如下：

1. 利用 TEJ 資料庫的數據，提供一個實證的基礎，用以證實中長線的投資，是可以取得更好的報酬率和勝率。一般投資大眾都應該把投資週期拉長。
2. 運用均線策略搭配法人籌碼確實可以提高報酬率和勝率。
3. 運用三個機器學習模型搭配價格突破的策略以及不同的成交量、均線等背景條件下，得出短中長期的模型歸納使用。
4. 研究結果證實，多元羅吉斯迴歸模型，在短中線上的篩選優於其他兩者模型，若是長線的投資，XGBOOST 的模型更勝一籌，顯示短中長線不同投資週期時的模型選用，為未來研究提供參考方向。

第三章研究方法

3.1 資料來源

本研究所使用的資料，來自台灣經濟新報(TEJ)之還原股價資料庫，為了產生所需均線參數，日期從 2006/07/01，開始至 2022/12/31，選取台灣所有的上市櫃公司不含(TDR、ETF)總計 1779 家之每日的開盤價、收盤價、最高價、最低價、成交量，還有對應的加權股價之資料。但實際的研究期間為，2007 年 1 月 1 日至 2022 年 12 月 30 日，扣除缺失值後，總資料筆數為 5,458,616 筆。

3.2 策略設計

本策略初始設計為，收盤價突破 60 日之高點，並設置箱型區間的概念，在突破日的前 20 天之高低點震盪為 20%以內，以及當日需漲幅大於 5%，且收實體紅棒。如圖 3-1 所示。藉此設計來了解，當股價長紅棒突破箱型區間時，後續帶來的報酬利潤，並在接著往下分別加入濾網，分別去檢視在不同環境下此策略所持有短中長期帶來的報酬，並在加入設置停損點出場去檢視績效。

以下分別三點策略濾網：

1. 當大盤在季線上，且季線上揚

大盤季線，60 日移動平均線(60MA)，又稱做生命線，為一季的成本線，也會視為多頭的情況，當大盤在季線上時，會有較多的投資人，進場買單，因此選擇大盤在季線上且季線上揚，研究策略的表現。根據葛蘭碧八大法則，移動平均線在下降後逐漸走平或上揚時，可作為股票買賣的指標。

2. 當突破的成交量是 20 日均量的 2 倍、3 倍以上

股市中常存在著量價關係，量先價行，田鑫一(2008)探討過股市量價關係，提到了“天量見天價，地量見地價”價漲量增、價跌量縮，（游英裕，2004）研究結果為價量俱有雙向因果關係，選股時成交量可做為參考依據，若是行情發動突破時，也代表有資金買盤進場。

3. 當均線糾結時的突破

當一段盤整時間長短周期均線糾結後，此時的成本都是很接近的，股價也暫時沒有方向，因此突破的爆發力往往比一般來得強，也通常是波段漲跌幅的起點，因此本節加入均線糾結的條件，將均線彼此的差距都限縮在 5% 以內。

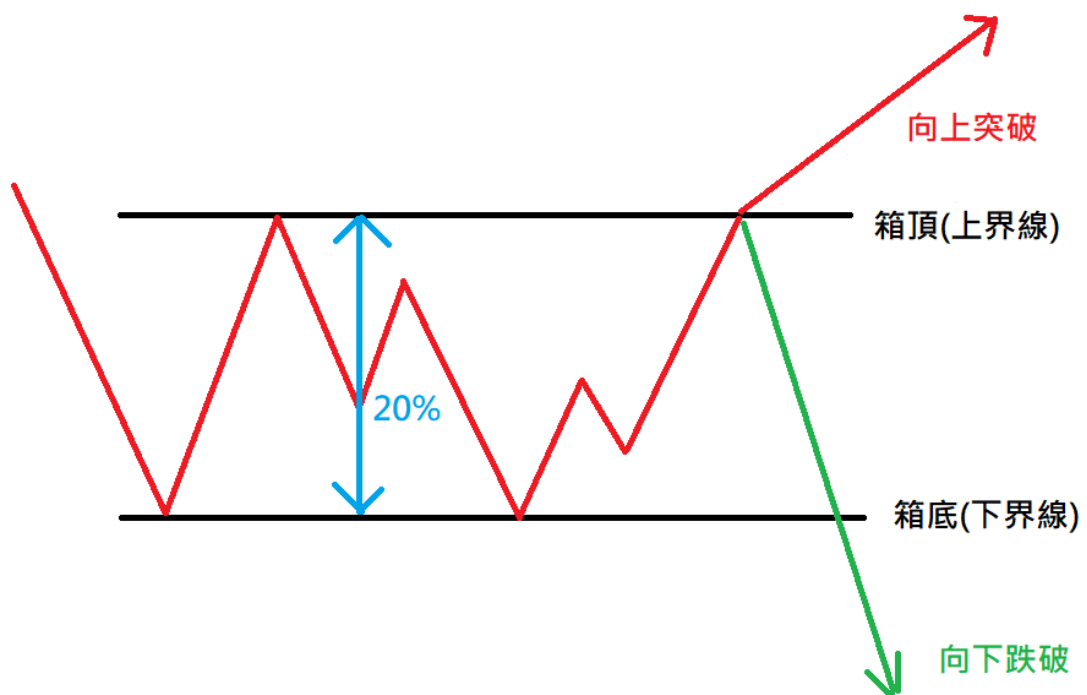


圖 三-1 箱型區間示意圖

資料來源:本研究整理

3.3 機器學習模型

機器學習(Machine Learning, 簡稱 ML)，是人工智慧的一個分支，在這一詞的出現道發展至今已過 40、50 年，隨機漫步理論中，認為股價的波動是隨機沒有規律，且難以預測，因此想藉由機器學習並給予其股價的價格特徵、技術指標……等，是否可以預測出較高勝率和報酬率的個股。機器學習又分為監督式學習和非監督式學習，所謂監督式學習：是指使用一組已知答案的例子來訓練模型，然後用訓練好的模型去預測新的數據。非監督式學習：只有輸入數據 (X)，但並沒有對應的目標變量(Y)。目標是通過未標籤的數據來找出某種結構或關聯。這種學習方法的主要挑戰是我們需要自己去理解和解釋結果。我們想了解當下股票突破後，持有短中長期的報酬，用以現有的數值去做預測，因此採用了以下三種監督式學習的模型：

3.4 極限梯度提升(XGBOOST, eXtreme Gradient Boosting)

極限梯度提升(XGBOOST, eXtreme Gradient Boosting)，是目前較常見到的算法模型，由陳天奇所提出，一種集成學習系統，具有出色的效能和效率，它可以快速且有效地處理大型數據集以解決預測問題。XGBoost 被廣泛應用在各種數據科學、機器學習和人工智能領域的競賽和項目中，因為它結合了各種先進的優化技術和算法以實現優越的模型訓練速度和預測準確度。

模型的核心結構與目標函數設計：XGBoost 主要是一種基於梯度提升機制的決策樹模型。模型的目標函數總體上包含了三個部分：Loss Function、Tree Complexity、和 Regularization。這三個部分的結合不僅使得模型在學習過程中追求最佳的預測效果，同時也在防止過擬合問題上起到了很好的作用。

特色的優化技術：XGBoost 引入了 Column Block 的概念，利用基於 Column Block 的 Out-of-core Computing，將資料儲存在硬碟的磁碟上，有效地解決了記憶體無法存儲大數據的問題。此外，XGBoost 還實現了並行處理技術，能夠在構建每個分裂節點時，並行地掃描和比較所有的特徵，進一步提高了運算效率。

自定義彈性與強大的學習能力：XGBoost 提供了豐富的系統配置和超參數設置，讓使用者能夠靈活地定義自己的損失函數和評估函數，並且能夠對模型的複雜度、學習率、樣本權重等方面進行細節調整。它的這些特性，使得模型既能夠處理線性問題，也能夠解決非線性問題，能夠滿足不同的學習。

3.5 輕量梯度提升(LightGBM, Light Gradient Boosting Machine)

LightGBM，全名為 "Light Gradient Boosting Machine"，也是一種以樹為基礎的梯度提升學習算法。這種模型由 Microsoft 提出，與 XGBoost 相似，都是在工業和學術領域被廣泛使用的集成學習方法。與 XGBoost 相比，LightGBM 的主要特色是在保持高預測準確性的同時，進一步提高了計算效率和記憶體使用效率。以下為幾個特點：

樹的成長策略：LightGBM 採用了一種稱為"深度優化"的策略，也被稱為"葉子優化"，這與傳統 GBM 算法採用的"寬度優化"策略（或者叫層級優化）有很大不同。這種策略在每一次分裂時，選擇具有最大梯度損失的葉子進行分裂，可以使模型更快地對誤差進行學習。

優化的特徵選擇：在決定分裂點時，LightGBM 採用了直方圖差分的方法，只需要計算相鄰 bin 之間的差值即可，大大減少了計算量。這種優化策略提高

了處理大型數據集和高維數據的能力。且支持類別特徵：對於具有類別型特徵的數據，LightGBM 可以自動對這些特徵進行轉換和優化，無需事先進行獨熱編碼或者其他形式的轉換。

並行學習與缺失值處理：LightGBM 支持並行學習，可以更快地構建模型。對於缺失值，LightGBM 會自動學習決定將缺失值分配到左子樹還是右子樹，這樣可以更好地處理數據缺失的問題。

3.6 多元羅吉斯迴歸

多元羅吉斯迴歸，是一種被廣泛應用於多分類問題的機器學習模型。它是二元羅吉斯迴歸的擴展，適用於目標變量有多於兩個分類的情況。以下將從多元羅吉斯迴歸的概念、應用、以及相關技術來進行介紹：

模型基礎：多元羅吉斯迴歸模型假設每個類別的對數比率能夠被解釋變數的線性組合所表示。該模型評估每個分類機率與解釋變數間的關係，並利用最大概似估計來進行模型參數的估計。

多元策略：多元羅吉斯迴歸主要有兩種形式，一種是"一對多"（One-vs-Rest, OvR）策略，另一種是"多對多"（Multinomial, Mn）或"Softmax"策略。OvR 策略將一個多分類問題拆分成多個二元分類問題，每個分類都有自己的羅吉斯迴歸模型；Mn 或 Softmax 策略則是直接將多個類別的概率建模為一個多元分布。

應用場景：可以處理各種多分類問題，例如：新聞文章的主題分類、醫學影像的疾病分類、手寫字體的識別等。

模型評估與優化：和其他監督學習模型一樣，多元羅吉斯迴歸的表現可以通過交叉驗證、混淆矩陣、準確度、F1 分數等多種方法進行評估。對於模型的

優化，可以嘗試使用正則化來防止過擬合，或者進行特徵選擇和特徵工程以提高模型的預測能力。

3.7 機器學習特徵選用

張瓊如 (2010) 的研究指出，技術分析是一種根據歷史股價和成交量等數據進行的計算方式，其目的是生成技術指標，以推斷市場未來的走勢和趨向。這種分析方法基於一個重要的假設，那就是歷史會以某種方式重演，而市場的行為模式和趨勢將在未來反映出來。再參照黃光廷 (2002) 的研究，我們深入探討了技術分析和基本分析在投資組合績效上的相關研究，並採用了由他列出的國內外學者普遍使用的指標和參數。我們選取了籌碼和相關的技術指標，包括移動平均線(MA)、隨機指標(KD)、指數平滑異同平均線(MACD)以及相對強弱指數(RSI)。這些技術指標各有其獨特的計算方式和解讀方法，能有效地描繪出股價的走勢和可能的市場轉折點。

移動平均線(MA)能平滑化股價數據，幫助投資者識別長期趨勢；隨機指標(KD)可預測可能的價格反轉，從而為買賣點提供依據；MACD 能反映出股價的中期趨勢，以及可能的買賣訊號；相對強弱指數(RSI)則用於確定市場是否過度買入或過度賣出，提供了市場可能反轉的信號。

基於這些指標，我們加入了一些數值的調整和組合，生成了約 20 個特徵，以便於我們對股票進行更全面、更精確的評估。這種技術指標的選取和調整，旨在將機器學習的精度提升到最高，以更好地服務於投資決策的制定。

表 三-1 特徵選用彙整

籌碼面	個股
三大法人當日買賣超	個股與短中長期均線距離
法人買賣超占比	個股當日成交量/20 日均量
外資	20 天振福
投信	60 天振福
自營商	
融資使用率	
大盤	技術指標
是否在月線上	KD K 值與 D 值
是否在季線上	RSI
成交量	MACD
當年年 K 是否為紅 K	OBV 量能潮
大盤融資餘額	

資料來源:本研究整理

3.8 機器學習時間設計

在股價預測中，機器學習算法已成為一種重要的工具，理論上給的訓練資料越多，能訓練出較好的模型，然而股價資料屬於時間序列的資料，市場也同時有多空的背景，因此預測模型的訓練，我們也必須處理時間序列數據的特性。因此這研究採用**時間窗格法**（Time-window Method）。

時間窗格法是一種涉及連續時間段或序列數據的方法，用於創建訓練和測試數據集。該方法把一個較長的時間序列數據切分成一系列較短的窗格或者區段，每一個窗格包含固定數量的連續數據點。優點是保持了數據的時間序列結構。在股價預測的問題中，因為股票市場的動態和變化往往具有時間序列性質。此外，透過時間窗格法，我們能更好地模擬投資者在實際操作中所面臨的情境。我們可以利用窗格的前一部分數據訓練我們的模型，然後在窗格的後一部分數據上驗證其性能。這種方法也被稱為滾動時間窗格驗證，它能夠更好地模擬模型在實際市場環境中的表現。

本研究的時間範圍為 2007 年~2022 年，採用每 5 年當一訓練集，再往後平移一年當測試集，依此類推，因此本研究總共會有 11 個樣本期間。如表 4-2，最後再將 11 個樣本的結果取平均數。

表 三-2 時間窗格法表格

	訓練集期間	測試集期間
樣本一	20070101-20111231	20120101-20121231
樣本二	20080101-20121231	20130101-20131231
樣本三	20090101-20131231	20140101-20141231
樣本四	20100101-20141231	20150101-20151231
樣本五	20110101-20151231	20160101-20161231
樣本六	20120101-20161231	20170101-20171231
樣本七	20130101-20171231	20180101-20181231
樣本八	20140101-20181231	20190101-20191231
樣本九	20150101-20191231	20200101-20201231
樣本十	20160101-20201231	20210101-20211231
樣本十一	20170101-20211231	20220101-20221231

資料來源:本研究整理

第四章實證結果與分析

4.1 總樣本敘述性統計

本研究之資料來源來自臺灣經濟新報資料庫(以下簡稱 TEJ)，使用還原股價資料庫。本研究選擇了 2007 年 1 月 1 日至 2022 年 12 月 31 日這段時間內台灣上市、上櫃的 1811 檔個股每日股價資料(不含 TDR、ETF)作為樣本。在扣除缺失值後，總資料筆數為 5,458,616 筆。經過本研究所設計的交易策略篩選後，共有 13,754 筆資料符合策略條件。

表 4-1 展示了原始樣本的敘述統計。該表包含了持有天數、勝率、最大報酬率、最小報酬率、中位數報酬率、平均報酬率、標準差以及期間贏大盤勝率等指標。根據表一，可以看到在不同的持有天數下，各項指標的表現有所不同。

對於持有天數為 5、10、20、60 和 120 天的資料，勝率分別為 39.65%、40.40%、41.51%、45.47%和 48.82%。可以觀察到，隨著持有天數的增加，勝率也呈現上升趨勢。此外，最大報酬率隨著持有天數的增加而上升，分別為 57.38%、110.70%、187.29%、319.44%和 798.43%。然而，最小報酬率也隨著持有天數的增加而下降，分別為-33.39%、-40.29%、-54.69%、-65.73%和-81.13%。

中位數報酬率在不同持有天數下表現不一，分別為-1.37%、-1.72%、-2.05%、-1.68%和-0.68%。平均報酬率在持有天數增加的過程中呈現上升趨勢，分別為-0.31%、-0.14%、0.50%、3.05%和 6.82%。標準差隨著持有天數的增加而增加，分別為 7.52%、10.10%、14.28%、24.38%和 38.91%。這表明，在增加持有天數的情況下，報酬率的波動性也有所提高。

在考慮期間贏大盤勝率時，我們可以看到在不同持有天數下，分別為 38.80%、38.82%、39.73%、41.43%和 44.44%。這些數據顯示，在增加持有天數的過程中，期間贏大盤勝率也呈上升趨勢，但是整體而言贏大盤的機率都不超過 50%。

根據表 4-1 的敘述統計數據，可以看出隨著持有天數的增加，勝率和期間贏大盤勝率都有所上升，但報酬率的波動性也隨之增大。此外，研究還發現，隨著持有天數的增加，最大報酬率和最小報酬率之間的差距也在擴大，這意味著風險和收益的潛在變化也隨著持有天數的增加而變大。

綜上所述，本研究的敘述性統計分析揭示了在不同持有天數下，報酬率、勝率以及風險等指標的變化趨勢。這些結果對於瞭解交易策略在不同持有天數下的有效性以及市場行為具有重要的啟示作用。在後續的章節中，將進一步探討加入不同的濾網會得到的結果，以及如何利用這些發現來使做實際的交易策略和決策。

表 四-1 原始樣本敘述統計表

持有天數	5	10	20	60	120
勝率	39.65%	40.40%	41.51%	45.47%	48.82%
最大報酬率	57.38%	110.70%	193.86%	319.44%	798.43%
最小報酬率	-33.39%	-40.29%	-54.69%	-65.73%	-81.13%
中位數報酬率	-1.37%	-1.72%	-2.05%	-1.68%	-0.68%
平均報酬率	-0.31%	-0.14%	0.50%	3.05%	6.84%
標準差	7.52%	10.10%	14.28%	24.38%	38.91%
期間贏大盤勝率	38.80%	38.82%	39.73%	41.43%	44.44%
樣本數	13754				
樣本期間	2007-2022				

資料來源：本研究整理

4.2 大盤季線上揚時的策略表現

本節我們在原有的策略篩選條件基礎上，加入了大盤收盤價在季線上且季線呈現上揚的條件。篩選後的資料筆數 10320 筆。表 4-2 展示了大盤季線上揚條件下，各個持有天數的策略表現。相較於表一，各個持有天數的勝率均有所下降，同時平均報酬率也呈現降低的趨勢。然而，在大盤季線上揚條件下，策略的風險表現較佳，各個持有天數的標準差均有所降低。

此外，期間贏大盤勝率在大盤季線上揚條件下亦呈現下滑趨勢。這些結果顯示，在大盤季線上揚的條件下，應更加謹慎地選擇股票，並可能需要調整策略以提高收益率和降低風險。綜合分析，大盤季線上揚條件下的策略表現相對較差，勝率、平均報酬率和期間贏大盤勝率均有所下降，而風險則稍微降低。

表 四-2 季線上揚敘述統計表

持有天數	5	10	20	60	120
勝率	38.69%	39.13%	40.29%	44.81%	48.24%
最大報酬率	57.38%	110.70%	187.29%	319.44%	798.43%
最小報酬率	-33.39%	-37.28%	-42.90%	-60.44%	-81.13%
中位數報酬率	-1.52%	-1.96%	-2.34%	-2.01%	-0.92%
平均報酬率	-0.42%	-0.27%	0.33%	3.04%	6.63%
標準差	7.43%	10.06%	14.57%	24.96%	39.52%
期間贏大盤勝率	37.81%	37.92%	38.58%	40.42%	43.59%
樣本數	10320				
樣本期間	2007-2022				

資料來源：本研究整理

4.3 突破時成交量明顯放大

這節將加入成交量的濾網，套用在本策略中加入另一個條件為，突破時當天成交量須超過20日均量的兩倍以上和三倍，分別跑敘述統計，表四-4樣本數三倍以上較表四-3兩倍少了2093筆，三倍的整體的勝率和報酬率都略微下滑，也可以得出突破時，成交量大部分至少都為20日均量的2倍或3倍以上，也有可能是因為股票在盤整區間交投較不熱絡，因此成交量也萎縮，當有資金買盤進場時，成交量很容易就爆量。

表 四-3 成交量至少兩倍敘述統計表

持有天數	5	10	20	60	120
勝率	39.15%	40.04%	41.20%	45.21%	48.76%
最大報酬率	57.38%	110.70%	187.29%	297.78%	798.43%
最小報酬率	-33.39%	-40.29%	-54.68%	-63.30%	-81.13%
中位數報酬率	-1.44%	-1.75%	-2.13%	-1.76%	-0.69%
平均報酬率	-0.34%	-0.15%	0.48%	3.03%	6.83%
標準差	7.48%	10.06%	14.31%	24.01%	38.78%
期間贏大盤勝率	38.38%	38.43%	39.42%	41.39%	44.43%
樣本數	12198				
樣本期間	2007-2022				

資料來源：本研究整理。

表 四-4 成交量至少三倍敘述統計表

持有天數	5	10	20	60	120
勝率	38.44%	39.54%	40.71%	44.69%	48.46%
最大報酬率	57.38%	110.70%	187.29%	297.78%	798.43%
最小報酬率	-25.31%	-40.29%	-54.69%	-63.30%	-81.13%
中位數報酬率	-1.51%	-1.84%	-2.20%	-1.89%	-0.78%
平均報酬率	-0.41%	-0.22%	0.41%	2.89%	6.52%
標準差	7.48%	10.06%	14.03%	23.61%	38.53%
期間贏大盤勝率	37.69%	37.74%	38.84%	41.04%	44.38%
樣本數	10106				
樣本期間	2007-2022				

資料來源：本研究整理。

4.4 均線糾結時

本節在最初的策略加入了均線糾結的條件，表四-5 樣本數也大幅減少至 5338 個，可以發現勝率都相較前幾個情境下降許多，但最小報酬率也降低，以及標準差也縮小，可以得知納入這個條件後勝率並沒有提升，有可能是股價也為處於高檔盤整階段，進而遇到假突破而賠錢，因此只納入均線糾結這個條件，並沒有如期的使整體勝率及報酬率來的較好。

表 四-5 均線糾結時敘述統計表

持有天數	5	10	20	60	120
勝率	36.79%	37.51%	39.07%	42.94%	47.65%
最大報酬率	57.14%	83.57%	116.62%	297.78%	798.43%
最小報酬率	-25.31%	-37.13%	-54.69%	-57.97%	-71.82%
中位數報酬率	-1.69%	-2.06%	-2.25%	-2.12%	-1.10%
平均報酬率	-0.60%	-0.61%	0.08%	2.12%	5.48%
標準差	7.23%	9.31%	12.94%	21.90%	36.30%
期間贏大盤勝率	36.09%	35.97%	36.85%	38.90%	43.36%
樣本數	5338				
樣本期間	2007-2022				

資料來源：本研究整理。

4.5 設定停損條件出場

本節採符合篩選條件後買入，持有並再加入一個止損條件為，當日收盤時若是累積損失達條件，則在下一期的開盤價賣出，設定止損條件的好處為可以減少單筆投入因股票大崩跌，而造成損失擴大，當股票下跌 10% 需漲回 11%，回到原點，15% 需要 18%，20% 則需 25%，30% 時，更需要 43%，因此本節使用 10%、15%、20%、30% 四項條件分別去做測試，並持有 120 天，可以從表 4-6 發現，出場條件設的越小，越容易打到止損點出場，因此持有到期天數的個股也越少，可以發現每一項的最大報酬率都相同，表示篩到的個股自從突破後就沒跌超過進場價的 10%。

表 四-6 設定停損條件敘述統計表

出場條件	-10%	-15%	-20%	-30%
勝率	34.40%	42.37%	46.02%	48.25%
最大報酬率	801.13%	801.13%	801.13%	801.13%
最小報酬率	-26.42%	-30.80%	-34.57%	-42.67%
中位數報酬率	-9.82%	-5.45%	-2.34%	-0.81%
平均報酬率	4.73%	5.72%	6.29%	6.64%
標準差	34.31%	36.35%	37.58%	38.40%
持有到期數量	5569	7930	9776	11944
樣本數	14013			
樣本期間	2007-2022			

資料來源：本研究整理。

4.6 策略加入濾網 持有天數比較

此節整理各天數加入濾網後的表現，首先持有 5 天時，可以看到表四-7 勝率，還是最初的策略最高，可以看到最大報酬率前四者都一樣，顯示該個股突破時，大盤站上季線且呈現上揚，成交量也至少三倍以上，，中位數報酬率和平均報酬率，也是最初策略來的好，在標準差的部分，加入了均線糾結時，個股的波動會來的小一點，且最小報酬率也較其他來的好些。

持有 10 天表四-8 的狀況，和 5 天大同小異，也是最初的勝率較佳，且中位數報酬率和平均報酬率也是最初較好，標準差的部分，仍然是加入均線糾結後，來的較低。在短天期中，兩者結果相似，因此在後續機器學習部分，只採用 5 天，來做為短天期機器學習的部分。

以勝率來說，幾乎都是最初策略的勝率和報酬率最高，只有在持有 120 天表四-11，成交量兩倍的濾網中，平均報酬率有略勝 0.01。隨著持有時間越長報

酬率也相對放大，在納入均線糾結這個濾網，其中的標準差都是最低，顯示加入這個濾網對於個股的波動報酬來的較小。下一個章節，將探討加入機器學習來預測，各項濾網的勝率和報酬率變動。

表 四-7 持有 5 天敘述統計表

	最初策略	季線上揚	成交量兩倍	成交量三倍	均線糾結
勝率	39.65%	38.69%	39.15%	38.44%	36.79%
最大報酬率	57.38%	57.38%	57.38%	57.38%	57.14%
最小報酬率	-33.39%	-33.39%	-33.39%	-25.31%	-25.31%
中位數報酬率	-1.37%	-1.52%	-1.44%	-1.51%	-1.69%
平均報酬率	-0.31%	-0.42%	-0.34%	-0.41%	-0.60%
標準差	7.52%	7.43%	7.48%	7.48%	7.23%
期間贏大盤勝率	38.80%	37.81%	38.38%	37.69%	36.09%

資料來源：本研究整理。

表 四-8 持有 10 天敘述統計表

	最初策略	季線上揚	成交量兩倍	成交量三倍	均線糾結
勝率	40.40%	39.13%	40.04%	39.54%	37.51%
最大報酬率	110.70%	110.70%	110.70%	110.70%	83.57%
最小報酬率	-40.29%	-37.28%	-40.29%	-40.29%	-37.13%
中位數報酬率	-1.72%	-1.96%	-1.75%	-1.84%	-2.06%
平均報酬率	-0.14%	-0.27%	-0.15%	-0.22%	-0.61%
標準差	10.10%	10.06%	10.06%	10.06%	9.31%
期間贏大盤勝率	38.82%	37.92%	38.43%	37.74%	35.97%

資料來源：本研究整理。

表 四-9 持有 20 天敘述統計表

	最初策略	季線上揚	成交量兩倍	成交量三倍	均線糾結
勝率	41.51%	40.29%	41.20%	40.71%	39.07%
最大報酬率	193.86%	187.29%	187.29%	187.29%	116.62%
最小報酬率	-54.69%	-42.90%	-54.68%	-54.69%	-54.69%
中位數報酬率	-2.05%	-2.34%	-2.13%	-2.20%	-2.25%
平均報酬率	0.50%	0.33%	0.48%	0.41%	0.08%
標準差	14.28%	14.57%	14.31%	14.03%	12.94%
期間贏大盤勝率	39.73%	38.58%	39.42%	38.84%	36.85%

資料來源：本研究整理。

表 四-10 持有 60 天敘述統計表

	最初策略	季線上揚	成交量兩倍	成交量三倍	均線糾結
勝率	45.47%	44.81%	45.21%	44.69%	42.94%
最大報酬率	319.44%	319.44%	297.78%	297.78%	297.78%
最小報酬率	-65.73%	-60.44%	-63.30%	-63.30%	-57.97%
中位數報酬率	-1.68%	-2.01%	-1.76%	-1.89%	-2.12%
平均報酬率	3.05%	3.04%	3.03%	2.89%	2.12%
標準差	24.38%	24.96%	24.01%	23.61%	21.90%
期間贏大盤勝率	41.43%	40.42%	41.39%	41.04%	38.90%

資料來源：本研究整理。

表 四-11 持有 120 天敘述統計表

	最初策略	季線上揚	成交量兩倍	成交量三倍	均線糾結
勝率	48.82%	48.24%	48.76%	48.46%	47.65%
最大報酬率	798.43%	798.43%	798.43%	798.43%	798.43%
最小報酬率	-81.13%	-81.13%	-81.13%	-81.13%	-71.82%
中位數	-0.68%	-0.92%	-0.69%	-0.78%	-1.10%
平均報酬率	6.82%	6.63%	6.83%	6.52%	5.48%
標準差	38.91%	39.52%	38.78%	38.53%	36.30%
期間贏大盤勝率	44.44%	43.59%	44.43%	44.38%	43.36%

資料來源：本研究整理。

4.7 加入機器學習

4.7.1 原始策略使用機器學習

這章節開始使用機器學習來優化上面的選股策略，使用三種模型：極限梯度提升(XGBoost)、輕量化梯度提升(Light GBM)、多元羅吉斯迴歸(Logistic Rgression)並加入特徵做機器學習選股的預測，將前述持有至到期日，以及加入前述章節其他條件濾網，一一使用機器學習來做預測，並分別比較各個勝率以及報酬率。

這章節將使用時間窗格法來做預測，從 2007 年開始五年，然後往下測試一年，依此類推至 2017 訓練至 2021，最後測驗期為 2022 年 12 月 31 日。並且使用三種閾值 0.5、0.55、0.6，分別檢驗測試。

表 四-12 原始策略時間窗格法資料筆數

	訓練集期間	訓練集筆數	測試集期間	測試集筆數
樣本一	20070101-20111231	3578	2012	430
樣本二	20080101-20121231	3481	2013	1228
樣本三	20090101-20131231	4139	2014	1488
樣本四	20100101-20141231	5095	2015	867
樣本五	20110101-20151231	4935	2016	591
樣本六	20120101-20161231	4604	2017	1037
樣本七	20130101-20171231	5211	2018	921
樣本八	20140101-20181231	4904	2019	751
樣本九	20150101-20191231	4167	2020	956
樣本十	20160101-20201231	4256	2021	1148
樣本十一	20170101-20211231	4813	2022	762

資料來源：本研究整理。

本節在探討利用極限梯度提升（XGBoost）、輕量化梯度提升（Light GBM）和多元羅吉斯迴歸（Logistic Regression）三種機器學習模型在選股策略上的表現。基於不同閾值和持有天數，我們分析了這些模型的平均報酬和勝率，並將其與未使用機器學習時進行了比較，本節開始均列舉樣本一至十一的平均報酬及報酬率對照同期未使用機器學習(以下稱原始)之數據，。

根據表 4.13. 閾值 0.5 的結果，多元羅吉斯迴歸在持有 5 天的情況下表現最佳，平均報酬為 1.62%，勝率達到 58.96%，明顯優於其他兩種模型和大盤指數。在持有 20 天和 60 天的情況下，多元羅吉斯迴歸的平均報酬分別達到 1.62%和 5.24%，勝率分別為 49.30%和 49.26%，仍然優於其他模型。然而，在持有 120

天的情況下，極限梯度提升的平均報酬和勝率分別達到 7.33% 和 50.56%，略高於多元羅吉斯迴歸，但 120 天的部分報酬率則都輸給原始。

在表 4.14 閾值 0.55 的情況下，多元羅吉斯迴歸在持有 5 天的情況下依然表現最佳，平均報酬為 2.29%，勝率更達到 62.24%。此外，在持有 20 天、60 天和 120 天的情況下，多元羅吉斯迴歸的平均報酬和勝率也均較高。在持有 60 天的部分因為閾值提高至 0.55，三個模型的報酬率表現均有大幅增加。

甚至多元羅吉斯的 60 天報酬率更勝過 120 天。然而，在表 4.15 閾值 0.6 的情況下，多元羅吉斯迴歸的持有 5 天部分勝率也是提高至 60%，但勝率和報酬率卻較閾值 0.55 時來的低。但閾值 0.6 圖 4-5、圖 4-6 的時候，多元羅吉斯的勝率和報酬率都是明顯優於其他兩者。綜合以上分析，可以看出多元羅吉斯迴歸在大部分情況下均表現較好，特別是在短期持有期內，但也可以發現多元羅吉斯的交易次數是相對其他來的少很多。在不同閾值下，極限梯度提升極限梯度提升也表現出一定的競爭力。

特別是在閾值 0.5 和 0.55 圖 4-1~圖 4-4 的情況下，極限梯度提升在長期持有期（120 天）內的平均報酬和勝率均略高於多元羅吉斯迴歸。這表明在不同的閾值和持有天數下，選擇合適的機器學習模型對投資策略的成功至關重要。

輕量化梯度提升（Light GBM）相對而言在各個閾值和持有天數下的表現較為普通，並未在任何情況下明顯優於其他兩種模型。然而，考慮到不同模型可能在不同的市場環境和投資策略下表現出不同的特點，因此在實際應用中，投資者仍需根據具體情況進行適當的選擇。

此外，原資料作為一個基準，我們可以看到，在大部分情況下，機器學習模型的投資策略有著更高的平均報酬和勝率。這說明機器學習模型在選股策略上具有一定的優勢，值得進一步研究和應用。本節分析了極限梯度提升、輕量化梯度提升和多元羅吉斯迴歸三種機器學習模型在選股策略上的表現。

結果表明，多元羅吉斯迴歸在大部分情況下表現較好，尤其是在短期持有期內。然而，在某些情況下，極限梯度提升也具有競爭力。因此，投資者在制定投資策略時，應根據具體市場環境和投資目標，選擇合適的機器學習模型進行選股，以提高投資回報。

表 四-13 原始策略閾值 0.5

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	0.11%	45.24%	1112	-0.18%	41.27%	2014	1.62%	58.96%	197	-0.39%	38.39%
持有 20 天	1.02%	44.29%	2245	0.15%	41.21%	2612	1.62%	49.30%	599	0.43%	40.69%
持有 60 天	3.88%	45.30%	2936	2.76%	43.39%	3589	5.24%	49.26%	1880	2.94%	44.34%
持有 120 天	7.33%	50.56%	5007	6.30%	48.08%	5065	5.94%	49.45%	5020	7.66%	49.19%

資料來源：本研究整理。

表 四-14 原始策略閾值 0.55

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	0.54%	45.72%	590	-0.02%	42.14%	1257	2.29%	62.24%	66	-0.39%	38.39%
持有 20 天	1.04%	43.28%	1009	0.20%	41.97%	1696	3.68%	54.68%	185	0.43%	40.69%
持有 60 天	5.60%	48.70%	1330	2.73%	44.57%	2447	7.75%	53.23%	372	2.94%	44.34%
持有 120 天	8.40%	52.20%	3140	6.85%	48.49%	3883	7.23%	51.55%	2609	7.66%	49.19%

資料來源：本研究整理。

表 四-15 原始策略閾值 0.6

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	0.56%	46.95%	239	0.26%	43.67%	762	1.61%	60.00%	32	-0.39%	38.39%
持有 20 天	0.18%	43.59%	298	-0.04%	42.28%	1032	3.62%	52.80%	70	0.43%	40.69%
持有 60 天	5.23%	48.30%	643	3.55%	45.17%	1561	7.08%	64.04%	80	2.94%	44.34%
持有 120 天	6.90%	52.18%	968	7.73%	49.59%	2745	14.23%	56.07%	860	7.66%	49.19%

資料來源：本研究整理。

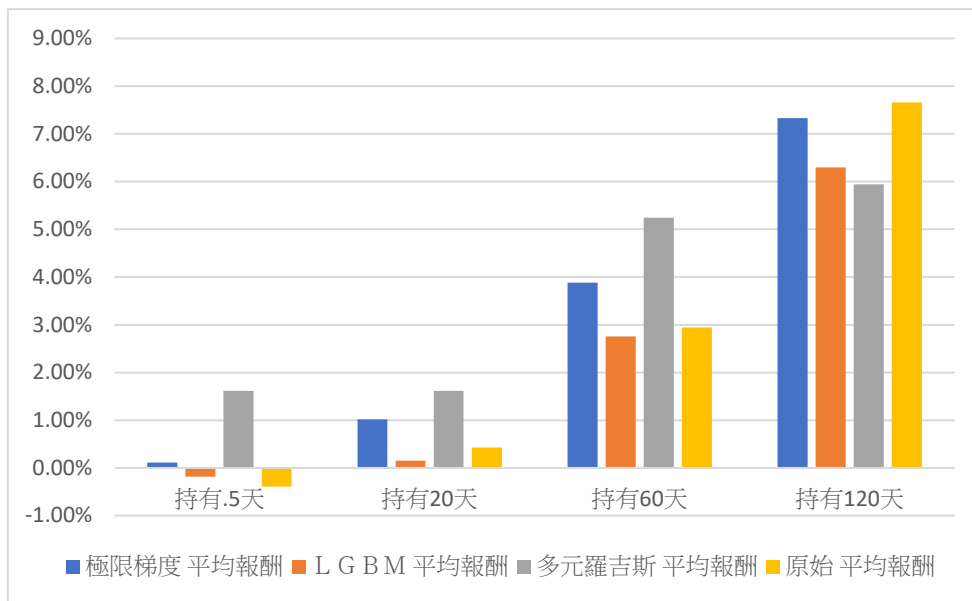


圖 四-1 閾值 0.5 報酬率對照

資料來源：本研究整理。

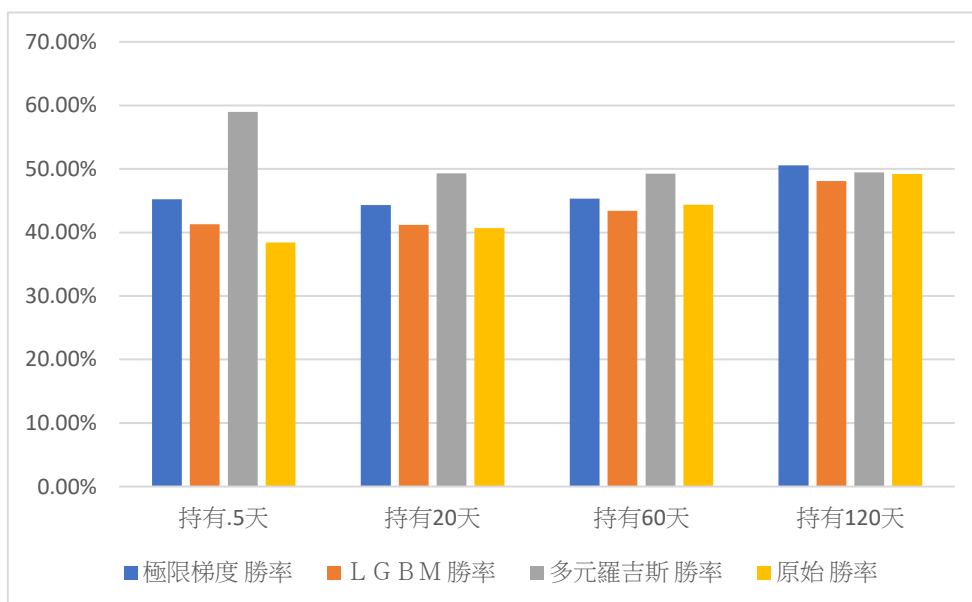


圖 四-2 閾值 0.5 勝率對照

資料來源：本研究整理。

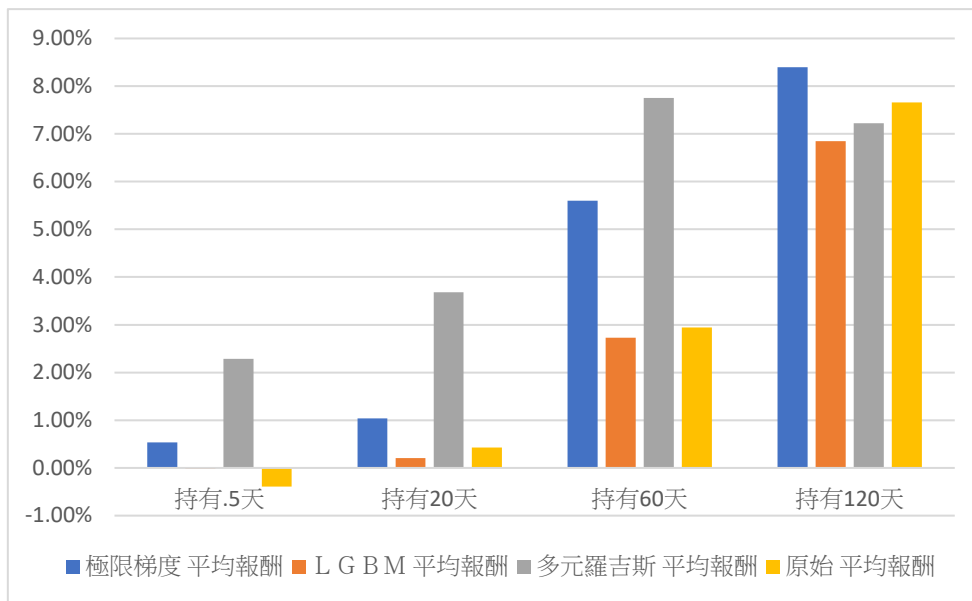


圖 四-3 閾值 0.55 報酬率對照

資料來源：本研究整理。

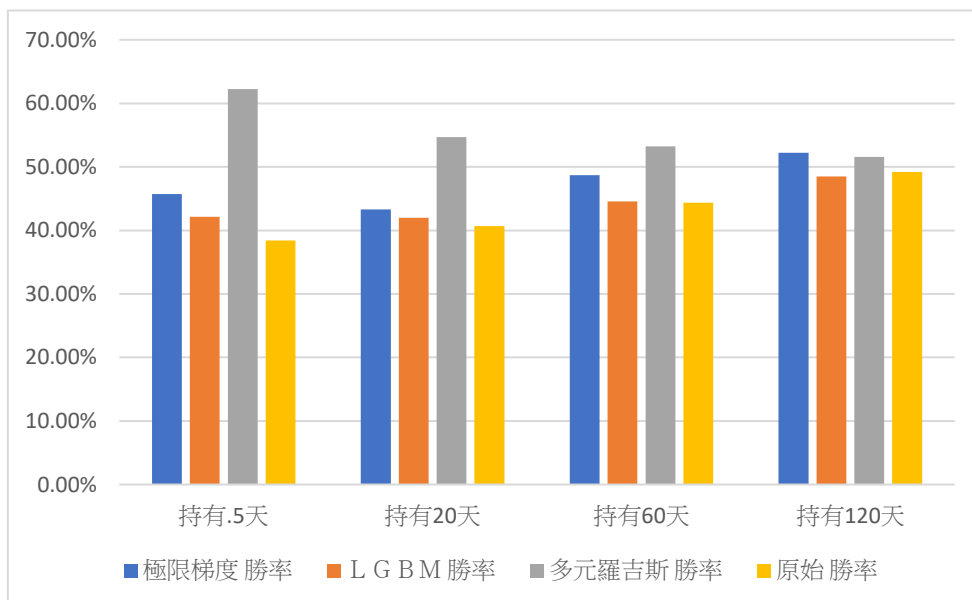


圖 四-4 閾值 0.55 勝率對照

資料來源：本研究整理。

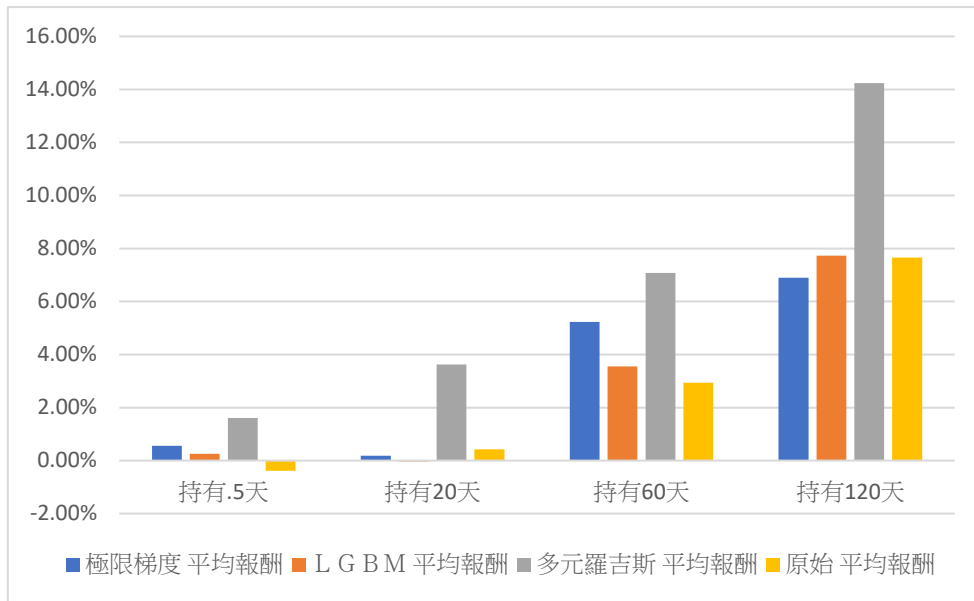


圖 四-5 閾值 0.6 報酬率對照

資料來源：本研究整理。

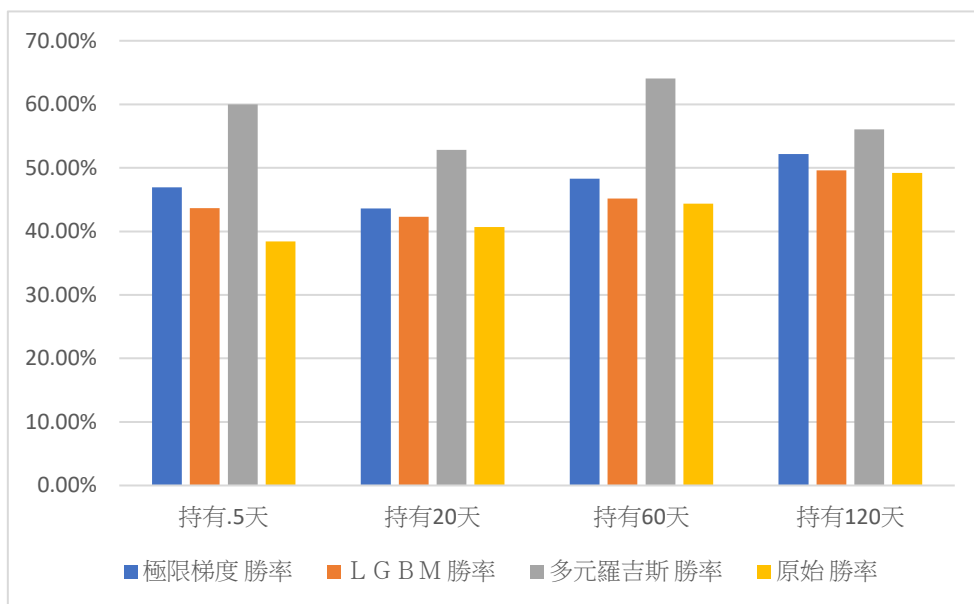


圖 四-6 閾值 0.6 勝率對照

資料來源：本研究整理。

4.7.2 大盤季線上揚加入機器學習

這一節中，我們將研究大盤站上季線且大盤季線上揚符合條件的個股，極限梯度、LGBM 和多元羅吉斯模型的表現。在表 4-17、圖 4-7、圖 4-8 閾值 0.5 時，多元羅吉斯的模型，持有 5 天勝率高達 73.35% 和平均報酬率 5.61%，優於其他兩者，以及持有長天期的報酬率。LGBM 反而更比原先未使用機器學習的報酬率和勝率來的低。持有 20 天，多元羅吉斯和極限梯度的效果較好，60 天的部分則是多元羅吉斯較佳，在這邊可以觀察到這個策略，在閾值 0.5 時，長期 120 天持有的情況下，機器學習的效果都不是太好，還是原始的報酬率勝率較佳。

閾值提高至 0.55 表 4-18、圖 4-9、圖 4-10 時，原先的多元羅吉斯模型 5 天的報酬率和勝率都比閾值 0.5 時來的低，交易次數也減少許多，但整體還是贏原始與其他兩者。LGBM 在 20 天和 5 天都輸原始未使用機器學習的績效，60 天的時候多元多吉斯的效果最好，持有到 120 天時，則是未使用機器學習的勝率、報酬率最佳。閾值提高至 0.6 表 4-19、圖 4-11、圖 4-12 時，三者平均報酬率與勝率均有提升，特別是多元羅吉斯的 120 天平均報酬上升至 9.83%，交易次數也隨著閾值拉高而大幅減少。但是季線上揚使用機器學習預測，長天期 120 天的情況下，普遍效果都沒有較原始來的好，反而短中期使用機器學習提升的效果較為明顯。

表 四-16 季線上揚時間窗格法資料筆數

	訓練集期間	訓練集筆數	測試集期間	測試集筆數
樣本一	20070101-20111231	2873	2012	214
樣本二	20080101-20121231	2614	2013	981
樣本三	20090101-20131231	3137	2014	1301
樣本四	20100101-20141231	4063	2015	659
樣本五	20110101-20151231	3867	2016	320
樣本六	20120101-20161231	3475	2017	881
樣本七	20130101-20171231	4142	2018	675
樣本八	20140101-20181231	3836	2019	291
樣本九	20150101-20191231	2826	2020	614
樣本十	20160101-20201231	2781	2021	1105
樣本十一	20170101-20211231	3566	2022	407

資料來源：本研究整理。

表 四-17 季線上揚閾值 0.5

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	0.29%	43.81%	985	-0.60%	37.05%	1566	5.61%	73.35%	73	-0.54%	37.27%
持有 20 天	0.22%	39.79%	1305	-0.14%	39.52%	1909	0.28%	36.22%	253	0.01%	38.87%
持有 60 天	2.13%	43.48%	2198	1.90%	40.17%	2586	3.23%	44.49%	967	2.34%	42.67%
持有 120 天	6.58%	48.66%	3469	5.33%	46.21%	3673	4.64%	47.07%	3477	6.86%	47.76%

資料來源：本研究整理。

表 四-18 季線上揚閾值 0.55

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	-0.07%	42.30%	506	-0.30%	37.98%	1038	3.39%	53.33%	22	-0.54%	37.27%
持有 20 天	0.72%	40.01%	937	-0.21%	39.26%	1336	0.41%	29.63%	46	0.01%	38.87%
持有 60 天	2.68%	41.61%	1136	1.99%	39.58%	1841	3.57%	54.58%	126	2.34%	42.67%
持有 120 天	6.56%	47.52%	2118	5.11%	45.86%	2848	4.63%	48.24%	1539	6.86%	47.76%

資料來源：本研究整理。

表 四-19 季線上揚閾值 0.6

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	0.04%	48.18%	195	-0.16%	39.16%	641	3.16%	49.05%	14	-0.54%	37.27%
持有 20 天	0.58%	42.06%	463	-0.07%	38.15%	871	2.66%	50.00%	19	0.01%	38.87%
持有 60 天	2.29%	47.92%	474	1.85%	39.32%	1255	3.73%	65.00%	18	2.34%	42.67%
持有 120 天	5.92%	48.23%	1179	5.59%	47.20%	2124	9.83%	50.67%	475	6.86%	47.76%

資料來源：本研究整理。

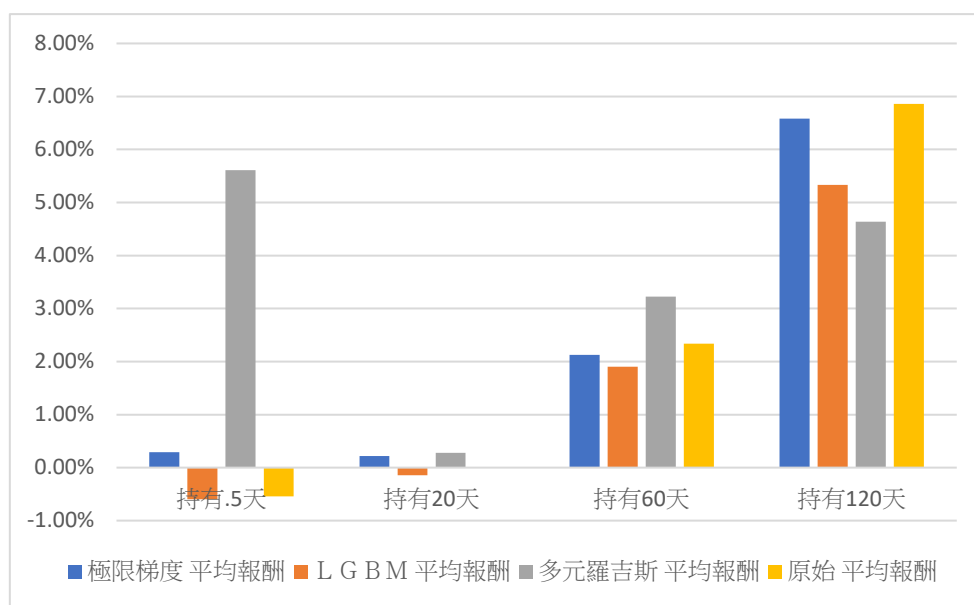


圖 四-7 閾值 0.5 季線上揚報酬率對照

資料來源：本研究整理。

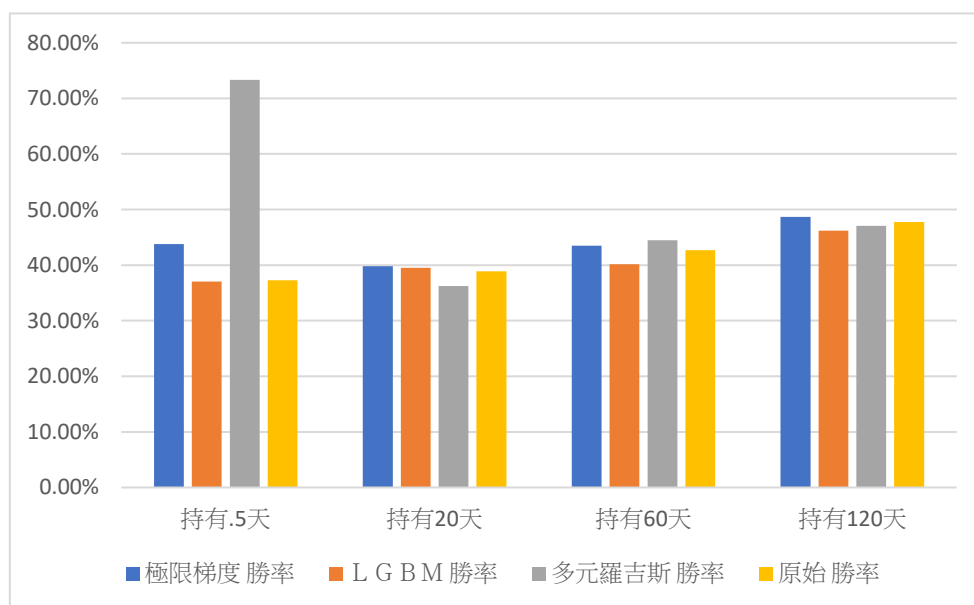


圖 四-8 閾值 0.5 季線上揚勝率對照

資料來源：本研究整理。

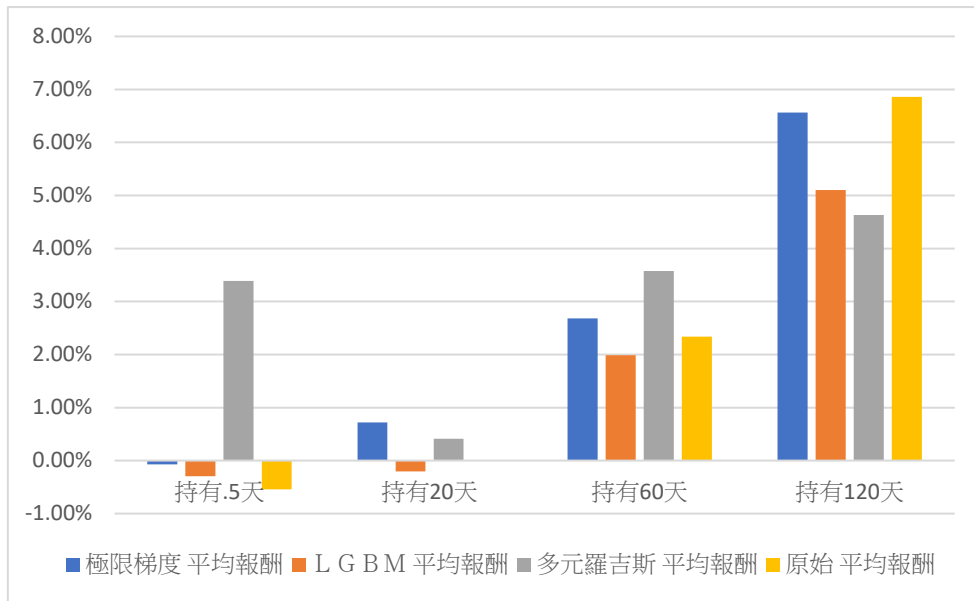


圖 四-9 閾值 0.55 季線上揚報酬率對照

資料來源：本研究整理。

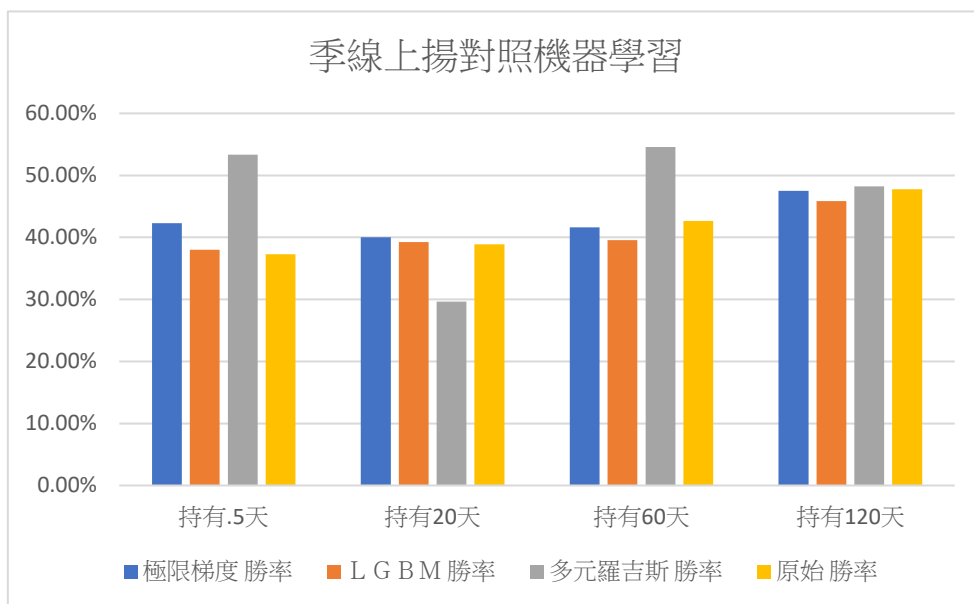


圖 四-10 閾值 0.55 季線上揚勝率對照

資料來源：本研究整理。

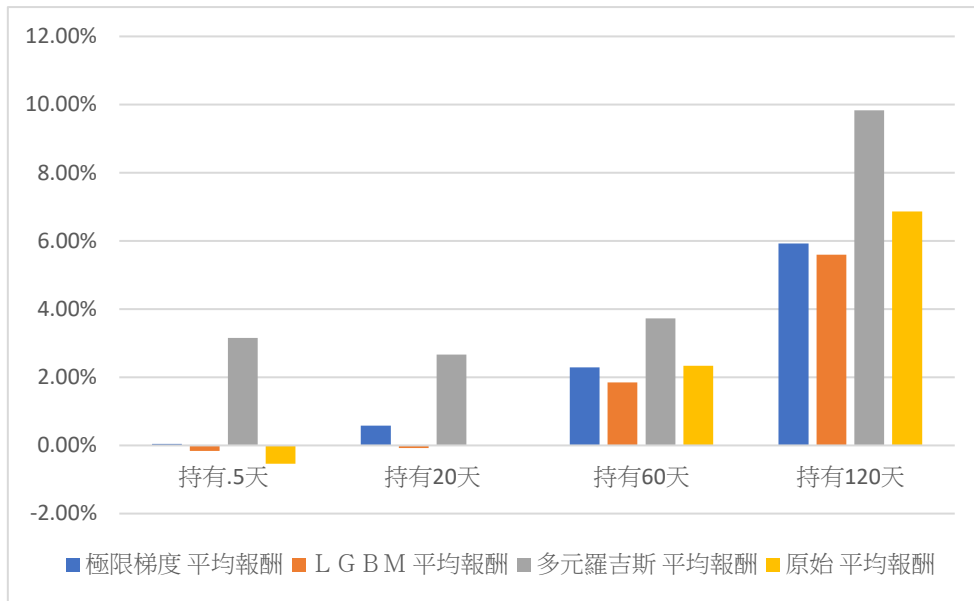


圖 四-11 閾值 0.6 季線上揚報酬率對照

資料來源：本研究整理。

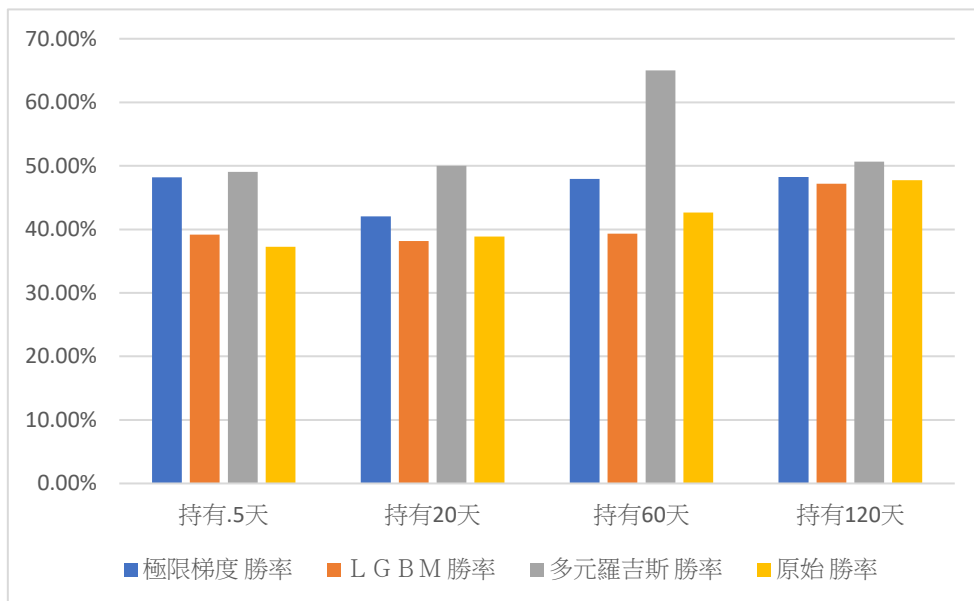


圖 四-12 閾值 0.6 季線上揚勝率對照

資料來源：本研究整理。

4.7.3 成交量放大

這小節將探討在成交量上升情況下，極限梯度、LGBM 和多元羅吉斯模型的表現。以下表格展示了不同閾值下成交量增長的機器學習模型預測結果。我們將對比不同閾值下的平均報酬和勝率。從表 4-22 到表 4-27、圖 4-13 到圖 4-24，我們可以觀察到在閾值從 0.5 到 0.6 的情況下，極限梯度、LGBM、多元羅吉斯和原始情況下的平均報酬和勝率均有所變化。

在兩倍成交量下，閾值為 0.5 時表 4-23，持有 5 天的平均報酬和勝率最高分別為 3.99% 和 62.53%，出現在多元羅吉斯情況下，遠大於其他兩者與未使用機器學習，在 0.5 的閾值下，短天期 5 天、20 天的效果多元羅吉斯均為最佳，60 天、120 天則是極限梯度模型。

閾值提高至 0.55 表 4-24，多元羅吉斯的表現依舊在 5 天、20 天最佳，60 天的報酬率也是三者較高，但是勝率也隨之下降不少。

隨著閾值提高到 0.6 表 4-25，持有 5 天的最高平均報酬和勝率分別上升至 5.00% 和 71.30%，同樣出現在多元羅吉斯的模型，平均報酬率與勝率在 0.6 的閾值，均是多元羅吉斯最佳。

在三倍成交量下，閾值為 0.5 表 4-26 時，持有 5 天的平均報酬和勝率最高分別為 4.27% 和 67.85%，出現在多元羅吉斯。

閾值提高到 0.55 表 4-27，多元羅吉斯的 5 天 20 天還是表現最好，但在 60 天時的平均報酬率卻是大幅下降。60 天和 120 天的報酬則是極限梯度模型表現最好。

隨著閾值提高到 0.6 表 4-28，持有 5 天的最高平均報酬和勝率分別上升至 3.92% 和 64.06%，仍然是在多元羅吉斯，60 天依舊大幅下降，並翻為負值。60 天和 120 天為極限梯度表現最佳。

綜合以上，在不同閾值下，各個模型的表現有所不同，多元羅吉斯模型在持有天數較短時的表現較好，而在持有天數較長時，極限梯度和多元羅吉斯模型的表現均有所提升，但極限梯度的模型表現更為優異。隨著閾值的提高，各個模型在不同持有天數下的平均報酬和勝率均呈現出一定程度的上升趨勢。因此，在選擇策略時，需根據自身的風險承受能力和投資目標，適當調整閾值以達到最佳投資效果。

表 四-20 兩倍成交量時間窗格法資料

	訓練集期間	訓練集筆數	測試集期間	測試集筆數
樣本一	20070101-20111231	2929	2012	351
樣本二	20080101-20121231	2833	2013	1117
樣本三	20090101-20131231	3513	2014	1357
樣本四	20100101-20141231	4574	2015	794
樣本五	20110101-20151231	4455	2016	524
樣本六	20120101-20161231	4143	2017	986
樣本七	20130101-20171231	4778	2018	867
樣本八	20140101-20181231	4528	2019	691
樣本九	20150101-20191231	3862	2020	837
樣本十	20160101-20201231	3905	2021	1042
樣本十一	20170101-20211231	4423	2022	704

資料來源：本研究整理

表 四-21 三倍成交量時間窗格法資料

	訓練集期間	訓練集筆數	測試集期間	測試集筆數
樣本一	20070101-20111231	2170	2012	251
樣本二	20080101-20121231	2080	2013	908
樣本三	20090101-20131231	2693	2014	1138
樣本四	20100101-20141231	3645	2015	672
樣本五	20110101-20151231	3632	2016	454
樣本六	20120101-20161231	3423	2017	883
樣本七	20130101-20171231	4055	2018	788
樣本八	20140101-20181231	3935	2019	594
樣本九	20150101-20191231	3391	2020	726
樣本十	20160101-20201231	3445	2021	892
樣本十一	20170101-20211231	3883	2022	630

資料來源：本研究整理

表 四-22 成交量兩倍閾值 0.5

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	0.35%	45.00%	1400	0.07%	41.48%	1908	3.99%	62.53%	141	-0.41%	38.14%
持有 20 天	0.49%	43.51%	1684	0.34%	41.32%	2382	1.28%	48.53%	491	0.36%	40.45%
持有 60 天	3.72%	46.97%	2512	2.26%	42.64%	3215	2.87%	45.80%	1524	2.73%	44.11%
持有 120 天	7.15%	50.02%	4935	6.82%	49.07%	4658	5.51%	49.12%	4623	7.38%	48.96%

資料來源：本研究整理。

表 四-23 成交量兩倍閾值 0.55

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	-0.42%	40.87%	556	0.07%	40.20%	1182	2.39%	61.14%	37	-0.41%	38.14%
持有 20 天	0.24%	41.31%	791	0.21%	41.78%	1580	3.27%	48.25%	111	0.36%	40.45%
持有 60 天	2.79%	45.14%	1540	2.26%	42.92%	2221	3.33%	43.66%	256	2.73%	44.11%
持有 120 天	7.89%	51.11%	2504	7.43%	49.71%	3616	6.57%	51.44%	2397	7.38%	48.96%

資料來源：本研究整理。

表 四-24 成交量兩倍閾值 0.6

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	1.25%	51.15%	322	0.33%	40.58%	685	5.00%	71.30%	22	-0.41%	38.14%
持有 20 天	0.10%	40.92%	444	0.32%	42.19%	983	5.98%	55.28%	32	0.36%	40.45%
持有 60 天	2.56%	53.58%	588	2.14%	42.67%	1443	4.54%	61.55%	40	2.73%	44.11%
持有 120 天	7.60%	47.82%	1929	7.51%	50.28%	2567	10.46%	57.60%	823	7.38%	48.96%

資料來源：本研究整理。

表 四-25 成交量三倍閾值 0.5

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	0.48%	44.25%	1263	-0.33%	40.35%	1739	4.27%	67.85%	172	-0.46%	37.63%
持有 20 天	1.32%	43.65%	1550	-0.30%	39.32%	2077	-0.56%	43.11%	442	0.29%	39.75%
持有 60 天	3.31%	44.39%	2086	1.90%	41.50%	2850	2.98%	45.51%	1132	2.62%	43.63%
持有 120 天	6.96%	50.12%	3815	5.84%	47.43%	3968	5.36%	48.30%	3935	7.23%	48.58%

資料來源：本研究整理。

表 四-26 成交量三倍閾值 0.55

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	-0.19%	43.93%	611	0.02%	41.89%	1186	3.61%	70.67%	49	-0.46%	37.63%
持有 20 天	0.22%	41.63%	1045	0.02%	40.61%	1424	2.61%	49.53%	126	0.29%	39.75%
持有 60 天	3.25%	44.00%	943	2.18%	42.16%	2053	0.61%	44.56%	232	2.62%	43.63%
持有 120 天	8.43%	50.55%	1910	6.18%	47.95%	3060	6.09%	49.09%	2107	7.23%	48.58%

資料來源：本研究整理。

表 四-27 成交量三倍閾值 0.6

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	-0.15%	39.96%	187	-0.55%	37.89%	738	3.92%	64.06%	25	-0.46%	37.63%
持有 20 天	-1.02%	35.54%	278	0.15%	41.75%	937	1.55%	51.36%	36	0.29%	39.75%
持有 60 天	4.40%	44.35%	341	1.99%	41.95%	1368	-0.57%	49.81%	32	2.62%	43.63%
持有 120 天	8.12%	49.59%	1358	6.09%	48.13%	2274	11.31%	56.21%	775	7.23%	48.58%

資料來源：本研究整理。

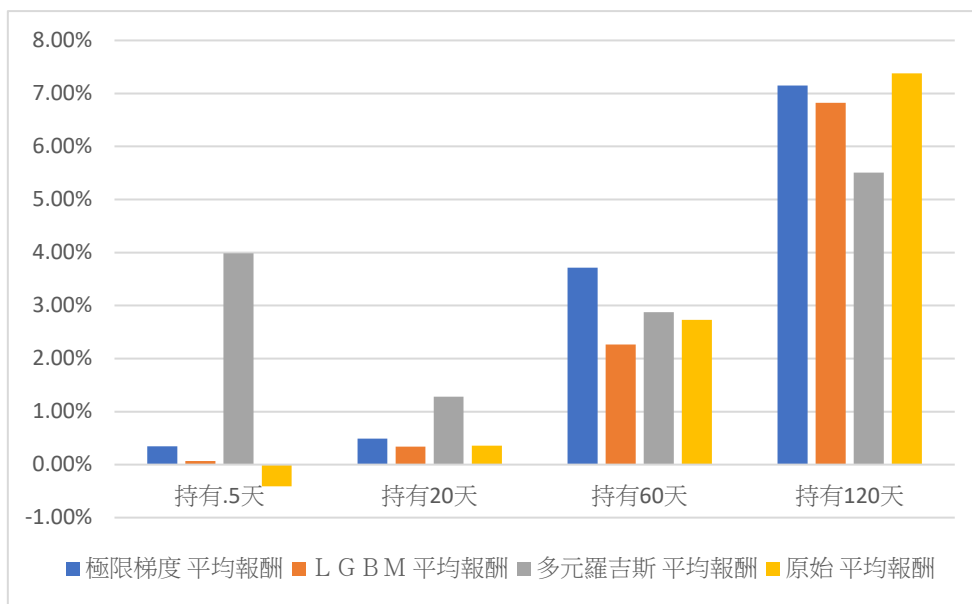


圖 四-13 閾值 0.5 兩倍成交量報酬率對照

資料來源：本研究整理。

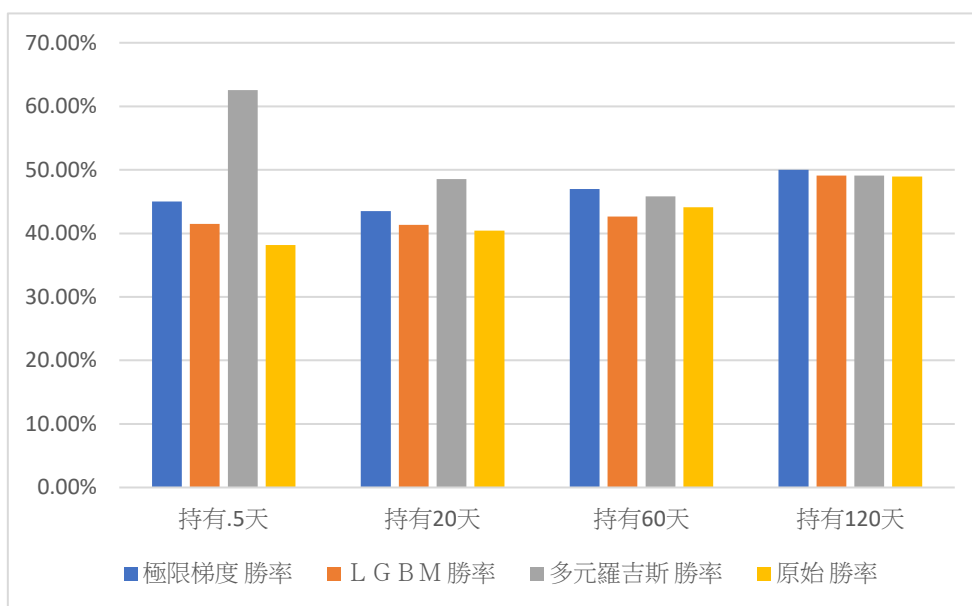


圖 四-14 閾值 0.5 兩倍成交量勝率對照

資料來源：本研究整理。

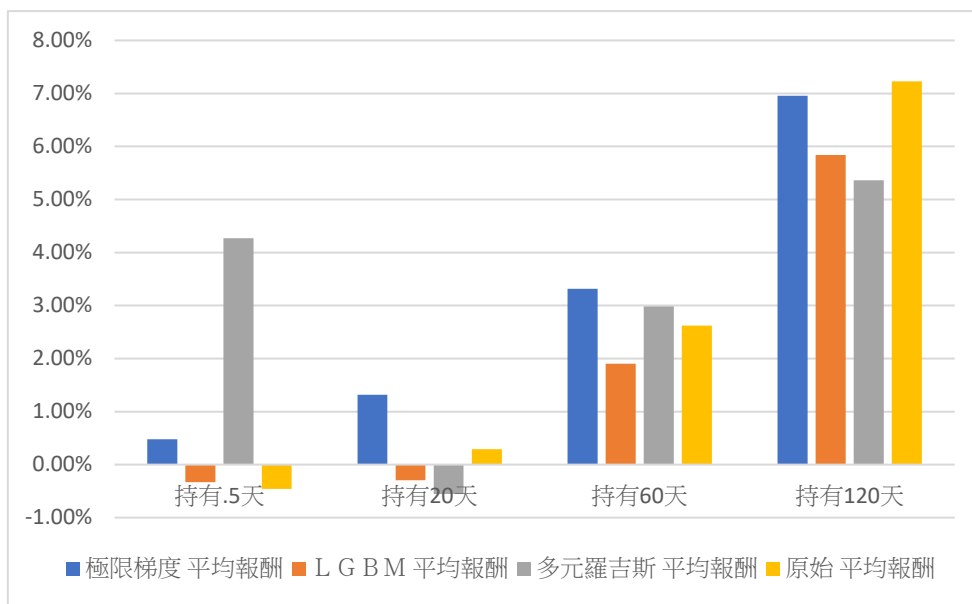


圖 四-15 閾值 0.5 三倍成交量報酬率對照

資料來源：本研究整理。

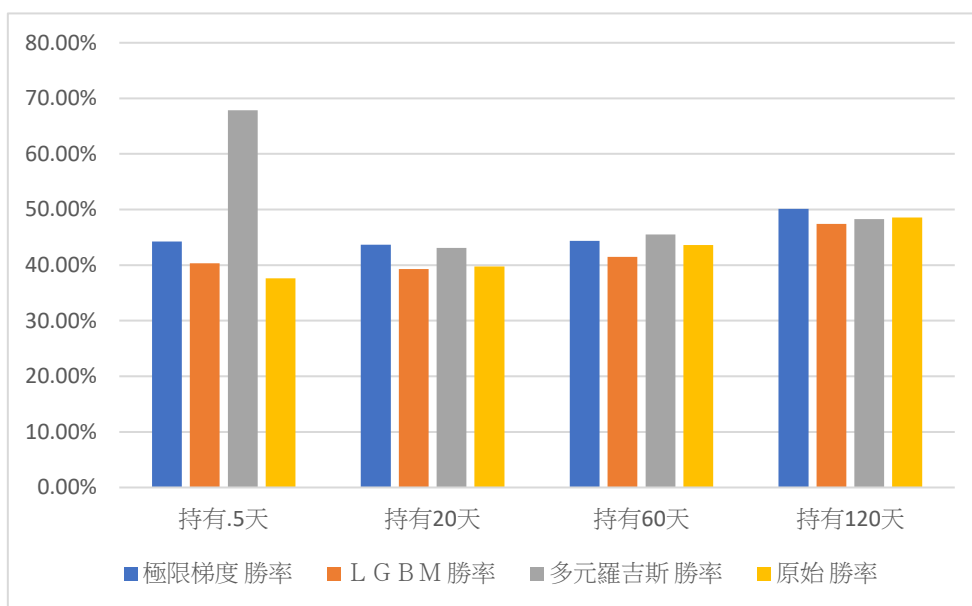


圖 四-16 閾值 0.5 三倍成交量勝率對照

資料來源：本研究整理。

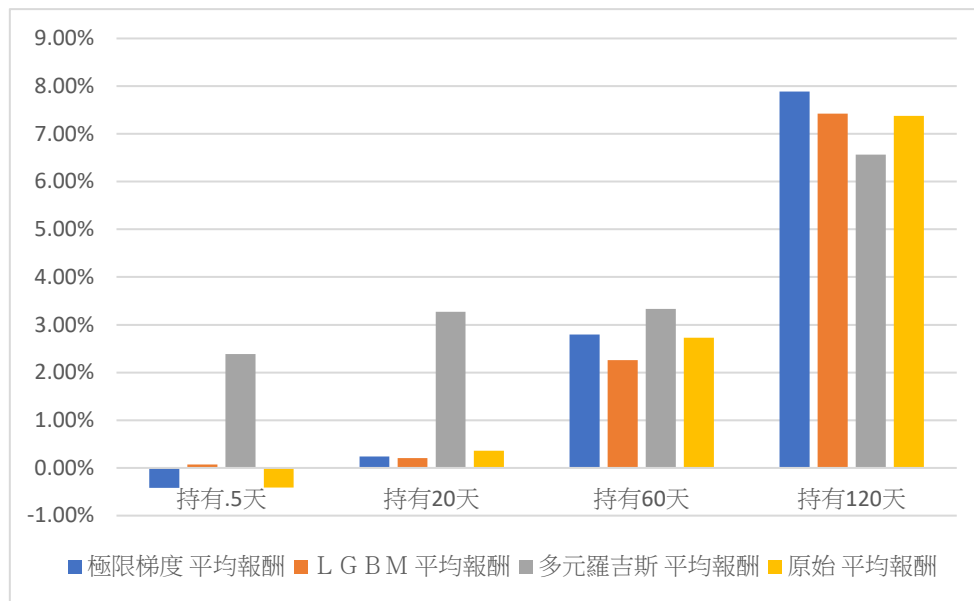


圖 四-17 閾值 0.55 兩倍成交量報酬率對照

資料來源：本研究整理。

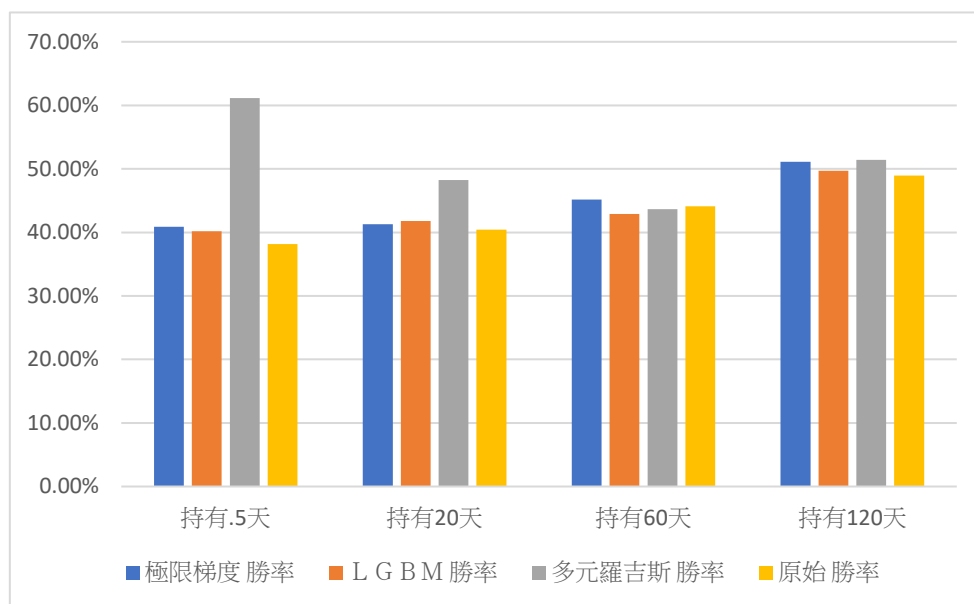


圖 四-18 閾值 0.55 兩倍成交量勝率對照

資料來源：本研究整理。

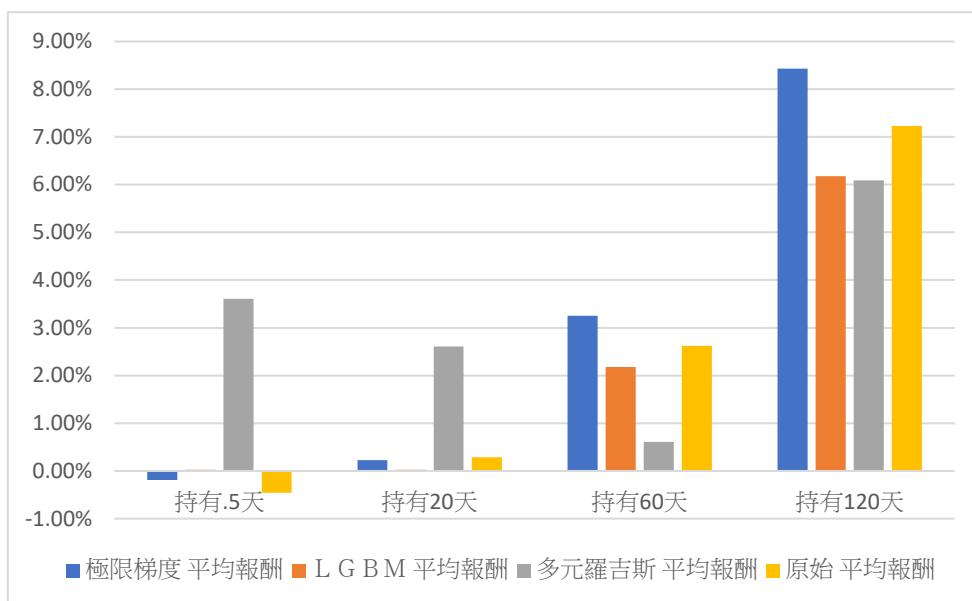


圖 四-19 閾值 0.55 三倍成交量報酬率對照

資料來源：本研究整理。

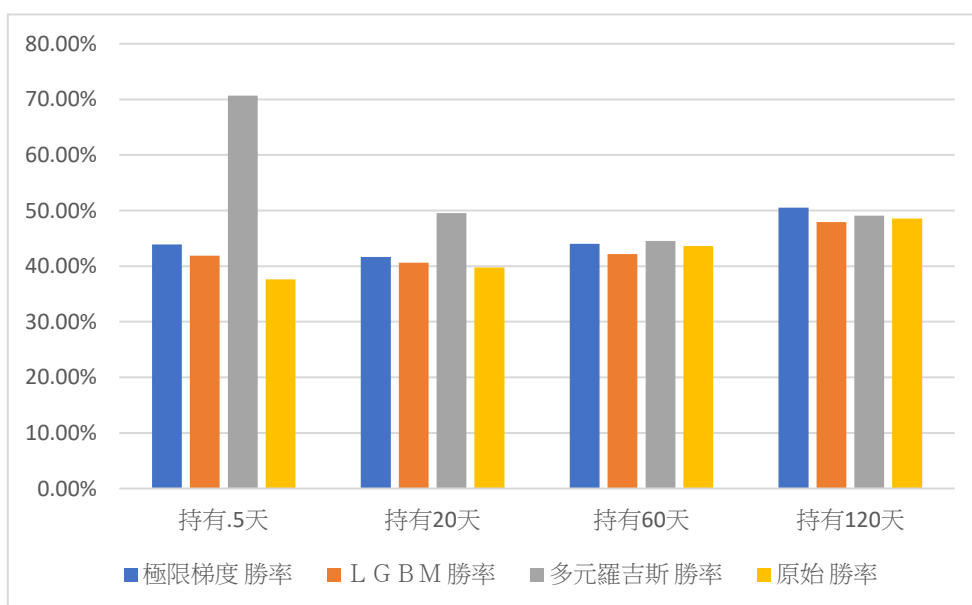


圖 四-20 閾值 0.55 三倍成交量勝率對照

資料來源：本研究整理。

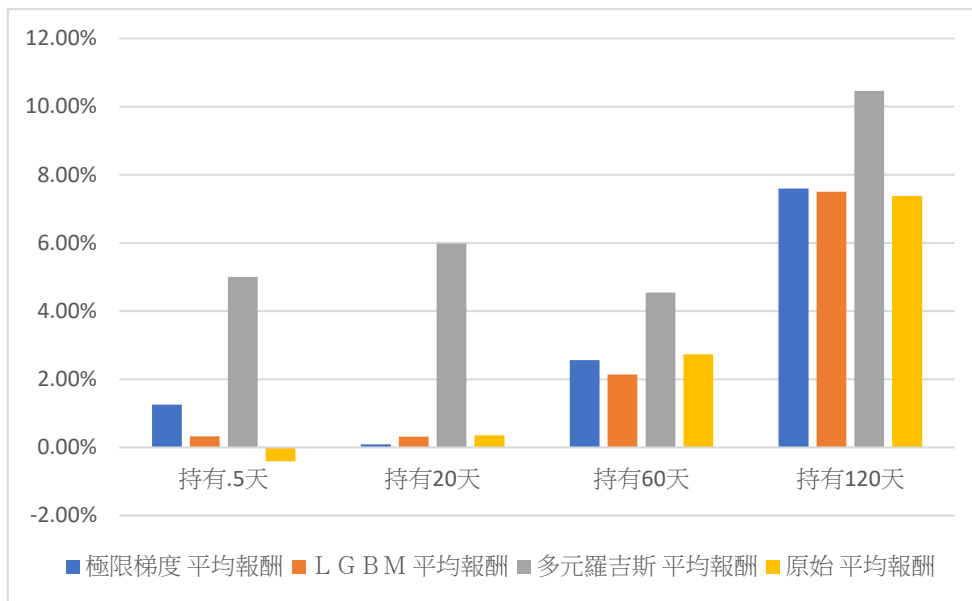


圖 四-21 閾值 0.6 二倍成交量報酬率對照

資料來源：本研究整理。

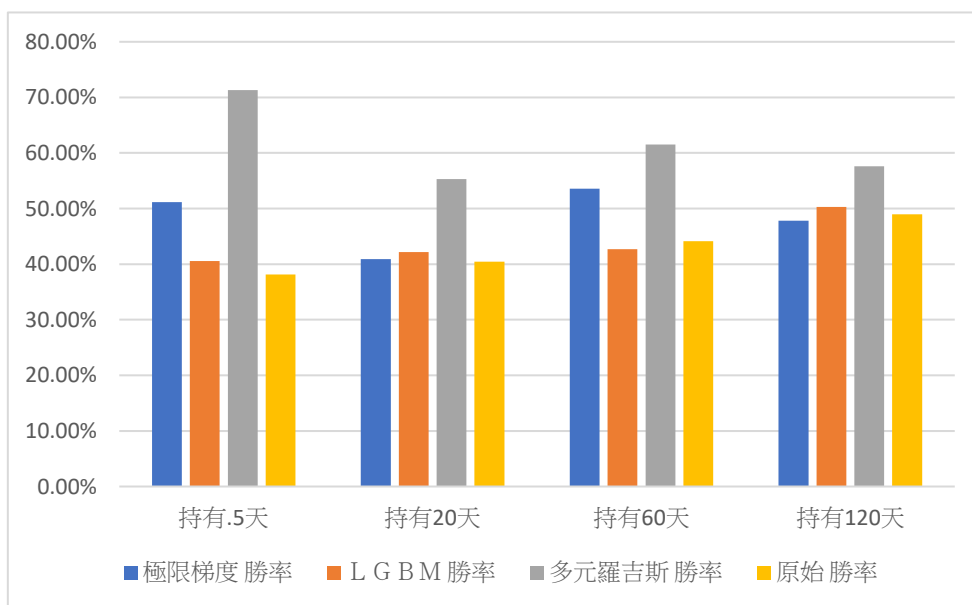


圖 四-22 閾值 0.6 二倍成交量勝率對照

資料來源：本研究整理。

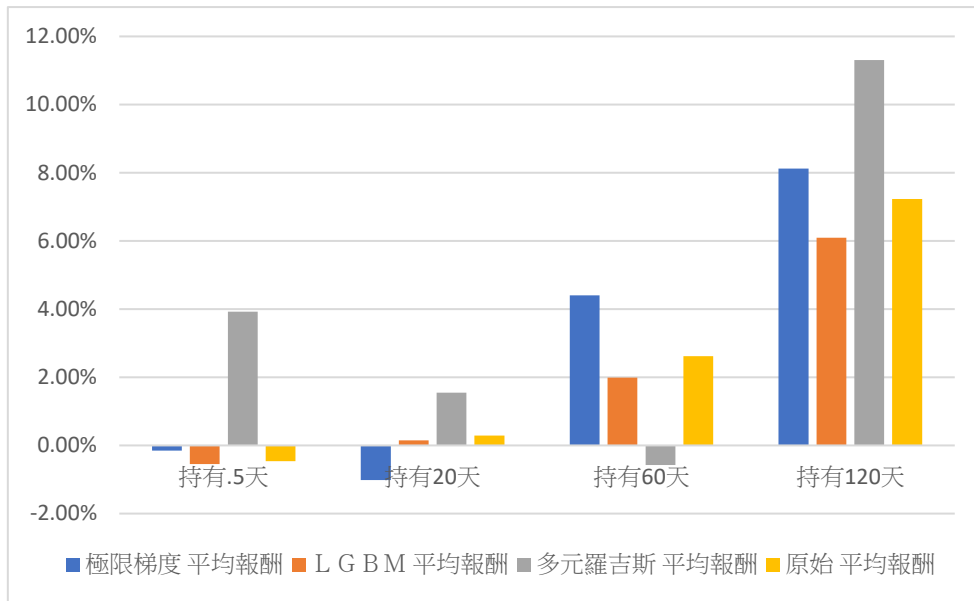


圖 四-23 閾值 0.6 三倍成交量報酬率對照

資料來源：本研究整理。

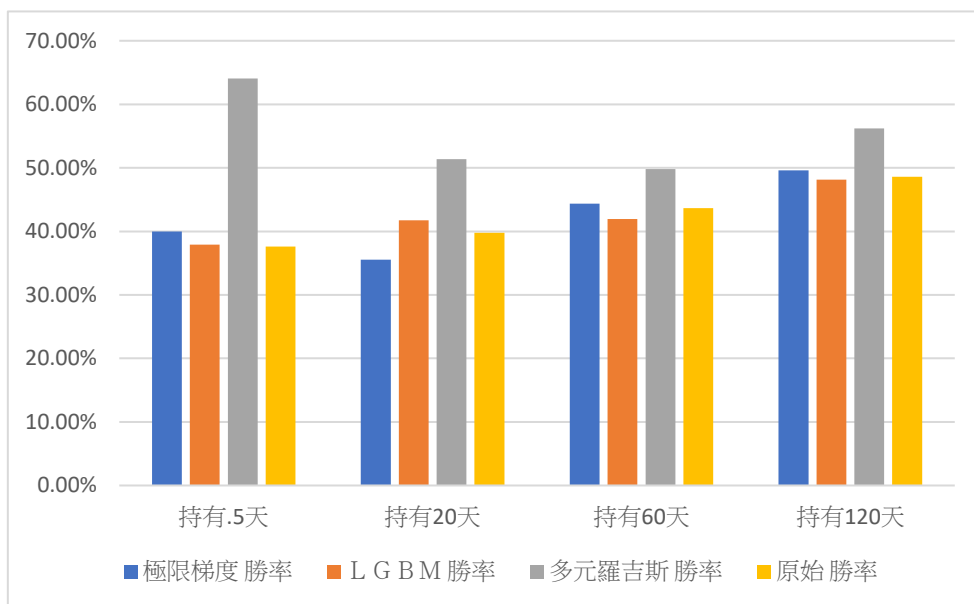


圖 四-24 閾值 0.6 三倍成交量勝率對照

資料來源：本研究整理。

4.7.4 均線糾結

本節旨在探討在均線糾結情境下，使用機器學習方法進行交易策略。研究中，我們比較了極限梯度、LGBM、多元羅吉斯迴歸和傳統方法的表現。

首先，觀察表 4-30、圖 4.25、圖 4-26 閾值 0.5 的結果，可以發現在持有期較短的情況下，多元羅吉斯迴歸模型的表現較佳，但隨著持有期的增加，極限梯度和 LGBM 在平均報酬方面表現更為突出，持有 120 天，極限梯度與 LGBM 均勝過原始報酬率和多元羅吉斯模型，但多元羅吉斯在 120 天的勝率還是最佳。在表 4-31、圖 4-27、圖 4-28 閾值 0.55 中，持有期較短時，多元羅吉斯迴歸在平均報酬和勝率上均表現出色，其他報酬均為負值，但是交易次數仍然是減少許多。

然而，在持有期較長時，多元羅吉斯平均報酬則下降許多，極限梯度模型在平均報酬上勝出，而多元羅吉斯在勝率上表現較好。最後，根據表 4-32、圖 4-29、圖 4-30 閾值 0.6 的數據，持有 5、20 天時，多元羅吉斯迴歸在平均報酬和勝率上均表現優越。然而，120 天時，極限梯度模型在平均報酬上的表現顯著優於其他模型。

可以得出以下結論：在不同閾值和持有期下，機器學習方法對於均線糾結情況的交易策略具有一定的優勢。隨著閾值的提高，模型在勝率上的表現差異逐漸擴大，多元羅吉斯迴歸在短期持有期內具有較好的表現。相比之下，極限梯度和 LGBM 在持有期較長時的平均報酬上表現更為突出。因此，在選擇交易策略時，投資者應該根據不同的閾值和持有期，選擇適當的機器學習模型以獲得最佳的投資組合。

表 四-28 均線糾結時間窗格法資料

	訓練集期間	訓練集筆數	測試集期間	測試集筆數
樣本一	20070101-20111231	959	2012	106
樣本二	20080101-20121231	913	2013	508
樣本三	20090101-20131231	1313	2014	572
樣本四	20100101-20141231	1790	2015	389
樣本五	20110101-20151231	1901	2016	246
樣本六	20120101-20161231	1821	2017	512
樣本七	20130101-20171231	2227	2018	493
樣本八	20140101-20181231	2212	2019	314
樣本九	20150101-20191231	1954	2020	375
樣本十	20160101-20201231	1940	2021	473
樣本十一	20170101-20211231	2167	2022	391

資料來源：本研究整理

表 四-29 均線糾結閾值 0.5

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	0.18%	45.59%	728	-0.50%	37.27%	1006	1.01%	50.47%	96	-0.67%	35.69%
持有 20 天	0.43%	43.86%	965	-0.31%	39.74%	1140	1.33%	46.87%	266	-0.08%	38.11%
持有 60 天	2.77%	44.37%	1152	1.84%	41.73%	1374	4.31%	45.23%	516	1.88%	41.80%
持有 120 天	6.23%	49.87%	1959	6.08%	47.74%	2072	5.76%	50.52%	1993	5.84%	47.59%

資料來源：本研究整理。

表 四-30 均線糾結閾值 0.55

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	-0.56%	37.48%	418	-0.64%	36.29%	758	3.30%	65.45%	35	-0.67%	35.69%
持有 20 天	-0.09%	39.48%	490	-0.23%	40.47%	874	4.71%	59.05%	62	-0.08%	38.11%
持有 60 天	3.95%	43.99%	759	1.23%	41.40%	1078	0.24%	52.74%	123	1.88%	41.80%
持有 120 天	6.36%	48.49%	1180	6.24%	47.81%	1749	5.31%	55.35%	954	5.84%	47.59%

資料來源：本研究整理。

表 四-31 均線糾結閥值 0.6

	極限梯度			L G B M			多元羅吉斯			原始	
	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率	交易次數	平均報酬	勝率
持有 5 天	0.49%	46.33%	228	-0.62%	36.62%	537	3.59%	76.93%	22	-0.67%	35.69%
持有 20 天	0.44%	41.70%	352	-0.38%	41.09%	634	6.41%	62.04%	17	-0.08%	38.11%
持有 60 天	0.67%	40.67%	345	1.09%	43.54%	845	-1.59%	36.36%	38	1.88%	41.80%
持有 120 天	9.59%	52.17%	588	5.44%	48.20%	1456	5.60%	51.84%	376	5.84%	47.59%

資料來源：本研究整理。

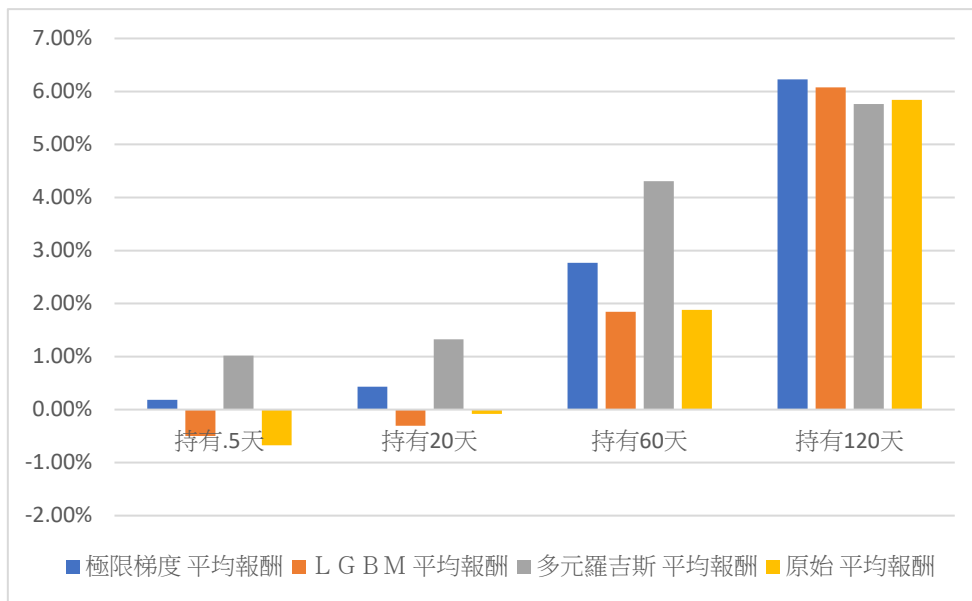


圖 四-25 閾值 0.5 均線糾結報酬率對照

資料來源：本研究整理。

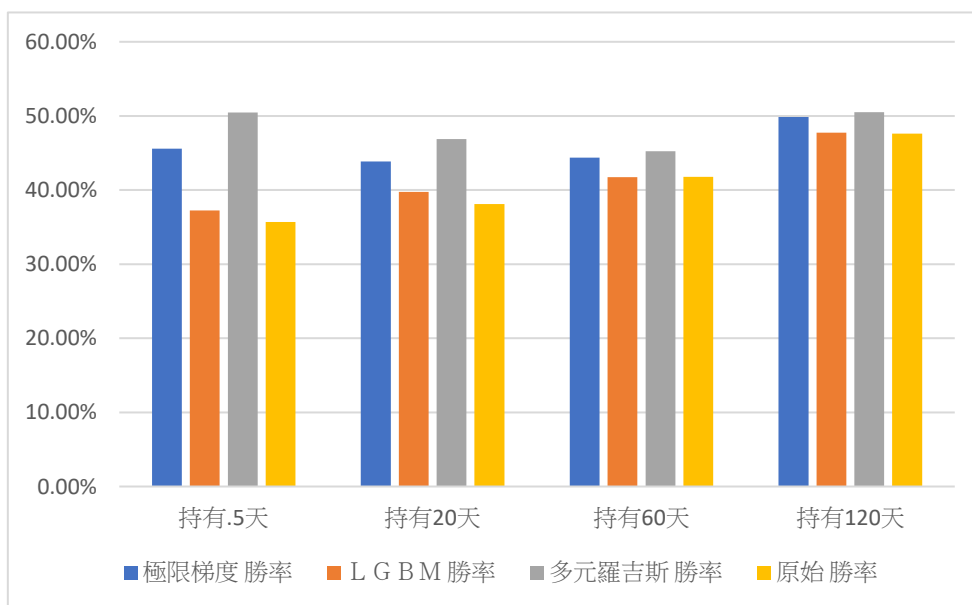


圖 四-26 閾值 0.5 均線糾結勝率對照

資料來源：本研究整理。

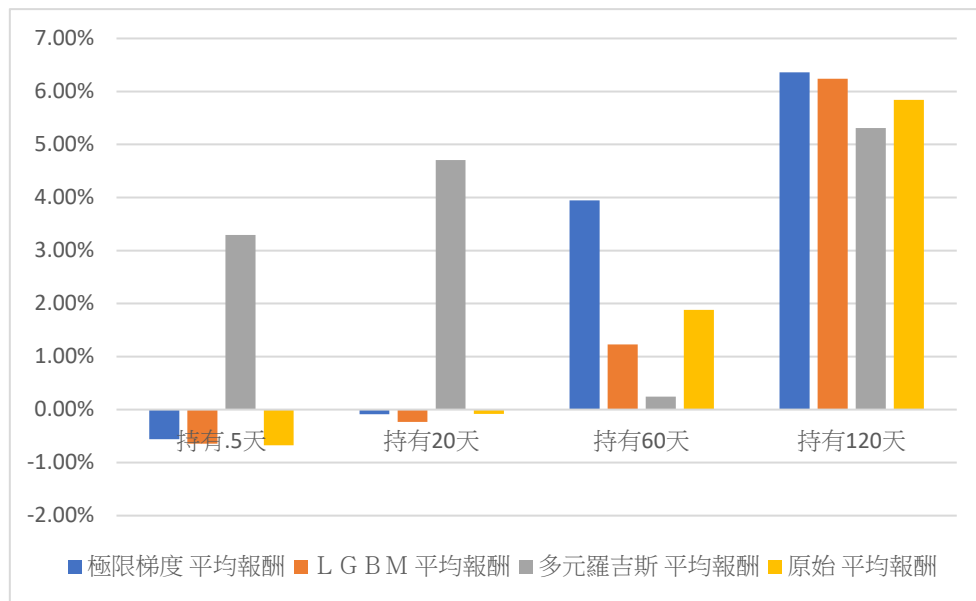


圖 四-27 閾值 0.55 均線糾結報酬率對照

資料來源：本研究整理。

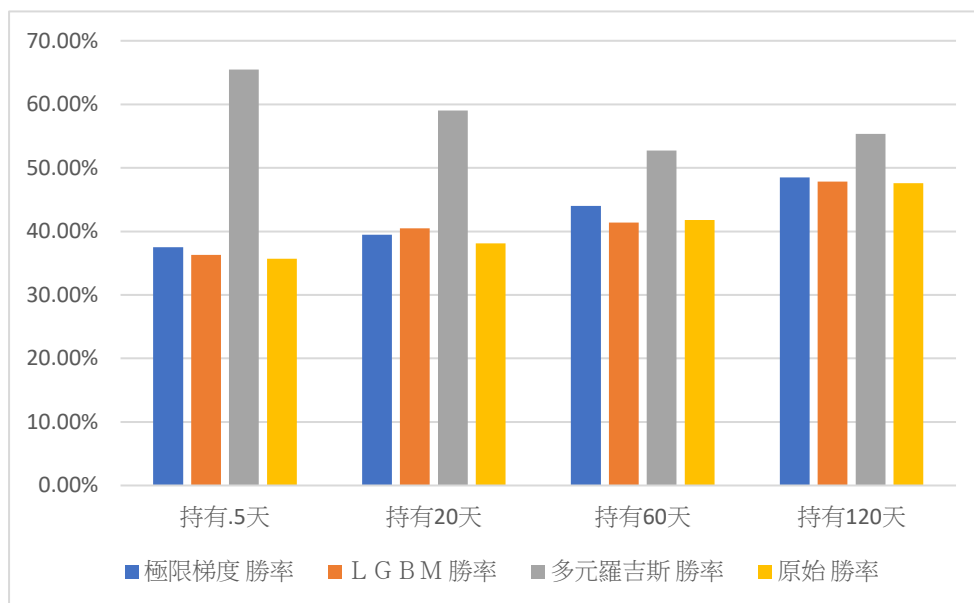


圖 四-28 閾值 0.55 均線糾結勝率對照

資料來源：本研究整理。

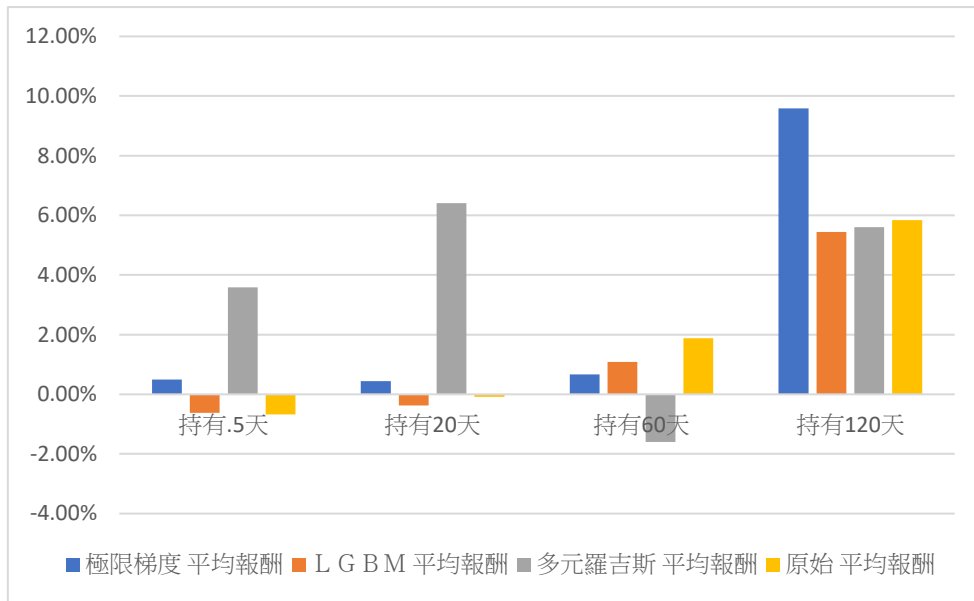


圖 四-29 閾值 0.6 均線糾結報酬率對照

資料來源：本研究整理。

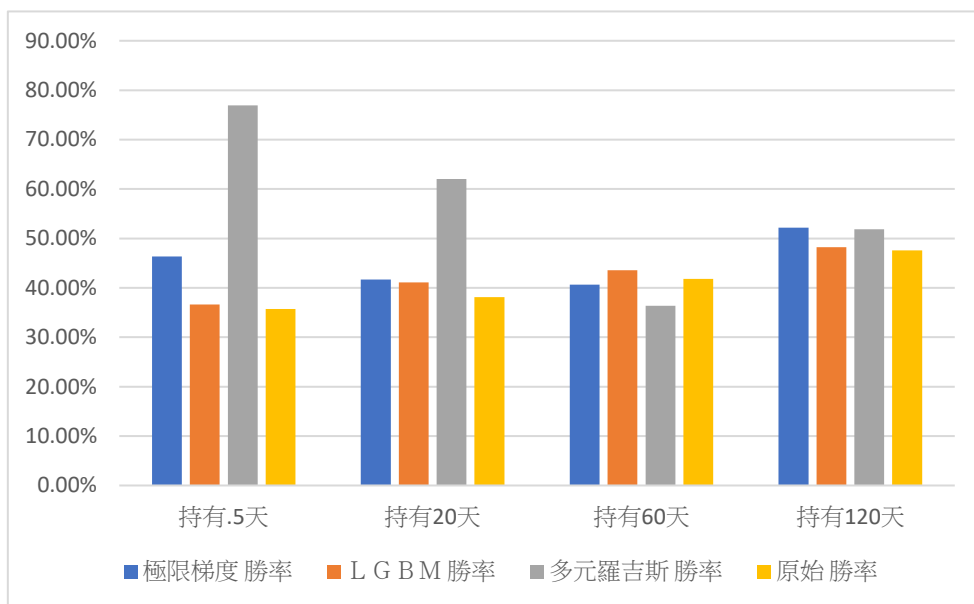


圖 四-30 閾值 0.6 均線糾結勝率對照

資料來源：本研究整理。

4.8 小結

本章節對機器學習在股票價格突破方面進行了預測，重點分析了其在不同天數的短、中、長均線策略中的表現。綜合研究結果，多元羅吉斯模型在短天數的均線策略中均取得了顯著的勝率提升和報酬率增加。此外，隨著天數的增加，整體模型報酬率和勝率呈現出上升趨勢，極限梯度模型特別在 60 天、120 天的表現更為突出，顯示出機器學習在短、中、長天數均線策略中具有顯著的潛力。本研究也發現隨著天數的增加，在機器學習多元羅吉斯模型的報酬率呈現下降趨勢，這可能是由於在更長的時間範圍內，市場波動和變化更為明顯，對機器學習模型的預測能力提出了更高的挑戰。

因此，在未來的研究中，我們可以探討如何改進機器學習模型以應對更大的市場波動和變化，進一步提高其在技術分析領域的預測能力。本研究僅為初步探討，未來可以對不同的金融市場、交易項目進行更深入的研究，以了解機器學習方法在更廣泛情況下的適用性。在選擇交易策略時，機器學習方法具有一定的潛力，但投資應根據具體情況靈活運用。進一步的研究可以拓寬機器學習方法在金融市場的應用範疇，為投資者提供更多有價值的信息和指導。綜上所述，機器學習技術在股票投資領域具有很大的應用潛力，尤其是在短、中、長天數均線策略中。本研究對機器學習模型在不同天數的均線策略中的表現進行了分析，希望為未來研究提供了有益的參考和啟示。

第五章結論

5.1 研究結論

本研究探討技術面價格突破後的買進策略，純粹的在價格突破後買進，經過回測後，並不能穩定地擊敗同期持有大盤的報酬率，勝率約落在 40%~48%，特別是追突破的短期報酬率都呈現負值居多，標準差也是比較大幅的波動，顯示股價在短期突破後，上下震盪也很大。反而是突破後持有中長期的報酬較佳，這也顯示股票市場長期向上，波段持有可獲得的報酬率也較好，但中長線的持有報酬最大跌幅也來到-82%，這也表示價格突破買進後，若未有停損的策略，則會有巨額的虧損，在有設定停損的情況下，則是可以有效控制損失。

在使用機器學習後，三個模型在短中長期，均有其表現的地方，比原先未使用時，勝率及報酬率都上升不少，以短線的預測，多元羅吉斯模型所得到的報酬率與勝率最佳，中長期則是 XGBOOST 的模型表現最好。多元羅吉斯雖然勝率有達到 6 成甚至 7 成，但是次數是少很多，本研究也發現隨著天數的增加，在機器學習多元羅吉斯模型的報酬率呈現下降趨勢，這可能是由於在更長的時間範圍內，市場波動和變化更為明顯，對機器學習模型的預測能力提出了更高的挑戰。

5.2 研究限制與建議

本研究以價格突破的條件去做策略選用，雖然有抓到大波動漲幅的個股，但損失也是占大多數，這是因為有可能價格突破時，股價已位處於高檔，股市有句話，拉高是為了出貨，因此少了檢驗股價位階的條件，則後續持有的報酬就帶來較大的負報酬，而且機器學習的模型也尚處於黑盒子狀態，每一次的篩選結果都不盡相同。對於此研究因此有幾個建議：

1. 納入判定股價當前的位階高低
2. 加入財務基本面的條件，如 EPS、營收
3. 若可以在加入型態的判讀，更可以提高策略之勝率

文獻參考

1. 中文文獻

1. 于慎為(2011)，價值型投資考量法人籌碼動能下的選股策略：以台灣股市為例，淡江大學財務金融學系碩士在職專班，碩士論文。
2. 王月玲(2003)，外資對台灣股市的影響，國立政治大學金融研究所，碩士論文。
3. 王震頡.(2020)，台股技術面交易法則之績效分析，國立高雄科技大學金融資訊系，碩士論文。
4. 林軒白(2016)，外資與投信之買超行為及資訊價值，國立臺灣大學財務金融學研究所，碩士論文。
5. 林睦融(2012)，K 線理論應用於台股市場之績效分析，東吳大學資訊管理學系，碩士論文。
6. 徐中琦、林皇瑞(2008)法人機構從眾行為之研究—台灣股市外資與自營商之比較，商管科技季刊，頁 251 - 276.
7. 張瓊如(2009)，KD 和 MA 技術指標在選擇權投資策略時機的應用—以台指選擇權為例，逢甲大學財務金融學所，碩士論文。
8. 許溪南、郭玫秀、鄭乃誠(2005)，投資人情緒與股價報酬波動之互動關係：台灣股市之實證，台灣金融財務季刊 6 卷 3 期，頁 107-121。
9. 陳淑玲、吳安琪、費業勳(2011)，臺灣股票市場技術指標之研究—不同頻率資料績效比較，東海管理評論 12 卷 1S 期，頁 187-225.
10. 游英裕(2004)，股價與成交量因果關係之研究—台灣股市的實證，義守大學管理科學研究所，碩士論文。
11. 黃光廷(2002)，技術分析、基本分析與投資組合避險績效之研究，國立成功大學會計學系碩博士班，碩士論文。
12. 楊淑惠(2007)，外資連續買賣超行為對股價報酬率之影響，國立屏東科技大學財務金融研究所，碩士論文。
13. 劉百耀(2020)，臺灣股市籌碼資料的資訊內涵—以籌碼指標建構投資組合之實證研究，國立政治大學財務管理學系，碩士論文。
14. 樓禎祺、何培基(2003)，股價移動平均線之理論與實證—以台灣股市模擬投資操作為例，育達研究叢刊 5&6 期，頁 27-51。
15. 蘇鵬翎(2000)，散戶與法人投資行為對股票報酬率與週轉率影響之探討—以台灣上市電子股為例，國立臺灣大學財務金融學研究所，碩士論文。

2. 英文文獻

1. FamaF.Eugene. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance, P383-417.
2. Brad M. Barber,Terrance Odean(2001) Boys will be Boys: Gender, Overconfidence, and Common Stock Investment. The Quarterly Journal of Economics, P261–292,.
3. Mark Minervini. (2015). Trade like a Stock Market Wizard.