

國立臺北大學財政學系
碩士論文

指導教授：劉彩卿 博士

汪志勇 博士

以深度學習建構台灣股市
當沖交易策略之實證研究

Intraday Trading Strategies Based on Deep Learning
-An Empirical Study of Taiwan Stock Market

研究生：何宇軒 撰

中 華 民 國 一 一 〇 年 六 月

國立臺北大學

財政學系(所)碩(博)士班

109 學年度第 二 學期畢業生論文

研究生：何宇軒 撰

業經本委員會審議通過

題目：以深度學習建構台灣股市當沖交易策略之實證研究

論文考試委員：(受疫情影響，論文口試採視訊方式，得由各委員分別簽章)

召集人

委員

委員

委員

指導教授

系(所)主任

論文口試及格日期：

中華民國 一 一 〇 年 六 月 十 七 日

謝誌

回想一開始選擇攻讀研究所的原因，雖說不是完全有達到當初預定的目標，但研究所這兩年也是盡心盡力的在努力著，花了大把課外時間實習、線上自學和參加培訓計畫等，讓自己確實累積了點東西，離目標更加接近。讀研究所是第一次離開台南，初次北漂到三峽，享受到自由的空氣，也被環境逼著獨立，在這一段沒意外是最後的學生時期，過得非常充實和自由，感謝這兩年遇到的所有人，感謝那些陪我一起聊天、玩樂和吃宵夜的同學們。

首先要感謝的是兩位幫助我許多的指導教授，劉彩卿老師和汪志勇老師。還記得第一次與彩卿老師在導生聚上的談話，便讓我對老師印象深刻，之後也有幸成為老師的指導學生，讓我選擇我自己有興趣的論文題目做研究，而彩卿老師除了論文上的指導，也很關心學生的職涯發展和生活狀況，多虧彩卿老師讓我能有機會更加認識資料分析在金融業的應用，也時常給予我鼓勵，讓我對自己保有信心。很感謝每次志勇老師都大老遠地跑來臺北大學和我討論論文進度，而老師在金融科技的经验也使我的視野更加開闊，論文題目的資料來源也是多虧志勇老師的合作計畫方能取得，除了論文方向的指導，私底下志勇老師也待我們不薄，除了有生日餐，還有籃球賽能看，學生何其有幸能擔任兩位老師的指導學生。

再來，感謝我的家人們，感謝奶奶每次的噓寒問暖，都帶給我很大的力量，雖然我沒有如你所願以公務員為目標，但孫子不會讓你擔心的。也感謝叔叔適時關心我職涯狀況，在有需要時給予我經驗上的建議。感謝爸爸支持我做的任何決定，給了我極大的自由做我想做的事。最後，感謝這兩年的自己，能夠持續保有熱情和積極地衝出舒適圈，希望未來我也能保持彈性的做自己。

何宇軒 謹誌於

臺北大學財政學系碩士班

中華民國一一〇年六月

國立臺北大學 109 學年度第二學期 碩士 學位論文提要

論文題目：以深度學習建構台灣股市當沖交易策略之實證研究 頁數：52

所 組 別：財政學 系(所) 實證 組 (學號：710872101)

研 究 生：何宇軒 指導教授：劉彩卿 博士
汪志勇 博士

論文提要內容：

金融科技 (Fintech) 主打以科技增進使用者體驗，透過科技增進金融服務效率。而台股在近幾年也因為「逐筆交易制」和「零股交易制」的實施，促使股票市場更有效率，也吸引眾多的年輕股民進入市場。本研究使用深度學習演算法的長短期記憶模型，學習技術指標背後的訊號，預測日內股票跳動趨勢，並搭配動能策略以建構當沖交易模式，再透過回測方式與大盤之報酬做比較。

本研究發現透過集成學習的堆疊法有助於交易訊號的預測效能，資料來源，為透過 Open API 取得 2021 年 1 月 18 日至 2021 年 5 月 13 日的即時日內資料進行分析。若加上動能策略，有最佳之交易績效表現。透過此一交易模式方法，機器人理財得到超額報酬的機率大幅提升。本研究主要貢獻有二點：第一，過往文獻大都注重在提升模型的衡量指標和不同演算法模型衡量指標的比較，本研究利用交易回測的實證結果，驗證堆疊法搭配動能策略能大過大盤報酬；第二，過去研究大都針對日資料做探討，少有當沖相關的實證研究。本研究透過 Open API 搜集日內資料，研究當沖的績效評估，為目前國內屈指可數的重要文獻。

關鍵字：金融科技、機器人理財、Open API、長短期記憶模型、當沖交易

ABSTRACT

Intraday Trading Strategies Based on Deep Learning

-An Empirical Study of Taiwan Stock Market

by

Yu Xuan, He

June 2021

ADVISORS: Dr. Liu, Tsai-Ching and Dr. Wang, Chih-Yung

DEPARTMENT: Department of Public Finance

MAJOR: Public Finance

DEGREE: Master of Arts

Fintech with AI technology creates a better performance of financial service, and dramatically increases the value of financial instrument that people invested. Whether the implementation of “Continuous Trading” and “Odd Lot Trading” improves the Taiwan Stock Market efficiently or not is an interesting issue. Thus, this paper focuses on the examination of this question using a Long Short-Term Model (LSTM) to catch trading signals from technical indicators. We construct a method of day trading with the prediction of intraday jumps and momentum strategies on back-testing. Finally, a comparison of the return level of TAIEX among the different methods will be included.

We find that the stacking method from ensemble learning increases the performance of trading signals prediction. The data used is through Open API, with intraday data from Jan, 18 to May, 13, 2021. The results find that the combination of prediction and momentum strategies has the best outcome in a cumulative return. The probability of excess return earned from a robo-advisor could increase substantially. In sum, the main contributions of this paper are as follows: First, the previous papers only aim to evaluate the different independent models on return, while this study verifies that the combination of stacking and momentum strategies has a better return than TAIEX. Second, this paper is, so far, the first empirical study using intraday data through Open API.

Key word : Fintech, Robo-Advisor, Open API, LSTM, Day Trading

章節目錄

第一章 緒論.....	1
第一節 研究背景.....	1
第二節 研究目的.....	7
第二章 文獻回顧.....	8
第一節 開放銀行和 Open API.....	8
第二節 機器人理財.....	11
第三節 台灣股市.....	15
第三章 研究方法.....	19
第一節 資料來源、處理與特徵選取.....	20
第二節 資料不平衡處理.....	28
第三節 長短期記憶模型.....	30
第四節 交叉驗證和集成學習.....	32
第五節 模型衡量指標.....	34
第四章 實證結果.....	36
第一節 模型訓練表現.....	36
第二節 交易回測.....	38
第三節 實證結論.....	45
第五章 結論.....	47
第一節 研究貢獻.....	47
第二節 研究建議.....	48
參考文獻.....	49

圖目錄

圖 1：全球投資活動總額和成交數（改自 KPMG, 2020）	2
圖 2：全球理財機器人公司管理資產總額（改自 Statista, 2020）	4
圖 3：理財機器人財富管理服務全球使用人數（改自 Statista, 2020）	4
圖 4：台灣股市每月平均當日沖銷交易統計趨勢圖.....	6
圖 5：銀行平台生態系中的各種開放性（Zachariadis & Ozcan, 2017）	9
圖 6：開放銀行架構（Premchand & Choudhry, 2018）	9
圖 7：研究方法流程架構圖.....	19
圖 8：主成分的累積貢獻比率.....	27
圖 9：SMOTE 示意圖	29
圖 10：LSTM 內部神經元（Neurons）運作模式（Oinkina, 2015）	30
圖 11：交叉驗證架構.....	32
圖 12：五大模型訓練集混淆矩陣.....	37
圖 13：交易回測虛擬碼.....	40
圖 14：類別一累積報酬比較圖.....	41
圖 15：類別二累積報酬比較圖.....	41
圖 16：類別三累積報酬比較圖.....	42
圖 17：類別四累積報酬比較圖.....	42
圖 18：表現最佳三大交易策略累積報酬圖.....	43
圖 19：表現最佳三大交易策略日報酬圖.....	43

表目錄

表 1：技術指標的計算公式.....	23
表 2：常見技術指標門檻定義原則.....	25
表 3：各模型類別比例.....	28
表 4：二元分類混淆矩陣.....	34
表 5：三分類混淆矩陣.....	35
表 6：五大模型訓練集衡量指標.....	37
表 7：各交易策略使用模型整理.....	46



第一章 緒論

第一節 研究背景

金融科技（Fintech）主打以科技增進使用者體驗，透過科技增進金融服務效率。由於科技日新月異的進步，人工智慧、區塊鏈、雲端運算等新興技術的出現，在金融服務領域產生許多應用，像是：電子支付、P2P 放貸、數位銀行到機器人理財顧問（Robo-Advisor），這些應用都在漸漸地改變我們對於金融服務的認知，甚至改變我們的生活模式。

圖 1 表示 2020 上半年全球金融科技的投資總額為 256 億美元，而全球創投對金融科技的投資金額達 200 億美元，雖然受到 COVID-19 影響，相比 2019 年有稍微下降，但整體依然保持著相當強勁的狀態（KPMG, 2020）。數位銀行市場規模在 2019 年約 8 兆美元，預期在 2026 年超過 12 兆美元（Preeti Wadhwani, Sachin Kasnale, 2020），金融科技在國際市場上正在快速發展。人工智慧、區塊鏈、數據平台、網路安全的監管科技等數位工具和策略夥伴關係將在數位轉型金融生態圈的銀行市場中成為必須存在的優勢關鍵（Khanna & Martins, 2018）。

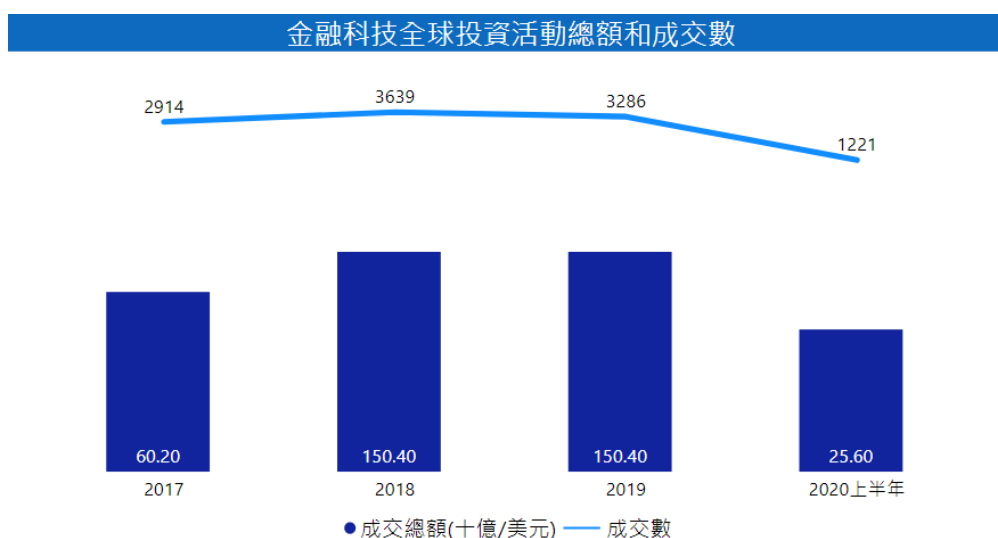


圖 1：全球投資活動總額和成交數（改自 KPMG, 2020）

在 2017 年，台灣的金融服務滲透率為 94%，世界排名 23（World Bank Group, 2018），因此金融科技在台灣帶來的影響是非常巨大的。安侯企業管理股份有限公司也在 2020 年 7 月發佈台灣金融科技發展 100 個關鍵問題，就身分驗證、支付、資本市場與財富管理等 9 大面向做探討。在 2020 年 8 月 27 日時，台灣的金融監督管理委員會發佈 90 頁的金融科技發展路徑圖白皮書，提及到四大目標：普惠、創新、韌性和永續，而這四大目標會透過八大方針輔助完成，分別是：單一窗口溝通平台、能力建構、法規調適、數位基礎建設、資源共享、監理科技、園區生態系、國際鏈結，替台灣未來三年的金融科技發展訂下具體目標。金管會引用 2020 年金融穩定學院（Financial Stability Institute, FSI）提出的金融科技樹（fintech tree）架構，結合台灣金融科技現況，將金融科技環境分為三類：樹冠－金融科技活動（fintech activities）、樹幹－賦能技術（enabling technologies）和樹根－政策輔佐（policy enablers）。依照這三類可分別對其訂定監理政策，其中機器人理財被列為重點監理金融科技活動之一。

機器人理財或稱智能理財，定義為：提供自動化、演算法驅動金融規劃服務之數位平台，僅需要很少，甚至無須真人監督。典型的機器人理財透過線上

調查，從客戶端收集他們個人財務狀況與未來目標等資訊，並使用這些資料進行分析後，提供客戶建議，且/或自動地投資客戶資產（財團法人中華民國證券暨期貨市場發展基金會, 2018）。全球的機器人理財公司管理的資產總額，將在 2020 年底達到 1.44 兆美元，且規模穩定成長（圖 2），至 2022 年全球將有 1 億用戶使用理財機器人自動化管理資產與投資，每人平均約投入 20,000 美元（圖 3）（Statista, 2020）。

台灣的機器人理財大致分成三類。第一種：是以基金投資為主的理財服務，提供線上投資理財服務；第二種：是以專門針對 ETF（指數型證券投資信託基金）推出的投資平台；第三種：是提供銀行理專挑選基金的量化系統（張明珠, 2019）。就本研究所知，截至 2020 年共有 7 家銀行（王道、中國信託、國泰世華、兆豐、台北富邦、台新、永豐）和 6 家投信投顧（富蘭克林、復華、野村、群益、中租、鉅亨）在市場上提供機器人理財服務，國內的機器人理財主要是透過大數據分析，給予投資人投資建議，但真正需要下單的還是投資人自己，屬於還在半自動化的階段，相較國外只需投資人設定好目標，系統會自動分析和下單，屬於全自動化的機器人理財服務，相比之下，台灣在機器人理財領域上仍處於早期階段，而自金管會在 2017 年開放機器人理財業務截至 2020 年 1 月以來，機器人理財管理的資產規模，僅有新台幣 7.1 億元。

機器人理財背後所使用的技術，主要為機器學習中的深度學習演算法，機器學習之定義為：「從過往的資料和經驗中學習並找到其運行規則，最後達到人工智慧的方法。」而深度學習則可視為實現機器學習的方法之一，常見的深度神經網絡便是深度學習演算法的一種。由於大數據時代來臨，讓深度學習演算法在各領域都有許多應用出現，而在機器人理財之應用大多可分為兩大部分，第一，用於自動預測資產組合配置權重調整，第二，用於預測市場走勢、股價等。

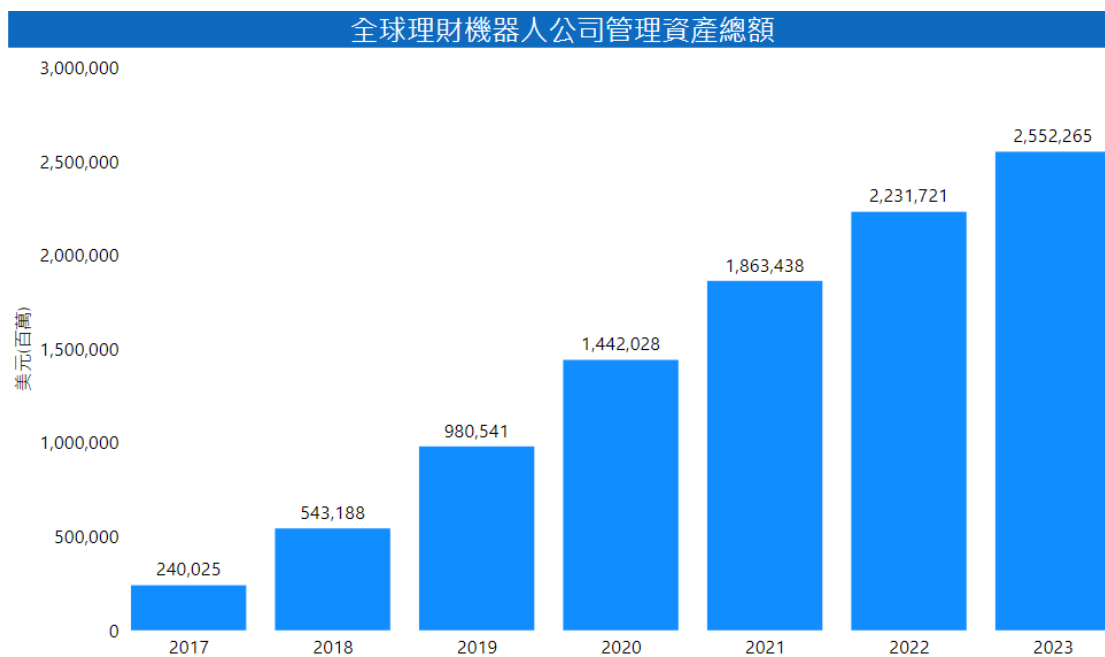


圖 2：全球理財機器人公司管理資產總額（改自 Statista, 2020）

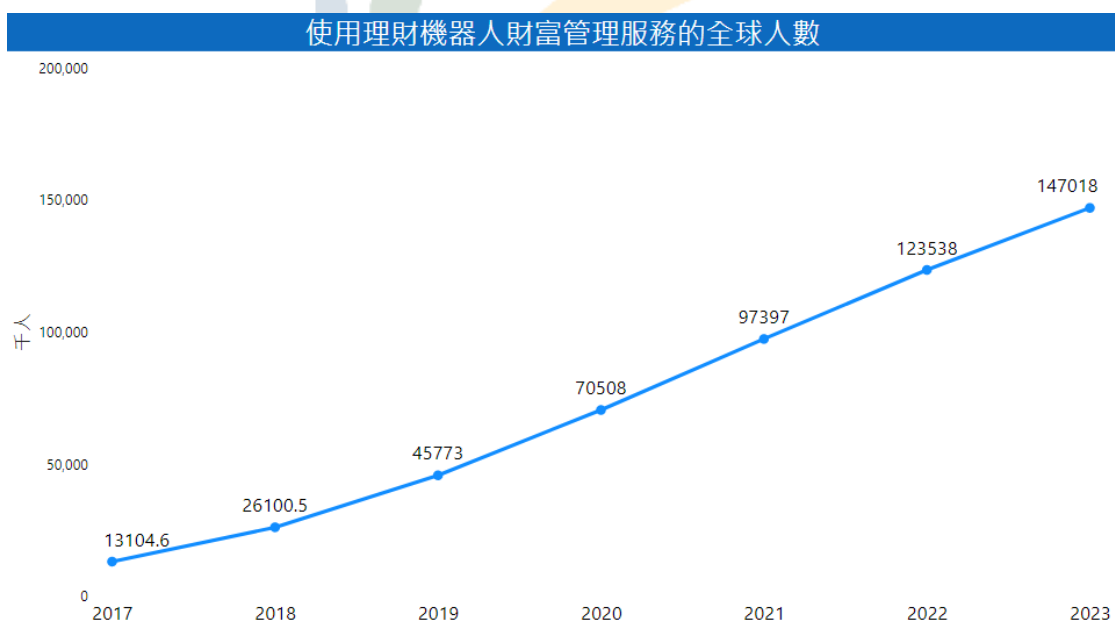


圖 3：理財機器人財富管理服務全球使用人數（改自 Statista, 2020）

除了機器人理財被列為重點監理金融科技活動外，開放應用程式介面（Open Application Programming Interface, 以下稱 Open API），也是支持台灣推行金融科技應用非常重要的基礎技術，API 是一種讓軟體之間互動的方式，Open API 則是提供一種基於公共標準規則的 API，目前主要透過財金資訊股份有限公司建置 Open API 平台，Open API 的推行讓台灣金融機構業者在自家伺服器端或是資料庫端所儲存之資料，能夠在符合公開標準和規範下，讓第三方服務提供者（Third-Party Service Providers, 以下稱 TSP）做使用，並透過 Open API 發展各種創新應用，以此推動台灣金融科技發展。

在 2020 年 3 月 23 日，台股開始實行逐筆交易制，不只跟國際接軌，也能提高交易效率、增加資訊透明度；在 2020 年 10 月 26 日，開放盤中零股交易，用銅板也能夠買股票，降低進入股市的門檻，吸引眾多年輕人投入股市。隨著股市中交易效率越來越高，散戶的進入門檻開始降低，大部分散戶會因為即時的大盤波動、新聞事件和財經老師們的建議等，影響到個人的投資判斷，也因此增加人為偏誤發生的機率，因此機器人理財顧問對於散戶的重要性也越來越高，金融科技的革新讓傳統高資產客戶才能享有的財富管理服務，也能透過機器人理財幫助小資族群和年輕人做好財富規劃。

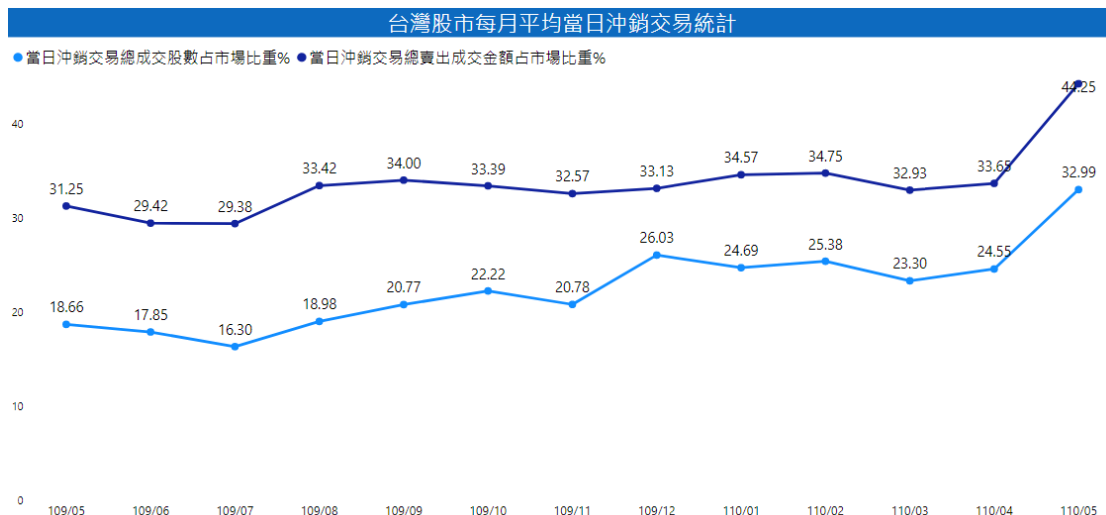


圖 4：台灣股市每月平均當日沖銷交易統計趨勢圖

圖 4 是台灣股市每月平均當日沖銷交易統計趨勢圖，可以發現台股的當沖交易成交股數和成交金額占比都在呈現上升趨勢，可能原因為疫情回溫股市呈現多頭趨勢，加上散戶增加與交易效率的提高所致，因此除了透過機器人理財分散人為的偏誤風險，希望未來也有機會讓機器人理財進行當沖交易，透過機器人理財降低因為當沖高風險帶來的門檻，也讓機器人理財能夠有不同面向的操作空間，故本研究想透過深度學習模型預測股價趨勢，並透過模擬實際交易狀況做回測，在減少人為偏誤下，讓機器人理財能有不同的交易方式獲取超額報酬。

第二節 研究目的

本研究希望設計一套架構讓金融科技樹中的政策（開放銀行政策），賦能技術（API、機器學習和人工智能），金融科技活動（機器人理財）做結合。透過金融科技創新園區提供的 Open API 串接台灣證券交易所的上市即時日內資料，結合深度學習中的長短期記憶模型（Long Short-Term Memory, 以下稱 LSTM）預估盤中股票的漲跌趨勢，讓機器人理財做到全自動化日內交易，以此流程實現開放銀行中透過 Open API 開發第三方服務，為之後機器人理財應用給予參考架構。透過本研究，希望能搭配模型預測結果和常見技術指標門檻定義原則建構交易策略，透過模型的判斷和程式化自動交易減少人為主觀因素造成的偏誤，提供能讓股市風險藉由模型來減緩的可能，也讓一般大眾能夠以較低門檻的方式透過機器人理財接觸當沖交易。因此本研究想了解以下四點：第一，現行第三方服務提供者透過 Open API 發展機器人理財應用的流程，Open API 在技術上或是應用上是否有所限制；第二，深度學習演算法是否能良好的從日內資料中找到當沖交易訊號，並從台灣股市獲取超額報酬；第三，除了深度學習演算法外，再搭配常見技術指標門檻定義原則之交易策略是否有助於交易上有更佳的績效表現；第四，使用集成學習堆疊法而建構之交易策略在回測績效上是否優於單一演算法預測結果所建構之交易策略。

第二章 文獻回顧

本章節第一節先就開放銀行和 Open API 目前在國際上和台灣的應用和機會做探討，並從中比較差異並尋找台灣能借鏡之處。第二節會針對機器人理財相關文獻的整理，包括機器人理財背後資產組合配置之理論和方向以及所運用之機器學習方法。最後一節會就台灣股市在交易策略、異常報酬和股市預測相關文獻做整理。

第一節 開放銀行和 Open API

巴塞爾委員會給開放銀行的定義：指銀行在客戶允許的情況下將客戶數據共享給第三方機構，以開發更有效率和透明的應用和服務。而開放應用程式介面則是實施開放銀行非常重要的基礎技術，應用程式介面是將讓軟體之間能互相交流的規則，以不同程式之間的應用介面呈現，促進軟體之間的互動；開放應用程式介面為提供一種基於公共標準規則的應用程式介面（BIS, 2019）。

Zachariadis & Ozcan（2017）整理了現存文獻中提到關於應用程式介面的應用，並且提出平台商業模式背後重要的理論依據：降低交易成本和網路外部性，將有助於銀行即平台（Banking As A Platform, BaaP）此類新興商業模式發展，並描述銀行即平台生態系下的各種開放程度（圖 5）。也提及在英國成立的開放銀行工作小組（the Open Banking Working Group, OBWG）和歐盟的第二號支付服務指令（The second Payment Services Directive, PSD2）的監管措施下，開放應用程式介面未來發展的機會和挑戰。

	Open Source API Platforms (e.g. Open Bank Project)	Banking API Platforms (e.g. Token, Fidor, etc.)	Challenger Banks (e.g. Monzo, Starling, No26, etc.)	Most Incumbent Banks
Demand-Side User (End Customer)	Open	Open	Open	Open
Supply-Side User (FinTechs)	Open	Open	Open	Closed
Platform Provider (Hardware/OS Bundle)	Open	Open	Closed	Closed
Platform Sponsor (Design & IP Rights Owner)	Open	Closed	Closed	Closed

圖 5：銀行平台生態系中的各種開放性 (Zachariadis & Ozcan, 2017)

Premchand & Choudhry (2018) 研究應用程式介面對於開放銀行的必要性，整理出開放銀行該有的架構 (圖 6)，並提出開放銀行透過開放應用程式介面進行價值創造的方式：授權第三方團體使用現存系統以創造新的應用、共同提供產品和服務、增強現有服務和產品、改善顧客互動、新的收益通路；也指出開放銀行的關鍵挑戰包含：贏得顧客信任、教育顧客了解開放金融資訊給第三方團體的好處、創造創新的產品和服務以便吸引那些尚未習慣數位化的顧客、防範因應用程式介面而出現的新型態網路攻擊。也整理出開放應用程式介面標準化 5 項現有的準則。

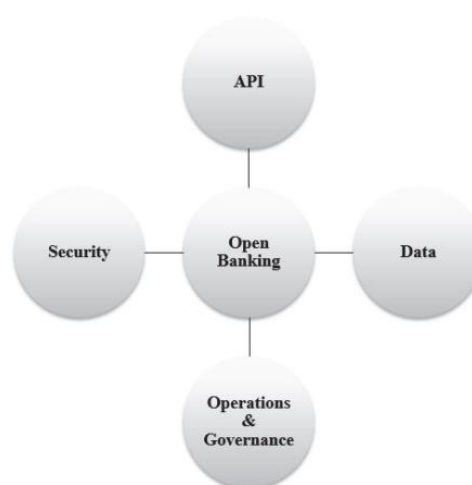


圖 6：開放銀行架構 (Premchand & Choudhry, 2018)

就台灣現況而言，金管會鼓勵金融機構基於營運策略與業務需求，以自願自律方式推動開放銀行服務，不同於英國和歐盟的強制開放方式，規劃採漸進式分階段推動，並在金融科技應用上由金融機構與 TSP 業者合作之方式推動，第一階段為「公開資料查詢」，如利率、匯率、ATM 位置、分行資訊、產品資訊等。TSP 業者透過 API 可串接到各銀行的商品資訊，消費者透過 App 就可輕鬆比較各銀行的相關商品服務；第二階段為「消費者資訊查詢」，在客戶同意下開放客戶資訊，TSP 業者可以提供帳戶整合服務，例如將消費者在不同銀行的房貸、存款、基金投資等資訊做整合；第三階段為「交易面資訊」，在客戶同意下開放交易與支付，TSP 可在整合帳戶後直接透過 APP 連結帳戶扣款、支付、調整或撥付帳戶資金。目前第一階段已於 2019 年 9 月開始運作，第二階段在 2020 年底開始上線（金融監督管理委員會, 2020）。台灣金融研訓院（2020）提及由於我國金融業並沒有像先進國家中存在足以壟斷市場的大型金融機構，因此我國環境相較於立法強制讓民營銀行遵守，自願自律的方便和彈性更適合我國現況。目前台灣由財金資訊公司和政治大學金融科技中心共同推動 Open API 生態系，包含成立 Open API 平台、TSP 業者的沙盒測試、定期舉辦開放銀行政策與法規之座談會等，主要充當金融機構和 TSP 業者之間溝通橋梁，滿足雙方對於 API 技術、資安和應用等需求。

第二節 機器人理財

機器人理財背後運作的依據可以拆解成：資產配置理論和機器學習演算法兩大部分。Beketov (2018) 整理出機器人理財基本的五大流程：資產範圍選擇、投資人輪廓確認、投資組合最適化、監控和調整、績效檢視和回報。而目前台灣的機器人理財大多還停留在半自動化類型，透過問卷了解客戶的投資風險偏好後，給予投資組合建議，但還是需要投資人親自下單，相比國外已經盛行的全自動化類型，台灣的機器人理財現況並不會做到自動下單和後續的投資組合再平衡。

一、資產配置理論

Markowitz (1952) 以平均值-變異數投資模型 (Mean-Variance Model) 建構投資組合，一系列的投資組合會形成一條曲線，稱為效率前緣 (Efficient Frontier)，位在效率前緣上的投資組合都能達到相同風險下，可獲得最高之預期報酬率或同預期報酬率下，相對風險最低，以此概念建構現代投資組合理論 (Modern Portfolio Theory, 以下稱 MPT)。而 MPT 也被廣泛做為機器人理財的資產配置理論依據，(陳彥安, 2019; Beketov et al., 2018; Semko, 2019; Obeidat et al., 2018) 都有針對 MPT 做相關的延伸應用研究。陳彥安 (2019) 以 MPT 為理論基礎結合機器學習演算法，依三大法人每日的買賣超籌碼變化，自動調整投資組合權重；Obeidat et al (2018) 也以 MPT 為基礎理論，搭配機器學習演算法並和傳統的投資組合管理方法：資本資產定價模型 (Capital Asset Pricing Model, 以下稱 CAPM)、三因子模型 (3 Factor Model)，以上三個經典的方法做比較；Beketov et al. (2018) 探討國際上的 219 個機器人理財顧問背後的理論依

據，並將這些機器人理財顧問公開資訊上提及的相關方法畫成文字雲。

MPT 在實務運用上有以下幾項缺點：資產配置過於集中，在做投資組合再平衡時，可能導致權重變換過大，增加不必要的交易成本（Black and Litterman, 1992）；對輸入參數過於敏感，像是對預期報酬率做小幅度的變化，模型結果卻會有大規模變化（Best and Grauer, 1991）；估計參數誤差過大（Michaud, 1989）。於是 Black et al.（1992）以 MPT 做延伸並結合 CAPM，提出 Black-Litterman model（以下稱 BL 模型）大幅修正以上缺點，並且能加入投資人主觀觀點進入模型。Semko（2019）以 BL 模型為理論基礎去結合機器學習演算法，觀察輸入參數對於中間各項處理單元的敏感度。雖然 BL 模型大幅修正 MPT 的缺點，但 Beketov et al.（2018）發現 MPT 仍然是現在機器人理財主要依循的理論，而比起開發全新的資產配置理論，市場主流還是去延伸和改進 MPT。

二、機器學習演算法

機器學習，是指透過數據和以往的經驗讓電腦能自主學習，以此優化算法找出背後可能的規律。早在 60 年代就曾被提及相關概念，但因為受限於當時的計算機能力有限且數據不夠龐大，因此沒有過多的應用出現，而近幾年由於電腦的出現加上大數據時代來臨，機器學習成了熱門領域和話題，機器學習則靠著強大的計算力著重在模型表現上，因此也被應用在許多商業情境和股票預測上。

股票預測可以拆解成預測股價的數值和預測股價的趨勢，（卓越, 2019；陳家蓁, 2020；Lai et al., 2019；Li et al., 2018；Liu et al., 2018）針對股票相關數值做預測；（施育霖, 2019；陳彥安, 2019；Ao et al., 2019；Huang et al., 2019；

Shen et al., 2020；Rosdyana Mangir Irawan Kusuma et al., 2019）則是針對股票趨勢做預測。

針對預測連續性資料，常見的機器學習演算法有：類神經網絡（Artificial Neural Network, 以下稱 ANN）、長短期記憶模型（LSTM）、支持向量迴歸（Support Vector Regression, 以下稱 SVR），陳家蓁（2020）使用共 19 年的台灣股市指數、電子類股指數和金融類股指數月資料，使用八項總體經濟因子和七個技術指標做為訓練模型的特徵，結合遺傳演算法（Genetic Algorithm）找出 ANN 各個神經元（Neurons）的最佳參數；Lai（2019）使用共 10 年的 5 家台灣上市櫃公司資料，將前五天的市場平均訊息做為訓練特徵，訓練 LSTM 模型預測後五天的市場平均值，並加入四種技術指標做為購買、出售或持有股票的依據；Li et al.（2018）將 LSTM 模型擴展成 Multi-Input LSTM，新增正相關和負相關的公司股價做為訓練特徵，強化能幫助股價預測的部分，同時減緩會傷害到模型表現的部分，並使用中國 300 支股票資料做測試，與原本的 LSTM 和遞歸神經網絡（Recurrent Neural Network, 以下稱 RNN）做比較，發現 Multi-Input LSTM 的效果比傳統的 LSTM 和大盤還要好；卓越（2019）選取 2009 年到 2018 年共 10 年的 170 個指標和收盤價資料，使用 SVR 結合主成分分析（Principal Component Analysis, 以下稱 PCA）和網格搜索法（Grid Search），進行降維和最佳參數尋找；Liu et al.（2017）對半年期黃金期貨和 S&P 500 期貨的數據，將交易量分解成兩部分來預測日內交易量百分比，特徵分別是前幾日同區間的交易量百分比殘差和同交易日的前幾區間的交易量殘差，放進 SVM 預測下一區間日內交易量殘差部分，並搭配 VWAP 交易策略做檢驗，發現整體表現是改善的。

而股票趨勢預測部分，可以將其視為分類問題，常見的機器學習演算法有：決策樹（Decision Tree）、K-近鄰演算法（K-Nearest Neighbor, 以下稱 KNN）、支持向量機（Support Vector Machine, 以下稱 SVM）、隨機森林

(Random Forest, 以下稱 RF)、貝氏分類器 (Naive Bayes Classifier, 以下稱 NBC)、ANN、LSTM，施育霖 (2019) 使用 2014 年至 2018 年共 5 年的 10 家金融公司，使用 20 個特徵做訓練，透過決策樹、SVM、KNN 預測隔天的股價趨勢是上漲還下跌；陳彥安 (2019) 使用 2008 年至 2019 年共 12 年的台灣 15 家企業做為投資組合，使用 6 個特徵做為 LSTM 訓練，進而學習以便於動態調整投資組合權重以提高投資報酬 Ao et al. (2019) 使用深圳證券交易所中 1,271 支股票 2014 年至 2017 年共 896 個交易天的 5 分鐘區間 level-2 高頻資料，輸入 130 個流動性測量特徵和 54 個技術指標特徵到 RF、SVM、ANN 和 KNN 做訓練，預測下一區間的股票趨勢，比較模型之間的表現，發現 RF 表現較佳；Huang et al. (2019) 透過 2000 到 2018 共 19 年的台灣股價指數資料，使用 10 個技術指標做兩種方式訓練模型，首先，先就連續性資料，之後則依照投資人觀點定義技術指標預測未來股價趨勢，將資料轉為類別型，並分別餵進 ANN、SVM、RF、NBC 做訓練，結果發現將資料預先處理成類別型資料有助於模型預測表現；Shen et al. (2020) 收集 2 年的中國 3,558 支股票資料，將原始的 20 個特徵使用特徵工程拓展成 54 個特徵，餵入 LSTM 做訓練，並結合 PCA 做降維，將前 9 天的資料拿來預測第 10 天的股價趨勢，並和 SVM、RF、NBC、ANN、羅吉斯迴歸做比較，發現 LSTM 表現較佳；Rosdyana Mangir Irawan Kusuma et al. (2019) 使用 2000 年至 2018 年 6 月中的台灣 50 家公司和印尼 10 家公司的資料，比較深度卷積網路 (Convolutional Neural Network, 以下稱 CNN)、殘差網路 (Residual Network, 以下稱 ResNet)、VGG Network 等深度學習模型和傳統機器學習的分類模型：RF 和 KNN 在預測股價趨勢上的表現做比較，本研究特色在於將不同區間的 K 線圖做為訓練資料餵入圖像辨識深度學習模型中，而最後結果也發現 CNN 表現最佳。

第三節 台灣股市

台灣股市，自從 1962 年開始。台灣股市一直以來也都是熱門的研究領域，以下整理關於台灣股市相關的研究文獻，主要分成三大部分：交易策略、股市現象、股市預測。由於台灣日內資料取得不易，因此大多數台灣股市相關文獻為日資料基礎所做之研究，對於日內資料基礎之當沖相關研究更為稀少，但本研究希望能借鏡現在已有之日資料基礎文獻，從中所使用的交易策略、台灣股市特有的現象和預測交易訊號所使用之特徵選取等，並由本研究應用至當沖交易上。

一、交易策略

市場上有各式各樣的交易策略被人們所尊崇，交易策略是一種準則，可能是依照技術指標出現買進賣出訊號，也可能是分散風險的方式等，常見的有像是配對交易，挑選兩個相關性較高的股票，一旦兩者之間出現了背離的走勢，且這種背離在未來是會得到修正的，那麼就可能產生套利的機會；動能策略，指買入那些過去某段時間報酬率相對強勢的股票，因為相信會保有動能持續上漲；反向策略，買入那些過去某段時間報酬率相對低落的股票，因為相信指是被低估，遲早會發生反彈。童韋傑（2020）將配對交易策略中的最小距離法透過機器學習去做拓展，以發展出延後交易策略；洪振虔（2011）探討結合過去交易量與報酬率如何預測價格反向現象的程度與持續性，分別建立強反向策略與弱反向策略，而發現過去交易量與報酬的關係，無論在價格具有動能或反向現象，都對於未來報酬率具有預測能力。

二、異常報酬

不同地區的股市會由於當地的文化差異、投資人組成方式而有些股市現象存在。林秀柔（2020）使用 2008 年至 2019 年共 12 年，1,618 間公司的歷史股票資料，對於七種常見的市場異象做異常報酬檢驗，包含：淨營運資產、毛利率、O-Score、帳面市值比、資產投資、價值低估減去價值高估、動能，並納入台灣股市中的三大法人交易資訊，判斷是否有顯著異常報酬存在，然後使用三因子模型解釋機構投資人對投資組合的異常報酬之影響。結果發現台灣的機構投資人並沒有積極利用市場異象賺取報酬，投資操作還是偏向保守；蘇偉榮（2018）利用事件研究法探討 2007 年至 2016 年，獲得天下雜誌 CSR 獎項上市櫃公司的股票報酬表現。並使用橫斷面迴歸分析，探討公司特性與 CSR 評分分數對報酬之影響。實證結果發現，在事件研究部分，多數的樣本顯示有提前反應的狀況，得獎公司在事件月前幾個月有顯著正向異常報酬，但事件後就無法透過得獎事件得到異常報酬；王克修（2018）以 2006 到 2016 各年度股價反轉訊號發生次數來觀察，在市場處於相對低檔的時候，其訊號發生的次數高於市場處於相對高檔的時候，表示運用三種均線特性產生交集後所設計出的判定方法，會在市場低檔的時候產生較多的進場機會；謝文良、林苑宜（2012）以 2006 至 2009 年共 4 年的所有台灣上市股票資料，並使用 8 個流動性指標，檢驗台股在各種層面上的流動性風險狀況，使用時間序列分析模型和就橫斷面取平均觀察流動性共變現象。實證結果發現，第一，超過 70% 股票都有流動性共變發生，且在台股中，是屬於跨產業的風險因素，反映流動性風險在台灣是不易被分散的系統性風險；第二，指數成分股中的流動性共變較非指數成分股間強烈；第三，流動性共變程度會受股票的規模所影響；第四，高波動性區段的流動性共變較低波動性強；第五，下跌市場中的個股對流動性敏感度稍微高於

上漲市場；K. Raesita & P.A. Mahadwartha（2020）探討台灣股票市場文化的價格群聚現象，因中華文化中有對於數字的迷信存在，於是透過台灣 2015-2019 年共 5 年的上市櫃公司收盤價資料，檢測股市是否有數字迷信的價格群聚現象發生，而不幸運數字在農曆新年期間和非農曆新年期間的出現頻率並無顯著差異；但幸運數字在農曆新年期間和非農曆新年期間的出現頻率有出現顯著差異。

三、股市預測

近幾年由於機器學習盛行，許多台股市場研究都是以機器學習做股市預測，包含常見的預測股價、預測股市趨勢、動態調整投資組合權重，甚至還有預測熊市發生時點和搭配交易策略的預測策略。（卓越, 2019；陳家蓁, 2020；Lai et al., 2019）都是對於股價預測做研究，卓越（2019）是針對台灣所有上市上櫃公司資料餵進 SVR 做訓練，並抓取其中的 5、10、15 支股票做回測，結果發現 SVR 對於賺取超額報酬是有所幫助的；陳家蓁（2020）針對台灣股市指數、電子類股指數、金融類股指數三個群組做預測績效比較，實證結果發現總體經濟因素和技術指標等變數之間的整體預測績效有些許差異，電子與金融類股的準確率相對較高，可能適合結合模型做出有意義的投資決策；Lai et al.

（2019）從台灣上市公司中抓取 5 家公司，預測未來五天市場訊息的平均，並且加入四種技術指標：KD、RSI、MACD、OBV，來考慮是購買、繼續持有或出售股票。（施育霖, 2019；Huang et al., 2019；Rosdyana Mangir Irawan Kusuma et al., 2019）則是針對股價趨勢上漲下跌做預測，施育霖（2019）探討金融類股的股價短期預測，收集台灣 10 家金融類股公司股票預測隔天股價趨勢；Huang et al.（2019）針對台灣股價指數資料，將輸入的 10 個技術指標分別做連續性和

類別性資料丟進模型做訓練，預測股價趨勢；Rosdyana Mangir Irawan Kusuma et al. (2019) 使用台灣 50 家公司資料，將其每日的市場資訊畫成不同區間的 K 線圖，透過圖像辨識深度學習的方式，用於預測股價趨勢，準確率高達 92.2%。而其他的像是童韋傑 (2020) 透過 SVM 預測配對交易下的兩支股票在未來是否會持續背離，以此發展出延後交易策略，實證結果發現延後交易策略的報酬率都會較一般配對策略好；陳彥安 (2019) 透過 LSTM 依照投資標的之三大法人每日買賣超的籌碼變化，動態調整 15 家上市公司投資組合權重，整體而言，投資績效表現優於臺灣 50 指數及效率前緣上 Sharpe Ratio 最大之最佳投資組合；李偉銘等人 (2015) 使用水泥、紡織、食品、造紙、塑化、機電、營建與金融等 8 大類股指數月收盤價資料，以 probit 模型分析描述預測熊市發生時段，並分析金融市場、整體物價水準、總體經濟活動、貨幣政策、與財政政策等狀態之總體經濟變數的樣本內與樣本外預測表現做研究。



第三章 研究方法

本章第一節先就實證資料來源、處理和特徵挑選的方式做說明，第二節介紹主成分分析，第三節說明資料不平衡會產生的問題，還有如何解決，第四節介紹本研究使用的深度學習演算法，第五節說明本研究引用集成學習（Ensemble Learning）的堆疊法（Stacking）所提出的架構，第六節講解使用什麼衡量指標判定模型好壞，圖 7 是本研究的方法流程架構圖。

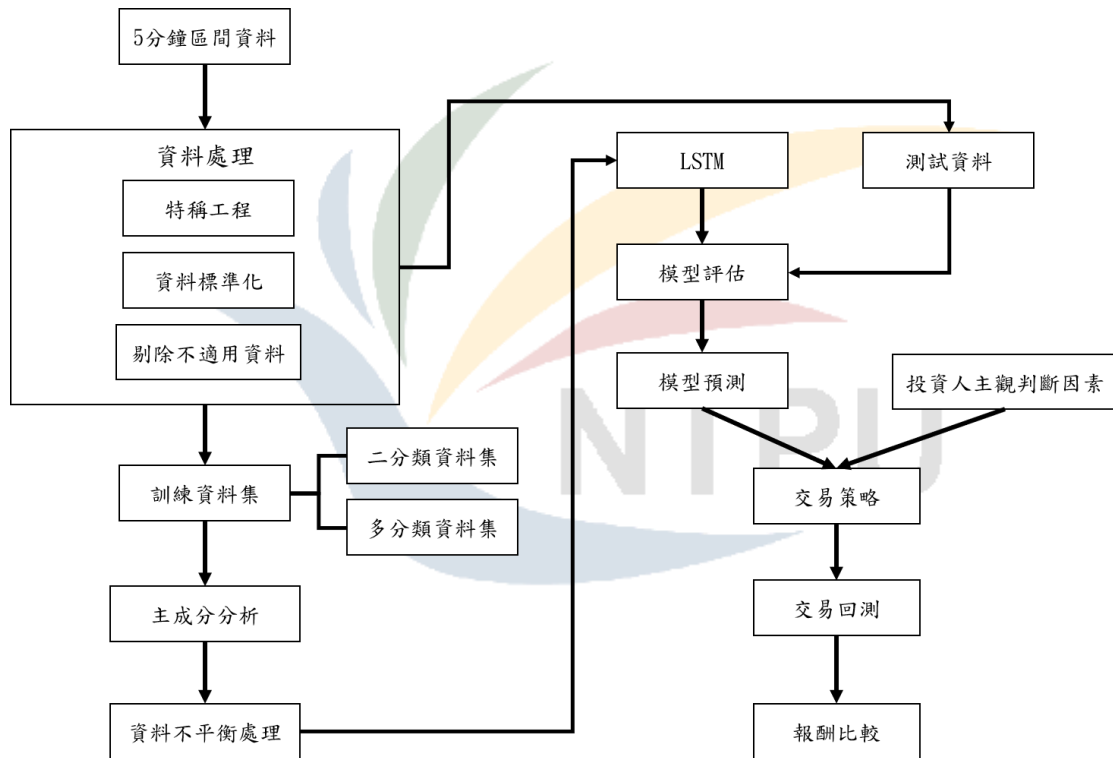


圖 7：研究方法流程架構圖

第一節 資料來源、處理與特徵選取

本研究資料來源為台灣證券交易所提供的 Open API 每日即時日內資料，由於資料是每日透過 Open API 取得，從零開始建立本研究之資料庫，資料期間為 2021 年 1 月 18 日至 2021 年 5 月 13 日，本研究將研究期間新上市之企業、至少一交易天無任何成交量的企業移除，加上由於台股有漲跌停限制，為了後續分析減少偏誤故將漲幅大於 9.95% 和跌幅小於 -9.95% 的資料給刪除，由於 2 月 17 至 2 月 19 日資料撈取出現技術性錯誤¹，為了避免後續技術指標計算上出現錯誤，將後續銜接的 2 月 22 日這一交易日刪除；4 月 8 日因技術性錯誤故刪除 4 月 9 日資料。本研究將 2021 年 1 月 18 日至 2021 年 3 月 12 日扣除掉春節、和平紀念日期間休市共 29 個交易日的台灣股市日內 5 分鐘區間資料共 911 家上市公司的 1,229,549 筆交易資料做為訓練模型資料集，之後再透過 2021 年的 3 月 15 日至 5 月 13 日的資料，扣除清明節、勞動節，共 41 個交易日的 1,795,756 筆交易資料為後續交易回測用，區間資料中包含個股每 5 分鐘的收盤價、開盤價、最高價、最低價和交易量（股）。

特徵選取參考 Ao et al. (2019) 所使用的技術指標，從中挑選出 13 個對預測股價區間趨勢較有效的技術指標：價格變動率、交易量變動率、移動平均 (MA)、指數移動平均 (EMA)、乖離率 (BIAS)、指數乖離率 (EBIAS)、移動平均價格震盪指標 (OSCP)、指數移動平均價格震盪指標 (EOSCP)、商品通道指標 (CCI)、聚散擺盪指標 (ADO)、真實波動幅度 (TR)、價量趨勢 (PVT)、能量潮指標 (OBV)，而前九個技術指標需要設定計算期間 q ，這裡以 $q=5, 10, 20, 30$ 分別去建立新的特徵，因此共有 40 個技術指標特徵，本研究

¹ API 不明原因無法撈取資料，因技術指標計算上會涉及到前幾天資料，為了避免資料缺失導致後續銜接上資料的計算偏誤，故將後一日資料也刪除。

也對這 40 個特徵依照個股自身資料分布做特徵標準化 (Normalization)，標準化有助於提升後續模型使用梯度下降法 (Gradient Descent) 最適化學習的收斂速度，且透過標準化將各特徵的距離壓縮至同樣區間，有助於模型預測的準確度，本研究採用平均值標準化 (Mean Normalization) 將特徵限縮在 -1 和 1 之間。

$$X_{normalization} = \frac{X_{raw} - \mu}{X_{max} - X_{min}} \in [-1, 1] \quad (1)$$

Huang et al. (2019) 提及將技術指標轉換成類別型指標，有助於加強未來股價趨勢預測能力。故本研究以 13 個技術指標為基礎進而建構 58 個常見技術指標門檻定義的類別型指標，因此本研究共使用 98 個特徵，表 1 列出各技術指標的計算公式， O_k 、 H_k 、 L_k 、 C_k 、 V_k 分別代表第 k 筆 5 分鐘區間的開盤價、最高價、最低價、收盤價和交易量。表 2 列出常見技術指標門檻定義原則。

做為預測下一區間的標籤，本研究分為五個模型，標籤分別是：

1. Model1：漲 = 1；平盤/跌 = 0
2. Model2：跌 = 1；平盤/漲 = 0
3. Model3：股價有跳動 = 1；無跳動 = 0
4. Model4：漲 = 1；平盤 = 0；跌 = -1
5. Model5：前三區間連續漲 = 1；其他 = 0；連續跌 = -1

Ao et al. (2019) 中將模型分為兩大類：跳動和不跳動、漲/平盤/跌，本研究引用這兩大類，並拓展三種額外標籤。前三項本研究定義為輔助標籤，第四項為主要標籤和最後的動能標籤，這裡運用集成學習中的堆疊概念，對於不同標籤分別建立 LSTM 模型預測下一區間，再進一步綜合各模型的預測結果，但稍微有所不同的在於本研究不把各預測值再放入新的模型中做訓練，而是直接透過不同模型的預測標籤，將它們組合並發展交易策略，增加對於股價趨勢的預測能力，讓結果能夠更加穩健，因此輔助標籤用於和主要標籤做堆疊，而動能標

籤則是因為動能相對於跳動方向是較具體的存在，故本研究也將動能趨勢作為標籤預測，而動能標籤也會搭配主要標籤做堆疊，發展交易策略。



表 1：技術指標的計算公式

技術指標	$q = 5, 10, 20, 30$
價格變動率 $\text{PROC}(q)_k$	$\text{PROC}(q)_k = \frac{C_k - C_{k-q}}{C_{k-q}}$
交易量變動率 $\text{VROC}(q)_k$	$\text{VROC}(q)_k = \frac{V_k - V_{k-q}}{V_{k-q}}$
移動平均 $\text{MA}(q)_k$	$\text{MA}(q)_k = \frac{\sum_{i=1}^q C_{k-i+1}}{q}$
指數移動平均 $\text{EMA}(q)_k$	$\text{EMA}(q)_k = \frac{2}{q+1} C_k + \frac{q-1}{q+1} \text{EMA}(q)_{k-1}$
乖離率 $\text{BIAS}(q)_k$	$\text{BIAS}(q)_k = \frac{C_k - \text{MA}(q)_k}{\text{MA}(q)_k}$
指數平均乖離率 $\text{EBIAS}(q)_k$	$\text{EBIAS}(q)_k = \frac{C_k - \text{EMA}(q)_k}{\text{EMA}(q)_k}$
移動平均振盪指標 $\text{OSCP}(q)_k$	$\text{OSCP}(q)_k = \frac{\text{MA}(q)_k - \text{MA}(q)_{k-1}}{\text{MA}(q)_{k-1}}$
指數移動平均振盪指標 $\text{EOSCP}(q)_k$	$\text{EOSCP}(q)_k = \frac{\text{EMA}(q)_k - \text{EMA}(q)_{k-1}}{\text{EMA}(q)_{k-1}}$
商品通道指標 $\text{CCI}(q)_k$	$\text{CCI}(q)_k = \frac{M_k - \text{SM}(q)_k}{0.015G(q)_k}$ <p>其中 $M_k = \frac{H_k + L_k + C_k}{3}$, $\text{SM}(q)_k = \frac{\sum_{i=1}^q M_{k-i+1}}{q}$,</p> $G(q)_k = \frac{\sum_{i=1}^q M_{k-i+1} - \text{SM}(q)_k }{q}$
聚散擺盪指標 ADO_k	$\text{ADO}_k = \frac{(H_k - O_k) + (C_k - L_k)}{2 * (H_k - L_k)}$
真實波動幅度 TR_k	$\text{TR}_k = \max\{H_k - L_k, L_k - C_{k-1}, H_k - C_{k-1}\}$
價量趨勢 PVT_k	$\text{PVT}_k = \frac{C_k - C_{k-1}}{C_{k-1}} * V_k$
能量潮指標 OBV_k	$\text{OBV}_k = \text{OBV}_{k-1} + \begin{cases} V_k, & \text{if } C_k > C_{k-1} \\ 0, & \text{if } C_k = C_{k-1} \\ -V_k, & \text{else} \end{cases}$

表一中提及的各技術指標，價格變動率和交易量變動率，都是衡量動量震盪程度的指標，計算出的值能夠反映價格和交易量的波動趨勢，而這兩種指標都需要設定參數 q ，表示要與多久之前的值作比較， $PROC(5)_{10}$ 代表第 10 筆 5 分鐘區間資料和第 5 筆資料的價格變動率，若第一筆資料從開盤後上午 9 點 05 分算起，為 9 點 50 分相對 9 點 25 分的某公司股價之價格變動率，交易量變動率同理。

移動平均和指數移動平均，主要用來衡量股價波動趨勢的指標，計算出的值能夠判斷未來潛在的股價趨勢， $MA(5)_{10}$ 代表第 6 筆 5 分鐘區間資料至第 10 筆之簡單平均，也可視為某公司在 9 點 50 分的 25 分鐘均線值。指數移動平均同理，只是計算上給予不同期間資料做了指數平滑的加權。

乖離率和指數平均乖離率，主要計算現行股價偏離趨勢狀況，當乖離率過大時有很大機率會在未來做出修正， $BIAS(5)_{10}$ 代表第 10 筆 5 分鐘區間資料的收盤價相對於第 10 筆資料的 25 分均線值偏離百分比，也可視為某公司在 9 點 50 分股價相對於現行 25 分鐘均線值的偏離比例，指數平均乖離率同理。

移動平均振盪指標和指數移動平均振盪指標，為衡量兩條平均線之間的偏離狀況， $OSCP(5)_{10}$ 代表第 10 筆區間資料的 25 分均線值相對第 9 筆資料的 25 分均線值之變化百分比，也等同於某公司 9 點 50 的 25 分鐘均線相較 9 點 45 分之 25 分鐘均線變化，指數移動平均振盪指標同理。

商品通道指標，為衡量股票超買或超賣狀況的震盪指標，使用上類似乖離率，當該股票是超買狀態，近期股價會下跌，反之被視為超賣，股價會上漲， $CCI(5)_{10}$ 可以視為第 10 筆區間資料現行的超買超賣狀況。

聚散擺盪指標，主要用於判斷買賣雙方拉鋸的力道，藉由雙方的相對關係判斷未來股市買氣的強弱程度，以此判斷股價趨勢， ADO_{10} 代表第 10 筆資料當下的買賣雙方力道狀況。

真實波動幅度，是用於衡量股票波動性幅度的指標， TR_{10} 代表第 10 筆資料當下的波動幅度；價量趨勢和能量潮指標都是同時衡量價格跟交易量變化的動能指標， PVT_{10} 和 OBV_{10} 代表第 10 筆資料當下的動能趨勢。

表 2：常見技術指標門檻定義原則

	常見技術指標門檻定義原則 (x=3, 5)
價格變動率	前 x 區間都 > 0 則預期下一區間會漲 前 x 區間都 < 0 則預期下一區間會跌
移動平均	MA(5), MA(10) 為短線；MA(20), MA(30) 為長線 前 x 區間 $MA(5) > MA(20)$ 則預期下一區間會漲 前 x 區間 $MA(5) < MA(20)$ 則預期下一區間會跌 前 x 區間 $MA(10) > MA(30)$ 則預期下一區間會漲 前 x 區間 $MA(10) < MA(30)$ 則預期下一區間會跌
指數移動平均	EMA(5), EMA(10) 為短線；EMA(20), EMA(30) 為長線 前 x 區間 $EMA(5) > EMA(20)$ 則預期下一區間會漲 前 x 區間 $EMA(5) < EMA(20)$ 則預期下一區間會跌 前 x 區間 $EMA(10) > EMA(30)$ 則預期下一區間會漲 前 x 區間 $EMA(10) < EMA(30)$ 則預期下一區間會跌
乖離率	前 x 區間都 > 0.08 則預期下一區間會跌 前 x 區間都 < -0.08 則預期下一區間會漲
指數平均乖離率	前 x 區間都 > 0.05 則預期下一區間會跌 前 x 區間都 < -0.05 則預期下一區間會漲
移動平均振盪指標	前 x 區間都 > 0 則預期下一區間會漲 前 x 區間都 < 0 則預期下一區間會跌
指數移動平均振盪指標	前 x 區間都 > 0 則預期下一區間會漲 前 x 區間都 < 0 則預期下一區間會跌
商品通道指標	前 x 區間都 > 100 則預期下一區間會跌 前 x 區間都 < -100 則預期下一區間會漲
聚散擺盪指標	前 x 區間都 > 0.5 則預期下一區間會漲 前 x 區間都 < -0.5 則預期下一區間會跌

表 2 對於常見技術指標門檻定義原則的說明，若是價格變動率前幾區間都持續大於 0，可以推測存在動能趨勢，因此預測下一區間也會漲，反之，小於 0 代表下一區間可能會跌，本研究將前 x 區間的參數定為 3 和 5，分別代表弱訊號和強訊號。

移動平均可分為短期均線和長期均線，若前 x 區間的短期均線皆大於長期均線，視為短線向上突破長線出現黃金交叉，也隱含動能趨勢存在，故下一區間判斷會漲，反之短期均先皆小於長期均線，視為死亡交叉，下一區間很有可能會跌，指數移動平均同理。

當乖離率過大時有很大機率會在未來做出修正，若是乖離率發生很大的正乖離時，判斷下一區間會發生修正因此下跌，反之，若出現較大的負乖離率，下一區間可能會修正上漲，指數平均乖離率同理，在這邊本研究將乖離率的門檻值設為 0.08，也就是乖離率大於 8% 時，則視為乖離過大，而指數平均乖離率的門檻值設為 0.05，也就是 5%。

移動平均振盪指標類似價格變動率，若前幾區間持續大於 0，代表移動平均變化是持續往上，也可推測存在動能趨勢，因此下一區間會漲，反之則會跌，指數移動平均振盪指標同理。

商品通道指標若大於 100 被視為該股票現階段為超買，故下一區間很有可能會下跌，若小於 -100 則視為超賣，故下一區間可能會上漲；聚散擺盪指標若大於 0.5 代表目前該股票買氣強勁，若前幾區間皆大於 0.5，可判斷該股的買氣強勁趨勢持續存在，故判斷下一區間股價會持續上漲，反之小於 0.5 時，代表買氣不足，下一區間股價很有可能會持續下跌。

本研究所使用特徵共 98 個，為了避免維度災難（curse of dimensionality）而導致模型過擬合（overfitting），所以使用主成分分析（Principal Component Analysis, PCA）進行降維，透過將現有特徵以線性組合方式組合出新的主成分，並且此主成分能夠保有原特徵最多的解釋力，也代表此主成分有最大的變

異量，並透過設定累積貢獻比率的門檻來選定要使用的主成分數量，本研究設定門檻為 99%，所以選擇 32 個主成分輸入進模型做訓練。使用主成分分析除了能避免過擬合，也能進而提升後續模型訓練速度，有助於應用在快速變動的實際交易上。圖 8 表達主成分的累積貢獻比率。由於主成分分析會因為資料單位不同，而產生誤差，本研究已將資料做標準化限縮在同樣區間，避免錯誤發生。

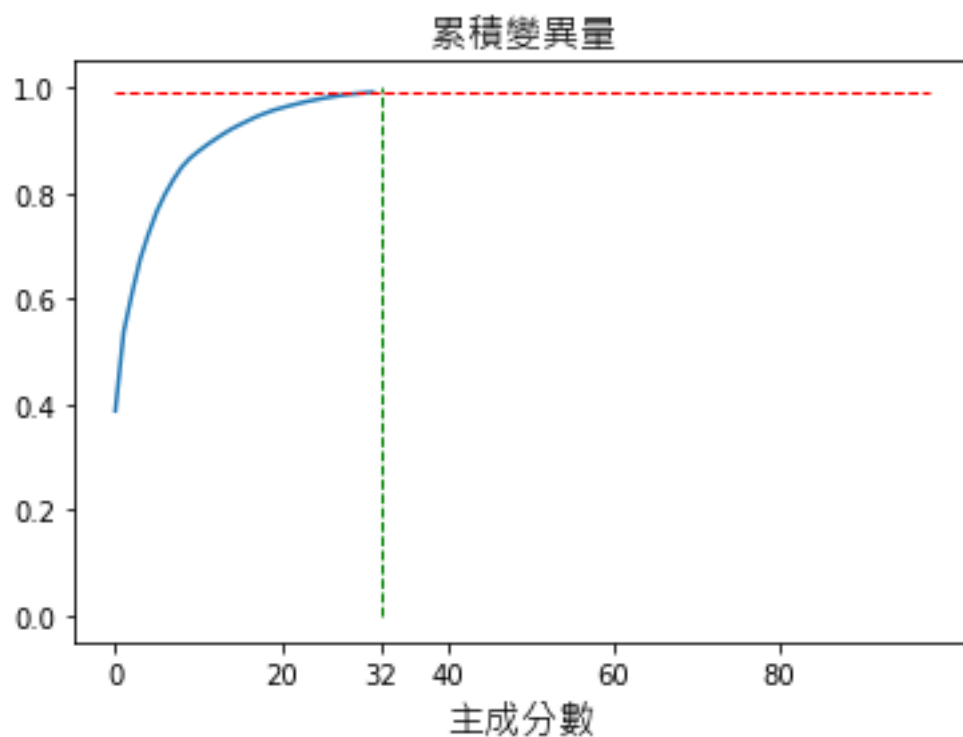


圖 8：主成分的累積貢獻比率

第二節 資料不平衡處理

從表 3 可以觀察到本研究主要分類的五大標籤各自的資料類別比例，可以發現 Jump1、Jump2、Jump5 有資料不平衡的狀況發生，不平衡資料（Imbalanced Data）在各領域一直是個挑戰，例如：信用卡盜刷、癌症診斷等，不平衡資料會導致模型只學習到類別比例較高的資料，進而忽略我們想觀察但類別比例較少的資料，使得模型預測能力不佳，若就衡量指標來觀測，則會有準確度悖論（Accuracy Paradox）產生，讓準確度（Accuracy）失效。

表 3：各模型類別比例

類別比例 (%)	Model1 漲/其他	Model2 跌/其他	Model3 跳/不跳	Model4 漲/平/跌	Model5 連續漲/其他/連續跌
1	22.45	24.79	47.3	22.45	2.27
0	77.55	75.21	52.7	52.77	95.79
-1	-	-	-	24.78	1.94
總計	100				

因此解決資料不平衡的處理方法相繼出現，He and Garcia（2009）整理出常用的幾種方法背後的運作原理、優缺點和相關文獻，包含：隨機過採樣和隨機欠採樣（Random Oversampling and Undersampling）、資料生成合成樣本（Synthetic Sampling with Data Generation）、代價敏感法（Cost-Sensitive Methods）等，其中 Chawla et al.（2002）所提出的合成少數過採樣技術（Synthetic Minority Oversampling Technique, SMOTE）被證實在廣大領域中在處理資料不平衡問題上都得到有力的表現，Ao et al.（2019）便是透過 SMOTE

來生成少數樣本，並輸入進分類模型中，增進預測股價跳動的分類表現，He and Garcia (2009) 中也提及以 SMOTE 為基礎延伸而出的許多處理方法，像是 Borderline SMOTE、SMOTE-ENN、SMOTE-Tomek Links 以不同面向優化 SMOTE 的方法，本研究考量到計算效能與增進模型效能之間的平衡，雖然實際操作上混合的方式真的有優化分類狀況，但幅度並不明顯，且上述三種在計算上過於複雜，考量到每次訓練計算時間，本研究選擇使用 SMOTE。

SMOTE 這個演算法執行步驟如下，圖 9 為 SMOTE 的示意圖呈現：

1. 找出與少數類別樣本 X_i 最近的 k 個少數類別樣本鄰近點
2. 在 k 個鄰近點中隨機選擇一個 X_j 做為生成新樣本的基底
3. 計算 X_i 和 X_j 距離差異， $\Delta = X_j - X_i$
4. 產生一個介於 0 - 1 之間的隨機亂數 δ
5. 生成新的樣本點 $X_{new} = X_i + \Delta * \delta$

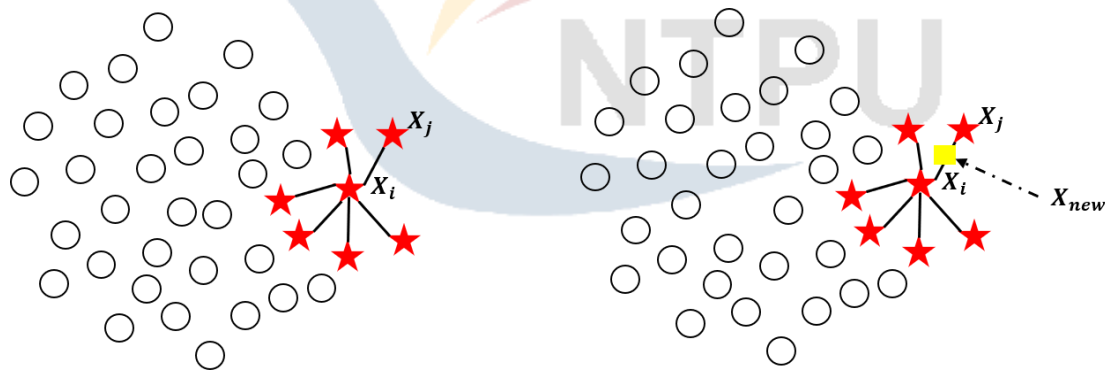


圖 9：SMOTE 示意圖

在二元分類模型中，透過 SMOTE 讓次要類別的資料個數過採樣至相等於主要類別的資料個數，本研究選擇對 Jump1、Jump2 進行處理；而在三分類模型中，透過 SMOTE 讓主要類別以外的兩類別資料個數都過採樣至主要類別的資料個數，本研究選擇對 Jump5 進行處理。

第三節 長短期記憶模型

長短期記憶模型（Long Short-Term Memory，以下稱 LSTM），是遞歸神經網絡（Recurrent Neural Networks，以下稱 RNN）衍生出來的模型，由 Hochreiter & Schmidhuber（1997）所提出，專門處理序列相關的資料，被廣泛地應用在：語音轉成文字、時間序列等問題上，而 LSTM 解決了 RNN 在長期記憶上表現不佳的缺點，因此較 RNN 來得頻繁被使用，圖 10 為 LSTM 的內部神經元運作模式。

LSTM 主要有三層匣門在控制記憶單元（Memory cell），分別是輸入匣層（Input gate layer）、遺忘匣層（Forget gate layer）和輸出匣層（Output gate layer）來決定哪些資料能被存進記憶單元。

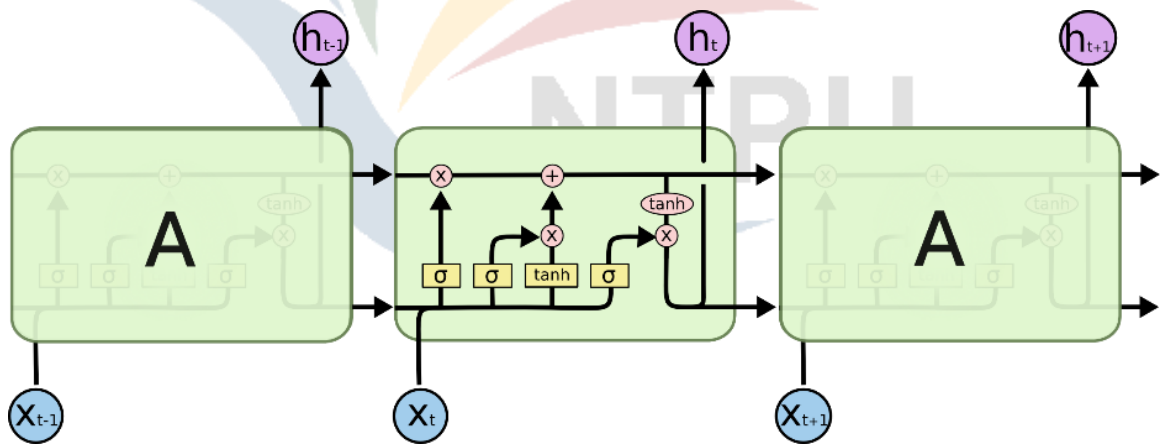


圖 10：LSTM 內部神經元（Neurons）運作模式（Oinkina, 2015）

首先是遺忘匣層，將輸入資料（ x_t ）和上段記憶單元的輸出值（ h_{t-1} ）乘上權重（ W_f ）加上殘差項（ b_f ）後，會經由激勵函數（Activate function），通常選用 Sigmoid 函數（ σ ），來決定資訊會不會通過，要留下多少資訊。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

接下來是輸入匣層，決定有哪些新資訊要被留下來，這邊要拆成兩部分，第一是 Sigmoid 函數用來決定哪些資訊需要被更新，另一個為 tanh 函數，創造一組新的向量值 (\tilde{C}_t)，在下一步會將兩者合併一起更新資訊。

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

再來就是更新記憶單元的步驟，將上次記憶單元的資訊 (C_{t-1}) 乘上遺忘匣層的 f_t 加上輸入匣層的值 ($i_t * \tilde{C}_t$) 成為新的記憶單元資訊 (C_t)。

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5)$$

最後透過輸出匣層決定有多少資訊要被輸出，讓輸入資料和上段記憶單元的輸出值先通過 Sigmoid 函數，決定有多少記憶單元資訊能夠通過 (o_t)，然後讓新的記憶單元資訊經過 tanh 函數後再乘上 o_t ，就會得到這次輸出值 (h_t)

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(C_t) \quad (7)$$

由於股市 5 分鐘區間資料是屬於高度序列相關的資料型態，因此本研究使用專門處理序列相關的 LSTM 模型作為主要模型，陳彥安 (2019)、Chun Yuan Lai (2019) 和 Jingyi Shen et al. (2020) 也都透過使用 LSTM 得到良好的預測表現。

第四節 交叉驗證和集成學習

傳統上會將資料集切成訓練集 (Training set) 和測試集 (Test set)，透過訓練集訓練出來的模型，再用訓練樣本外的測試集去測試模型表現，以此來確保模型沒有過擬合或欠擬合 (Underfitting) 的情況發生，但在特定情況下只切一次測試集做測試，其實還是有可能會產生錯誤，因此使用交叉驗證來檢定模型的表現可以更加確保模型的穩健性，而且交叉驗證可以確保資料都被使用到，而不會因為一次的拆分，只訓練到特定的資料。本研究使用分層 K 折交叉驗證 (Stratified K-fold Cross-Validation)，這裡的 K 指的是分群，本研究設定 $K = 10$ ，代表會進行 10 次的訓練集和測試集拆分，而分層代表的是每次拆分都會確保拆分後的資料集跟原資料集的類別比例是相同的，每次迭代都會重新訓練模型並透過測試集做測試，之後將 10 次的結果取平均，再用這個平均來衡量模型的好壞，圖 11 呈現分層 K 折交叉驗證的過程。

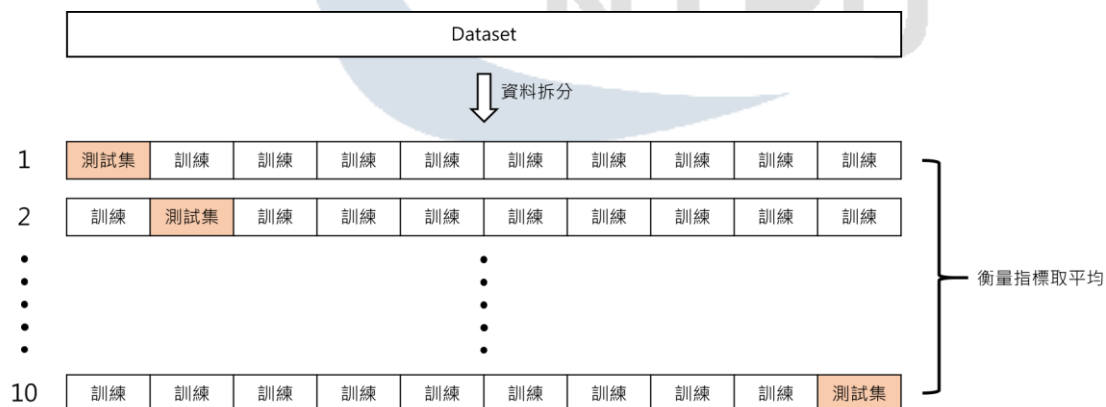


圖 11：交叉驗證架構

集成學習是透過組合多種模型、學習演算法來獲得比單獨的機器學習模型更佳的效能。常見的集成學習有拔靴集成法（Bootstrap Aggregating, Bagging）、多模激發法（Boosting）、堆疊法（Stacking）這三種，常見的分類模型隨機森林樹就是使用拔靴集成法。本研究運用堆疊法的概念，透過將股價趨勢跳動以不同形式轉化成標籤，並丟進不同的 LSTM 模型中做學習，再將預測的結果搭配常見技術指標門檻定義原則去建構交易策略，因為預期不同模型能夠學習到股價趨勢的某些部分，所以多模型構建出的策略能夠涵蓋較廣泛的交易資訊，本研究並沒有將預測結果再餵進新的模型中做學習，而是直接將多模型預測結果拿來使用。



第五節 模型衡量指標

本研究使用分類問題常見的衡量指標，準確率（Accuracy），代表在所有分類樣本中，預測正確之樣本相較總體樣本的比例有多少；召回率（Recall），旨在預測值中同一類樣本有多少樣本是被正確預測；精確率（Precision），在真實值中同一類樣本有多少被正確預測分類的比例為多少，和 F1 分數（F1 Score），其中準確率是最常見的衡量指標，而 F1 分數則是平衡召回率和精確率的指標，因此可以最佳的衡量模型整體分類能力，尤其是當資料有不平衡狀況發生時，F1 分數常常做為主要衡量指標。混淆矩陣（Confusion Matrix）呈現模型預測值和實際值的正確錯誤數量，並以矩陣內容為基礎去計算上述提及的衡量指標，本研究將各模型的主要跳動類別定為混淆矩陣中的 True Positive 做計算，表 4 呈現二元分類混淆矩陣，表 5 呈現三分類混淆矩陣，下方計算公式三分類指標只呈現類別一做範例公式。

表 4：二元分類混淆矩陣

二元分類混淆矩陣		預測值	
真實值		Class1	Class2
	Class1	True Positive (TP)	False Positive (FP)
	Class2	False Negative (FN)	True Negative (TN)

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

$$Recall = \frac{TP}{TP + FN} ; Precision = \frac{TP}{TP + FP}$$

$$F1\ Score = \frac{2 * Recall * Precision}{(Precision + Recall)}$$

表 5：三分類混淆矩陣

三分類混淆矩陣		預測值		
		Class1	Class2	Class3
真實值	Class1	a	b	c
	Class2	d	e	f
	Class3	g	h	i

$$Accuracy = \frac{a + e + i}{a + b + c + d + e + f + g + h + i}$$

$$Recall_{class1} = \frac{a}{a + d + g} ; Precision_{class1} = \frac{a}{a + b + c}$$

$$Recall_{all} = \frac{Recall_{class1} + Recall_{class2} + Recall_{class3}}{3}$$

$$Precision_{all} = \frac{Precision_{class1} + Precision_{class2} + Precision_{class3}}{3}$$

$$F1\ Score_{class1} = \frac{2 * Recall_{class1} * Precision_{class1}}{(Precision_{class1} + Recall_{class1})}$$

$$F1\ Score_{all} = \frac{F1\ Score_{class1} + F1\ Score_{class2} + F1\ Score_{class3}}{3}$$

第四章 實證結果

第一節 模型訓練表現

圖 12 為五個模型的訓練資料集混淆矩陣，表 6 為五大模型的訓練後衡量指標表現，在此選用加權出現頻率計算的衡量指標，從衡量指標可以發現在準確度上五大模型表現都算還不錯，其中 Model1、Model2、Model5 的準確度甚至在 70%以上，而 F1 Score 部分，除了 Model4 低於 60%，其他 Model 的 F1 Score 都表現不錯，Model5 甚至到 79%，綜合上述兩個指標的表現差異，本研究最主要的 Model4 在表現上只有尚可，推測可能原因一為股價趨勢預測在現有特徵中沒有存在有力的資訊幫助模型學習判別趨勢，因此本研究引用集成學習中的堆疊法，透過 Model1-3 這三個預測輔助標籤的模型的預測值，輔助 Model4 來制定交易策略，以減緩單一模型預測能力不佳的情況。也發現 Model5 在整體分類表現上表現是五大模型中最良好的，也推論動能可能存在於股市，且相較於股價跳動趨勢，動能是個較具體的存在，因此模型學習較良好。

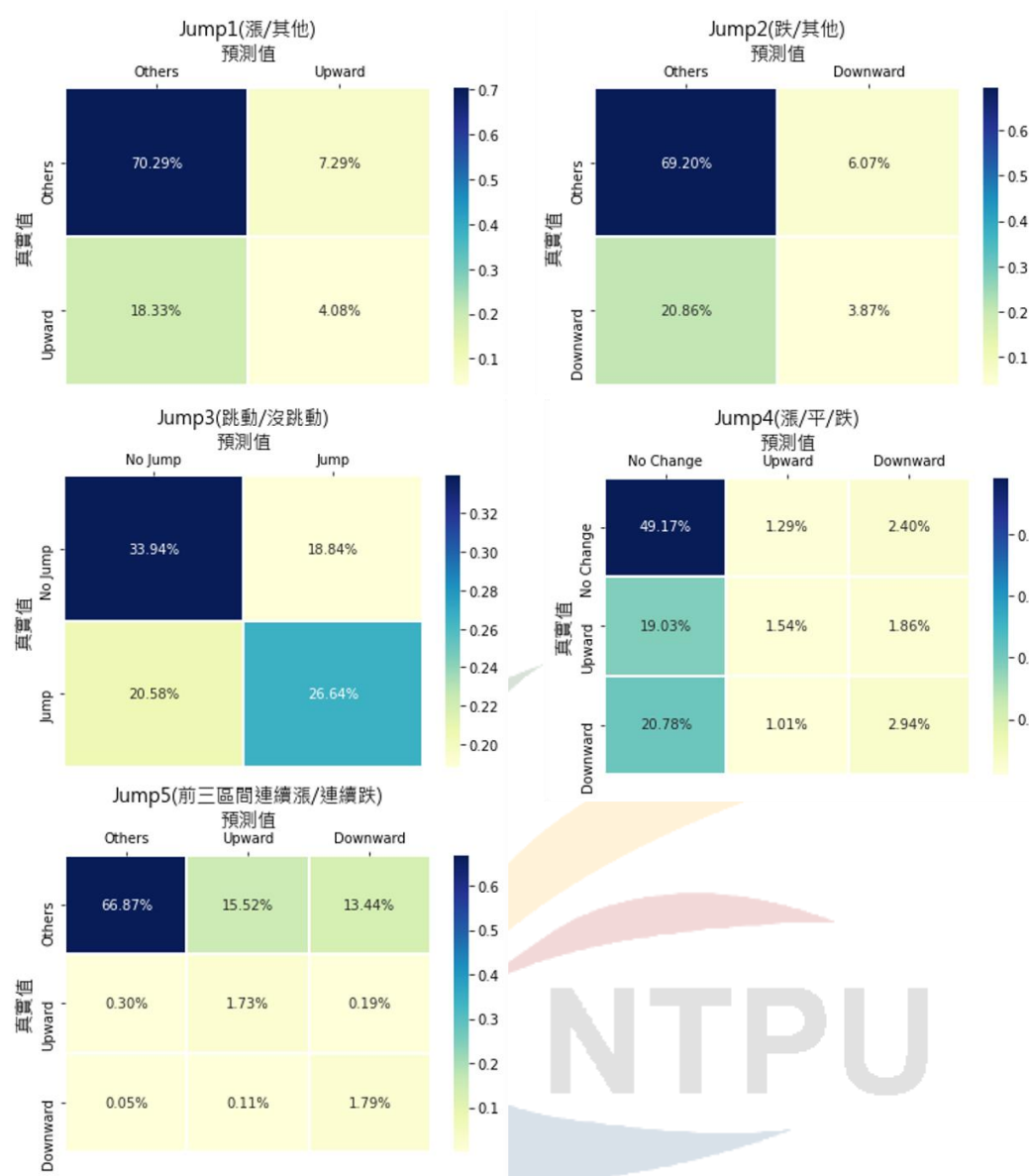


圖 12：五大模型訓練集混淆矩陣

表 6：五大模型訓練集衡量指標

(%)	Model1	Model2	Model3	Model4	Model5
Accuracy	74	73	61	54	70
Recall	70	67	61	48	96
Precision	74	73	61	54	70
F1 Score	71	69	61	44	79

第二節 交易回測

本研究在做交易回測時，盡可能地模擬實際交易場景會遇到的狀況，包含考慮到交易成本，買進時當沖手續費為 0.1425%，賣出時當沖手續費需再加上證券交易稅 0.15%，故賣出手續費為 0.2925%，當每 5 分鐘資料從 Open API 進入資料庫時立即就模型預測結果做出相對應的行動。至於資產組合配置部分，由於非本研究的重點，因此回測時直接就 911 家上市公司皆投入相同本金的方式（單看個股報酬率）做操作，再和從台灣經濟新報資料庫 TEJ 撈取的加權報酬指數（含息的大盤指數）日報酬做比較，由於加權報酬指數相較台灣加權指數來說更貼近真實報酬值，故本研究使用加權報酬指數。而本研究的回測雖然是當沖操作，但是可以接受股票持有至下一交易日，回測期間為 3/15 ~ 5/13 共 41 個交易日，除了模型預測區間趨勢之外，本研究也將傳統的動能策略加入交易策略中做比較，圖 13 為交易回測的虛擬碼。

以下為本研究所採四大類別的十一個交易策略：

類別一：純 Model 預測區間趨勢

1. Model14 預測區間趨勢結果為交易訊號
2. Model14 預測區間趨勢搭配兩模型
 - Model1 預測會漲為買入訊號
 - Model2 預測會跌為賣出訊號
3. Model14 預測區間趨勢搭配三模型
 - Model1 預測會漲 + Model3 預測會跳動為買入訊號
 - Model2 預測會跌 + Model3 預測會跳動為賣出訊號

類別二：純動能策略（本研究定義第一類為弱動能；第二類為強動能）

4. 前三區間價格變動率 >0 則為買入訊號，反之， <0 則為賣出訊號
5. 前五區間價格變動率 >0 則為買入訊號，反之， <0 則為賣出訊號

類別三：Model 預測區間趨勢+動能策略

6. Model4 預測區間趨勢搭配三模型+弱動能策略
 - Model1 預測會漲+Model3 預測會跳動+前三區間價格變動率 >0 為買入訊號
 - Model2 預測會跌+Model3 預測會跳動+前三區間價格變動率 <0 為賣出訊號
7. Model4 預測區間趨勢搭配三模型+強動能策略
 - Model1 預測會漲+Model3 預測會跳動+前五區間價格變動率 >0 為買入訊號
 - Model2 預測會跌+Model3 預測會跳動+前五區間價格變動率 <0 為賣出訊號

類別四：Model 預測動能趨勢

8. Model5 預測動能趨勢+弱動能策略
 - 前三區間價格變動率 >0 則為買入訊號，反之， <0 則為賣出訊號
9. Model5 預測動能趨勢+強動能策略
 - 前五區間價格變動率 >0 則為買入訊號，反之， <0 則為賣出訊號
10. Model5 預測動能趨勢+Model14 預測趨勢+弱動能策略
 - 前三區間價格變動率 >0 則為買入訊號，反之， <0 則為賣出訊號
11. Model5 預測動能趨勢+Model14 預測趨勢+強動能策略
 - 前五區間價格變動率 >0 則為買入訊號，反之， <0 則為賣出訊號


```

# 建立空白列表↓
賣出紀錄=[]↓
For 每一支個股代碼(x) in (911 支個股代碼列表):↓
    # Validation_data 為 3/15 ~ 5/13 回測資料↓
    stock_val = Validation_data[Validation_data['個股代碼'] == x]↓
    成本 = 0↓
    個股報酬 = 0↓
    交易量 = 0↓
    購買價格 = 0↓
    賣出價格 = 0↓
    For 每一 row in stock_val.iterrow():↓
        if (買入訊號):↓
            while 交易量 == 0:↓
                交易量 += 10000↓
                購買價格 = row 的 5 分鐘區間收盤價↓
                成本 = 購買價格 * 1.001425 ↓
            elif (賣出訊號):↓
                while 交易量 > 0:↓
                    賣出價格 = row 的 5 分鐘區間收盤價↓
                    交易量 = 0↓
                    個股報酬 = ((賣出價格*0.997075 - 成本) / 成本) * 100↓
                賣出紀錄.append(個股報酬)↓

```

圖 13：交易回測虛擬碼

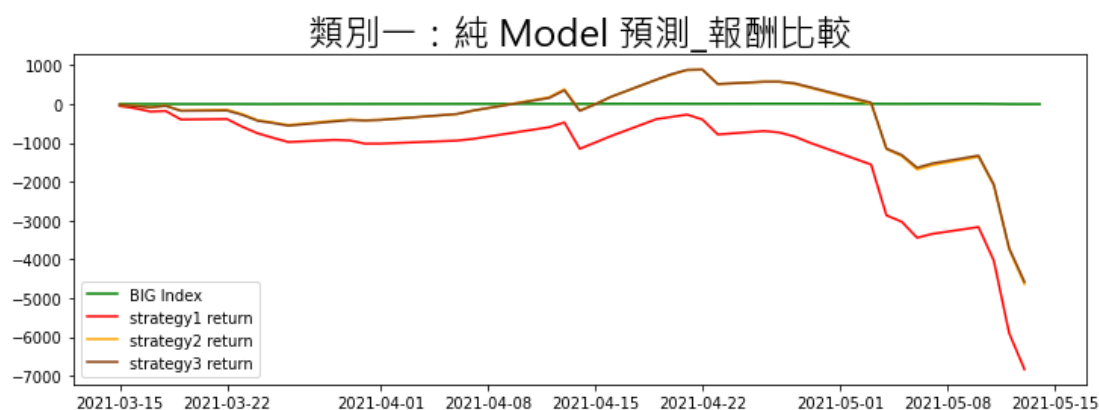


圖 14：類別一累積報酬比較圖

圖 14 主要是以 Model1 預測區間趨勢是否會漲、Model2 預測區間趨勢是否會跌、Model3 預測區間股價是否跳動、Model4 預測區間趨勢跳動方向，以上四個模型去做堆疊形成的三種交易策略。從圖中可以發現堆疊後的回測表現優於單一模型預測結果。

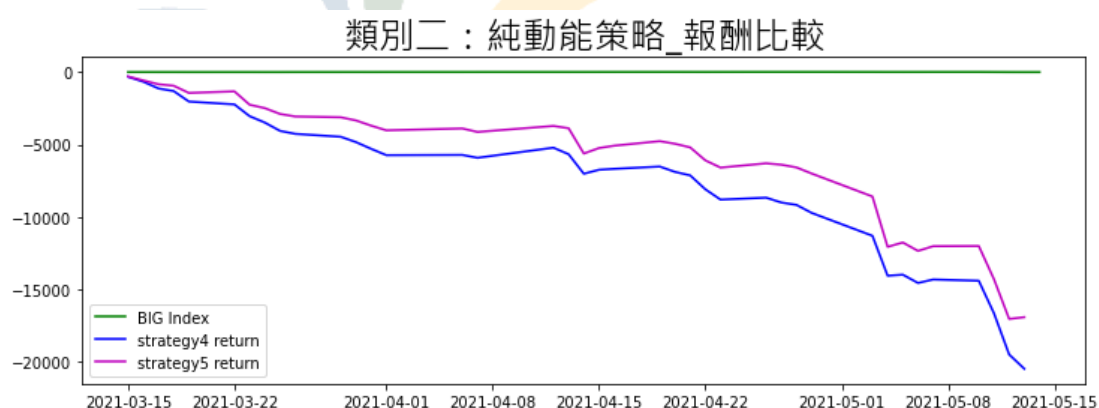


圖 15：類別二累積報酬比較圖

圖 15 是以常見的弱動能策略（前三區間價格變動率大於 0，判定下一區間會上漲）和強動能策略（前五區間價格變動率大於 0，判定下一區間會上漲）所建構的兩種交易策略。途中可以發現強動能策略的報酬表現優於弱動能策略，因此推斷股市確實存在動能趨勢，一但出現強動能趨勢，便能得到較良好的交易訊號。

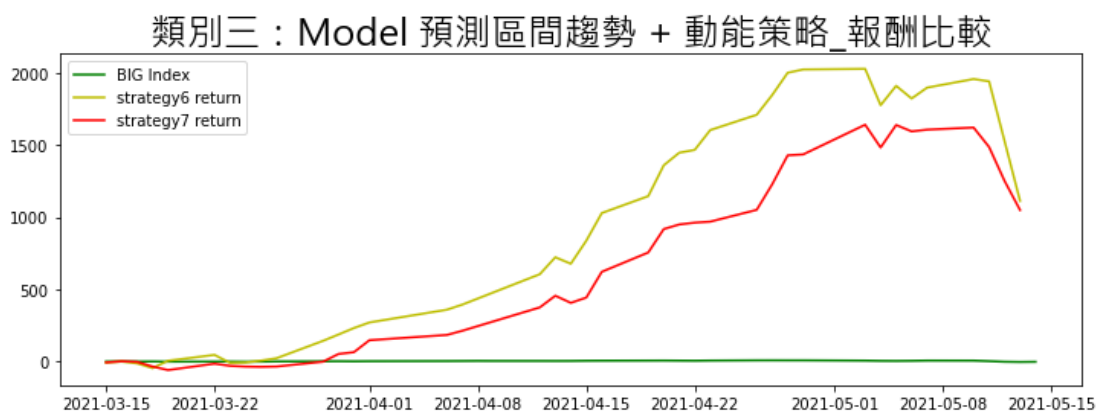


圖 16：類別三累積報酬比較圖

圖 16 是讓 Model1、Model2、Model3、Model4 所建構之堆疊交易策略，再分別與弱動能和強動能策略做搭配，建構兩種交易策略。從圖中能發現讓模型預測結果和動能策略做搭配所建構之策略，能在股市中獲取超額報酬並勝過加權報酬指數，回測表現也優於單純用模型預測結果，或是單純動能策略的表現。

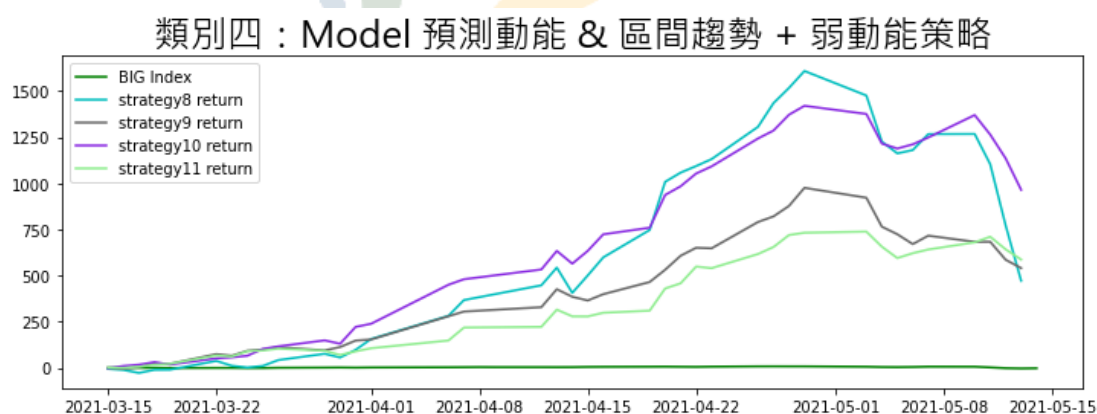


圖 17：類別四累積報酬比較圖

圖 17 讓 Model5 預測動能趨勢搭配弱動能和強動能策略，並進一步再與 Model1、Model2、Model3、Model4 做堆疊建構四種交易策略。從圖中能發現這四種策略能在股市中獲取超額報酬，再次驗證堆疊法，模型預測結果與動能策略搭配，在交易回測表現上優於單一模型或是單一方式做為交易訊號判定準則之表現。

圖 14 可以發現使用堆疊法建構的交易策略三優於只有一個模型預測的交易策略一，但也可以發現交易策略三在 4/12-4/16 這一週的累計報酬才開始贏過加權報酬指數，而 5/3 這一交易日開始，剛好遇上台灣股市的急跌趨勢，而 5/12 也發生台股盤中歷史最大跌點 1417 點，因此累積報酬大幅下降；圖 15 則可以發現強動能策略的報酬表現優於弱動能策略，也初步推測動能確實存在股市之中，但整體的累積報酬都低於加權報酬指數；圖 16 在加入了常見技術指標門檻定義原則之動能策略後，讓預測模型表現大幅提升，約 3/22 開始交易策略六和七的累積報酬就持續地高過加權報酬指數，而 5/3 開始的急跌趨勢，一開始表現平平無明顯降低，但 5/12 這天還是因整體股市影響導致累積報酬大幅降低；圖 17 中發現交易策略八和十表現都很良好，但一樣在 5/12 這天受到影響。

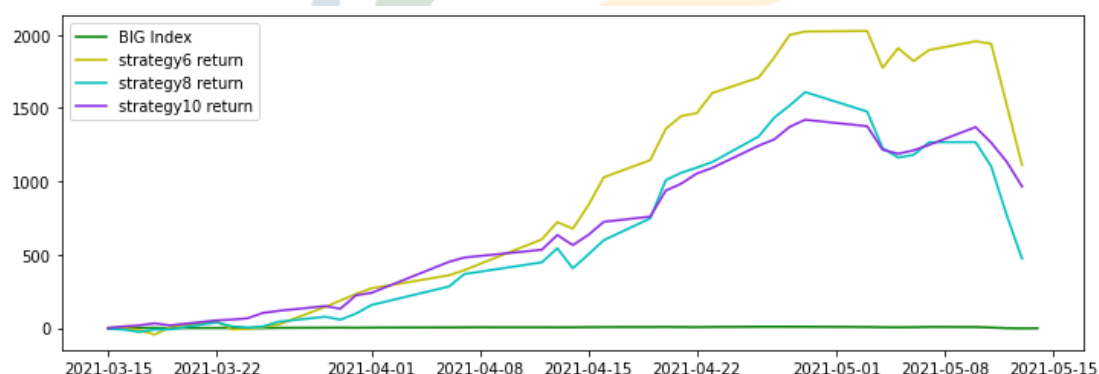


圖 18：表現最佳三大交易策略累積報酬圖

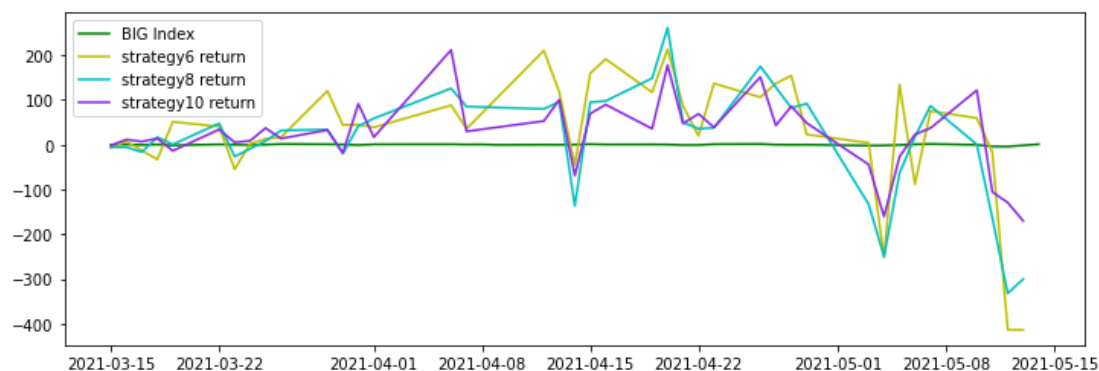


圖 19：表現最佳三大交易策略日報酬圖

接下來就表現最佳的三個交易策略做分析，圖 18 可以發現交易策略六整體累積報酬是最佳的，但圖 19 的日報酬圖中可以發現在 5/12 這天交易策略六卻是日報酬率最低的，而交易策略十的日報酬雖然為負但卻是三個策略中最高的，收盤表現 3/19 跌幅 1.34、3/24 跌幅 0.9%、4/13 跌幅 0.21%、5/3 跌幅 1.96%、5/12 跌幅 4.11%，有較大跌幅的這幾個交易日，交易策略的日報酬也受到影響，甚至低於加權報酬指數，但可以發現交易策略十雖然不是累積報酬最佳的策略，但在遇到趨勢走跌的狀況下，交易策略十的抗跌能力是三個策略中最佳的。



第三節 實證結論

表 7 為本研究各交易策略所依據的模型整理。從實證結果可以發現以下幾個特點：

1. 堆疊法建構的交易策略回測表現上優於單一模型預測的結果
2. 強動能交易策略相較弱動能交易策略表現較佳，可推測動能趨勢確實存在於股市
3. 預測動能趨勢的模型在分類衡量指標和交易回測上表現較佳，也證實深度學習模型善於學習隱藏在資料中相對具體的趨勢
4. 不管是哪個交易策略在遇到類似 5/12 的系統性風險下殺時，日報酬都會受到一定程度的影響
5. 交易策略六、八、十的表現中，雖然交易策略十不是累積報酬最佳之策略，但在面對系統性風險發生時，抗跌能力確是最好的

雖然模型四在分類衡量指標上表現不佳，但透過搭配模型一、二、三所建構之交易策略，彌補單一模型預測不佳的狀況，而在交易回測上也證實透過集成學習堆疊法所建構的交易績效表現是優於單一模型的，而再搭配常見技術指標門檻定義原則之動能策略判斷後，累積報酬甚至能超過加權報酬指數，由此可見單一深度學習演算法能從日內資料中學習到部分的隱藏交易訊號，若再加上常見技術指標門檻定義原則可使整體交易策略績效表現更佳。但是在遇到系統性風險發生的情況下，不管是哪個交易策略在抗跌能力上都是有待改善，由此可見，技術指標雖然能有效地預測出交易訊號，但深度學習演算法在特徵選取上還有能夠改善之空間，或是在常見技術指標門檻定義原則策略上，能建構除了動能策略外更加多樣化之策略，以捕捉股市的多樣性訊號。

表 7：各交易策略使用模型整理

交易策略	Model1	Model2	Model3	Model4	Model5	弱動能	強動能
1				V			
2	V	V		V			
3	V	V	V	V	V		
4						V	
5							V
6	V	V	V	V		V	
7	V	V	V	V			V
8					V	V	
9					V		V
10				V	V	V	
11				V	V		V

第五章 結論

第一節 研究貢獻

本研究發現透過集成學習堆疊法的方式確實有助於交易訊號的預測效能，也發現深度學習模型預測動能會較直接預測區間趨勢結果在模型衡量指標有較佳的表現，這也證明了深度學習模型善於從數據中學習隱藏趨勢，動能趨勢相對於區間漲跌趨勢是一個相對具體的存在，因此越具體的隱藏趨勢也能夠被深度模型較良好的學習。而後續的交易策略雖然只是搭配簡單的動能策略，也讓累積報酬有更良好的表現，模型雖然能夠抓出潛藏的交易訊號，但如果搭配常見技術指標門檻定義原則，除了能讓整體模型學習表現更好，在後續交易回測上也會有較佳的表現。本研究認為若能透過技術指標在台灣股票市場賺取超額報酬時，代表現行台股是一個雜訊較大的市場，若大眾可以透過機器人理財得到超額利潤，在現行制度面不變下，能使得市場的價值更接近真實價值。

從過往的深度學習預測股價或是股價趨勢的文獻中可以發現，過往文獻大都注重在提升模型的衡量指標和就不同演算法模型做衡量指標的比較，本研究加入了交易回測的實證結果，讓常見技術指標門檻定義原則交易策略能夠一起被考慮和運用，對於模型表現衡量除了傳統的衡量指標，也讓實務交易回測報酬作為衡量策略甚至模型的表現上。而就本研究目前認知文獻中，由於日內資料取得不易，多數研究都是針對日資料股價做探討，對於台股市場少有像是當沖相關的實證交易論文，但隨著 Open API 應用越加廣泛，往後的資料取得較容易後，期許關於台股當沖相關之研究能跟著增加。

第二節 研究建議

本研究建議後續的延伸方向能夠拆為兩大方向，分別是技術面和實務面

技術面：

1. 集成學習堆疊法能夠使用除了 LSTM 以外的演算法模型，像是：隨機森林樹、支撐向量機等
2. 能加入其他新穎的技術來拆解數據，像是：傅立葉轉換等
3. 資料不平衡處理，若能增加硬體效能或是使用其他大數據框架去做處理，可以嘗試像是 SMOTE-ENN、SMOTE-Tomek Links 等運算較複雜的方法

實務面：

1. 可以增加不同特徵，像是流動性、消息面等
2. 資產配置權重可以進一步去做設計
3. 可以考慮納入放空或是與其他交易市場做資產組合和避險

本研究透過 Open API 取得每日即時資料，以第三方角度提出這套當沖交易的框架，為機器人理財可用服務製作一個模板，而本研究也發現目前 Open API 可用的項目偏少、聯絡體系系統還不夠成熟、API 連線會有不穩的問題等，未來我國政府要持續發展開放銀行，這部分可能還有能夠改善的空間。在這個越來越多人接觸金融、進入股市的時代，本研究期許讓當沖不再是一個讓人覺得是高門檻的交易手段，能夠透過機器人理財降低進入門檻，透過群體的力量，減少個人投資風險。

參考文獻

中文文獻：

KGI Inside (2019)。金融科技現況與發展趨勢。

KPMG (2020)。台灣金融科技發展 100 個關鍵問題。

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。國立政治大學金融學系碩士學位論文。

台灣金融研訓院 (2020)。2020 金融趨勢關鍵議題。

張明珠 (2019)。淺談機器人理財在台灣未來之發展。證券暨期貨月刊，37(1)，14 - 25。

施育霖 (2019)。以機器學習方法預測股價：以台股金融類股為案例。國立中興大學資訊管理學系碩士學位論文。

李偉銘、吳淑貞、黃啟泰 (2015)。總體經濟變數對臺灣股市之大盤及類股熊市預測表現之探討。經濟研究，51(2)，171 - 224。

林秀柔 (2020)。機構投資人持股變化與台股市場異常現象之關係。國立屏東大學財務金融學系碩士學位論文。

洪振虔 (2011)。交易量和報酬之關係與交易策略。中山管理評論，19(2)，305 - 342。

王克修 (2018)。台股漲升趨勢起始判定之實證分析。國立中央大學財務金融學系碩士學位論文。

童韋傑 (2020)。配對交易結合 SVM 於台股市場之應用研究。國立臺北大學統計學系碩士學位論文。

蘇偉榮 (2018)。台股媒體效果對企業社會責任投資組合績效之影響。國立宜蘭大學應用經濟與管理學系碩士論文。

謝文良、林苑宜 (2012)。台灣股市之流動性共變現象。證券市場發展季刊，24(4)，135 - 186。

財團法人中華民國證券暨期貨市場發展基金會 (2018)。我國發展機器人理財顧問之研究。

金融監督管理委員會 (2020)。金融科技發展路徑圖。

陳家蓁(2020)。GA_ANN 模型對總體經濟因素與技術指標預測股市指數績效之研究。國立臺北大學企業管理學系博士論文。

陳彥安(2019)。深度學習於台股動態投資組合之應用。國立交通大學管理學院財務金融學程碩士論文。

英文文獻：

A. Kong, Zhu, H., & Azencott, R. (2020). Predicting intraday jumps in stock prices using liquidity measures and technical indicators Predicting intraday jumps in stock prices using liquidity measures and technical indicators. *Journal of Forecasting*, July, 1–23.

Beketov, M., Lehmann, K., & Wittke, M. (2018). Robo Advisors : quantitative methods inside the robots. *Journal of Asset Management*, 19(6), 363–370.

Best, M. J., & Grauer, R. R. (1991). On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *The Review of Financial Studies*, 4(2), 315–342.

BIS. (2019). *Basel Committee on Banking Supervision Report on open banking and application* (Issue November).

Black, F., & Litterman, R. (1992). Global Portfolio Optimization. *Financial Analysts Journal*, 48(5), 28–43.

Chawla, N.V., Chawla, N.V., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE:synthetic minority over-sampling technique. *W. Philip Kegelmeyer*, 16, 321–357.

H. Chin-Sheng, L. Y.-S. (2019). Machine Learning on Stock Price Movement Forecast: The Sample of the Taiwan Stock Exchange. *International Journal of Economics and Financial Issues*, 9(2), 189–201.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *Studies in Computational Intelligence*, 21(9).

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.

K. Raesita, & Mahadwartha, P. A. (2020). Superstitions and Price Clustering in the

Taiwan Stock Exchange. *17th International Symposium on Management (INSYMA 2020)*, 115, 483–488.

KPMG. (2020). *Pulse of Fintech H1 2020* (Issue September).

Lai, C. Y., Chen, R., &Caraka, R. E. K. O. (2019). PREDICTION STOCK PRICE BASED ON DIFFERENT INDEX FACTORS USING LSTM. *International Conference on Machine Learning and Cybernetics (ICMLC)*.

Li, H., Shen, Y., &Zhu, Y. (2018). Stock Price Prediction Using Attention-based Multi-Input. *Proceedings of Machine Learning Research*, 95, 454–469.

L. Xiaotao, &L. Kin Keung (2017). Intraday volume percentages forecasting using a dynamic SVM-based approach. *Journal of Systems Science and Complexity*, 30(2), 421–433.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91.

Michaud, R. O. (1989). The Markowitz Optimization Enigma : Is “Optimized” Optimal ? . *SSRN Electronic Journal*, 45(1), 31–42.

Obeidat, S., Shapiro, D., Lemay, M., MacPherson, M. K., &Bolic, M. (2018). Adaptive Portfolio Asset Allocation Optimization with Deep Learning. *International Journal on Advances in Intelligent Systems*, 11(1), 25–34.

Premchand, A., &Choudhry, A. (2018). Open Banking & APIs for Transformation in Banking. *International Conference on Communication, Computing & Internet of Things (IC3IoT)* .

PwC. (2020). *Global Fintech 2019*. <https://www.pwc.co.za/en/press-room/global-fintech-2019.html>

Rosdyana Mangir Irawan Kusuma, H. Trang-Thi, K. Wei-Chun, O. Y.-Y. and H. K.-L. (2019). Using deep learning neural networks and candlestick chart representation to predict stock market. *ArXiv*.

Semko, R. (2020). Machine learning for robo-advisors : testing for neurons specialization. *Investment Management and Financial Innovations*, 10(2019), 205–214.

Shen, J., &Shafiq, M. O. (2020). Short - term stock market price trend prediction using a comprehensive deep learning system. *Journal of Big Data*, 7(66).

Wadhwani, P., &Kasnale, S. (2020). *Digital Banking Market Size By Type (Retail*

Banking, Corporate Banking, Investment Banking), By Service (Transactional [Cash Deposits & Withdrawals, Fund Transfers, Auto-Debit/Auto-Credit Services, Loans], Non-Transactional [Information Security, Risk Man.

Wewege, L., Lee, J., & Michael C. Thomsett. (2020). The Digital Banking Transformation : Disruption, Synergy toward Fintech Frontier. *Centre for Financial and Management Studies*, 25.

World Bank Group. (2018). The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution. In *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*.

Zachariadis, M., & Ozcan, P. (2017). The API Economy and Digital Transformation in Financial Services: The Case of Open Banking. SSRN Electronic Journal.

網路資源：

Khanna, S., & Martins, H. (2018). Six digital growth strategies for banks. Mckinsey. , 2020 年 10 月 28 日取自網址：
<https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/six-digital-growth-strategies-for-banks>

Oinkina. (2015). The repeating module in an LSTM contains four interacting layers. Colah's Blog. 2020 年 11 月 2 日取自網址：
<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Preeti Wadhwani, S. K. (2020). Digital Banking Market Size By Type (Retail Banking, Corporate Banking, Investment Banking), By Service (Transactional [Cash Deposits & Withdrawals, Fund Transfers, Auto-Debit/Auto-Credit Services, Loans], Non-Transactional [Information Security, Risk Man. Global Market Insights. 2020 年 11 月 2 日取自網址：
<https://www.gminsights.com/industry-analysis/digital-banking-market>

Statista. (2020). Robo-Advisors. Statista. 2020 年 11 月 24 日取自網址：
<https://www.statista.com/outlook/337/100/robo-advisors/worldwide#market-revenue>

著作權聲明

論文題目：以深度學習建構台灣股市當沖交易策略之實證研究

論文頁數：52頁

系所組別：財政學系

研究生：何宇軒

指導教授：劉彩卿、汪志勇

畢業年月：110.06

本論文著作權為何宇軒所有，並受中華民國著作權法保護。

