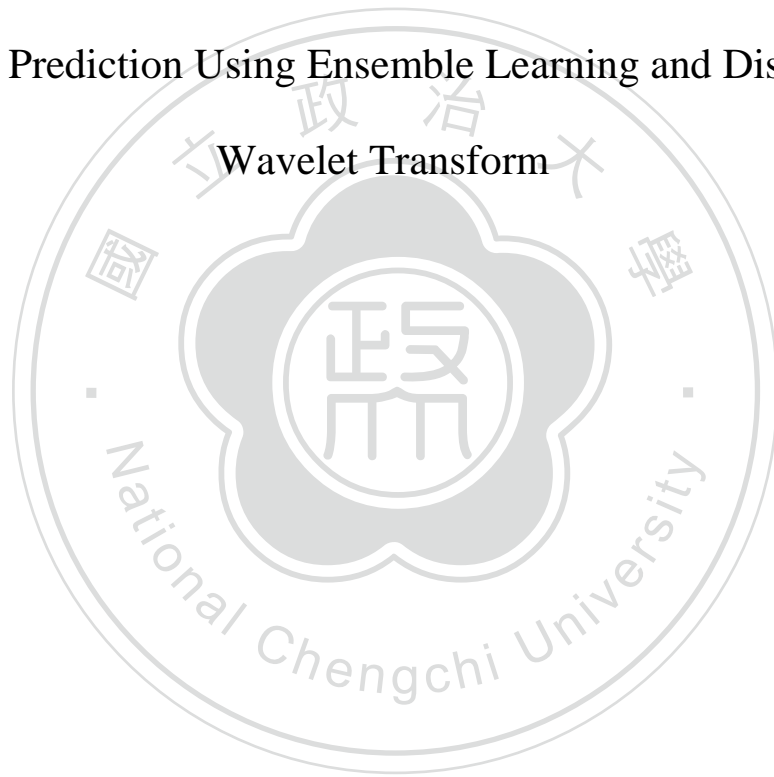


國立政治大學風險管理與保險學系

碩士學位論文

利用集成學習及離散小波轉換進行股票預測

Stock Prediction Using Ensemble Learning and Discrete
Wavelet Transform



指導教授：黃泓智 博士

研究生：張婷媛 撰

中華民國 111 年 7 月

摘要

本研究使用台灣上市公司股票之股價資訊、技術指標以及總體經濟指標以集成學習概念進行台灣股市個股漲跌預測、建立最適投資組合。本論文使用五個不同的機器學習模型：決策樹（Decision Tree）、極限梯度提升模型（XGBoost）、輕量化的梯度提升機（LightGBM）、支持向量機（SVM）以及多層感知器（MLP）進行個股的漲跌預測。為了使模型訓練結果更好，本研究利用集成學習（Ensemble Learning）的堆疊技巧（Stacking），將五個機器學習模型的預測結果整合並進行最終的漲跌預測，選出上漲機率較高的股票，接著組成股票投資清單。另外，本研究第二階段使用離散小波轉換（Discrete Wavelet Transform）去除股票收盤價之雜訊，並當作新的特徵加入模型，重新進行預測。實證結果發現，使用多種模型進行集成學習所建立的投資組合能夠獲得更好的績效，且加入小波轉換技術也有效提升模型的整體績效。

關鍵字：股市漲跌、集成學習、小波轉換、輕量化的梯度提升機、決策樹、極限梯度提升、多層感知器、支持向量機

Abstract

This research uses the stock price information, technical indicators, and macroeconomic indicators to predict the trend of individual stocks in the Taiwan stock market with ensemble learning and establish the optimal investment portfolio. This paper uses five different machine learning models: decision tree, XGBoost, LightGBM, SVM, and MLP. To make the model training results better, this study uses the stacking technique of ensemble learning to integrate the prediction results of five machine learning models and selects the stocks with high rising probability, then make up a stock investment list. In addition, in the second stage of this study, Discrete wavelet transform is used to remove the noise of stock closing price, and it is added to the model as a new feature. The empirical results show that the investment portfolio established using multiple models for ensemble learning can achieve better performance, and adding wavelet transform technology can also effectively improve the model's overall performance.

Keywords : Stock prediction, Ensemble learning, Discrete wavelet transform, Decision tree, XGBoost, LightGBM, SVM, MLP

目次

第一章 緒論.....	1
第一節 研究動機.....	1
第二節 研究目的.....	2
第三節 研究流程.....	3
第二章 文獻回顧.....	5
第一節 離散小波轉換文獻回顧.....	5
第二節 選用指標與股價預測文獻回顧.....	6
第三節 股價預測與機器學習模型文獻回顧.....	7
第四節 集成學習用於投資市場預測文獻回顧.....	8
第三章 研究方法.....	10
第一節 研究架構.....	10
第二節 指標變數選擇.....	12
第三節 離散小波轉換.....	16
第四節 資料預處理.....	17
第五節 機器學習模型.....	20
第六節 集成學習選股.....	23
第七節 績效指標說明.....	27
第四章 實證結果.....	30
第一節 離散小波轉換.....	30
第二節 集成學習.....	39
第三節 最終模型.....	42
第五章 結論與建議.....	49
參考文獻.....	50

表次

表 1、技術指標.....	15
表 2、訊號指標.....	16
表 3、訓練集與測試集區間.....	19
表 4、決策樹回測績效.....	32
表 5、極限梯度提升模型回測績效.....	33
表 6、輕量化的梯度提升機回測績效.....	35
表 7、支持向量機回測績效.....	37
表 8、多層感知器回測績效.....	38
表 9、10 檔股票投組回測績效.....	40
表 10、20 檔股票投組回測績效.....	41
表 11、30 檔股票投組回測績效.....	41
表 12、集成學習回測績效.....	42
表 13、離散小波轉換_10 檔股票投組回測績效.....	43
表 14、離散小波轉換_20 檔股票投組回測績效.....	44
表 15、離散小波轉換_30 檔股票投組回測績效.....	44
表 16、集成學習回測績效比較表.....	45
表 17、集成學習回測績效比較表.....	46
表 18、集成學習_離散小波轉換回測績效.....	48
表 19、最佳績效投資組合比較表.....	48

圖次

圖 1、研究流程.....	4
圖 2、研究架構.....	11
圖 3、離散小波分解示意圖.....	17
圖 4、模型訓練流程.....	23
圖 5、裝袋法示意圖.....	24
圖 6、提升法示意圖.....	25
圖 7、堆疊法示意圖（第一層）.....	26
圖 8、堆疊法示意圖.....	27
圖 9、決策樹回測績效.....	31
圖 10、決策樹_離散小波轉換回測績效.....	31
圖 11、極限梯度提升模型回測績效.....	32
圖 12、極限梯度提升模型_離散小波轉換回測績效.....	33
圖 13、輕量化的梯度提升機回測績效.....	34
圖 14、輕量化的梯度提升機_離散小波轉換回測績效.....	34
圖 15、支持向量機回測績效.....	36
圖 16、支持向量機_離散小波轉換回測績效.....	36
圖 17、多層感知器回測績效.....	37
圖 18、多層感知器_離散小波轉換回測績效.....	38
圖 19、集成學習回測績效.....	40
圖 20、集成學習_離散小波轉換回測績效.....	43
圖 21、集成學習_減少模型_離散小波轉換回測績效.....	46
圖 22、集成學習_離散小波轉換回測績效.....	47

第一章 緒論

第一節 研究動機

隨著科技的發展日益進步、電腦計算能力提升，人工智慧（Artificial Intelligence, AI）的發展越來越普遍、成熟，人工智慧是一種讓系統或電腦設備有模擬人類思考模式、邏輯與行為的能力，且能自行透過數據分析的過程，持續校正、進化。機器學習（Machine Learning, ML）是人工智慧下的一個分支，可處理大量歷史數據，並從資料中學習規律，使用數據進行學習和分析，從而對新樣本做分類或者預測。深度學習（Deep Learning, DL）為機器學習下的分支之一，以多層次的巨大神經網路搭配更進步的訓練技術、計算能力，學習更複雜的大數據，普遍應用在圖像辨識、自然語言處理等領域。

在這個瞬息萬變的金融市場中，若能把握其趨勢，將可得到良好的投資績效。因此，金融市場的許多參與者都想找到方法預測金融市場。許多研究顯示，機器學習和深度學習方法在預測金融市場上，取得了前所未有的亮眼表現。同時利用此方法，在交易上還不需使用太多的人力進行操盤，或是仰賴專業的理財顧問進行資產配置，由於金融業擁有許多金融資產的歷史資料，因此可利用機器學習和深度學習的方法從資料中學習與分析，進而對未來股價進行預測。

綜上所述，本研究欲使用機器學習以及深度學習模型將股價資料、相關技術指標以及總體經濟指標等放入各個模型，進行下個月的股價趨勢預測，其概念就如同投資人在金融市場進行交易時，會觀察股票相關的技術指標以及市場

上總體經濟指標以判斷如何選股、進行交易，本研究將五個模型預測出的上漲結果利用集成學習（Ensemble Learning）的方式進行整合，並做出整合後的最終上漲結果預測，再將預測上漲機率高的股票利用效率前緣的概念建立投資組合，期望能獲得良好的投資績效。

第二節 研究目的

現在許多人都有投資習慣或是投資需求，有些人會花許多時間看盤，尋找最佳時機做短線交易，有些人則是花費較高的手續費，尋求專業理財專員協助投資，不僅耗時、花費較高，甚至有可能投資失利。若能利用前一節提到的機器學習技術推出理財機器人，即可解決耗費時間多、手續費高以及投資績效不理想之問題。因此，本研究希望透過機器學習的技術訓練投資於台灣股市的理財機器人，目標為達到優於大盤的投資績效，期望能給予有投資需求的民眾或是公司，理財規劃方面的協助。

本研究欲達成以下研究目的：

1. 選股指標：利用總體經濟指標如油價、美國公債十年期殖利率，以及股票相關技術指標如簡單移動平均、布林通道指標選股，找出對於預測模型有幫助之指標作為模型輸入特徵。
2. 離散小波轉換技術：利用離散小波轉換技術去除收盤價雜訊，將此新特徵加入模型，探討是否可提升模型預測準確度以提升投資績效。
3. 機器學習模型：利用決策樹（Decision Tree）、極限梯度提升模型（XGBoost）、輕量化的梯度提升機（LightGBM）、支持向量機（SVM）以及多層感知器（MLP），使用前兩點得到之特徵資料，分別進行個股的

漲跌預測並進行回測，探討各模型之預測漲跌能力。

4. 集成學習：利用集成學習的堆疊技巧（Stacking）將各模型預測結果整合，得到每檔股票單一的預測結果，驗證集成學習的投資績效與單一模型的投資績效差異。

第三節 研究流程

本研究依據上述研究動機與研究目的，透過台灣上市公司之股價資訊進行資料處理，並使用台灣加權股價指標之漲跌趨勢作為預測目標。本研究將分為五個章節，個別敘述如下：

一、緒論

針對本論文利用集成學習與離散小波轉換進行股價預測之研究動機與研究目的進行論述。

二、文獻回顧

將本研究相關之文獻資料進行整理與回顧，包含離散小波轉換文獻、選用指標與股價預測文獻、股價預測與機器學習模型文獻以及集成學習用於投資市場預測文獻，並從上述文獻摘取對本研究有幫助之部分進行研究。

三、研究方法

介紹本研究所需技術指標與總體經濟指標、離散小波轉換概念、資料預處理方式、機器學習模型、如何運用集成學習進行股票預測以及績效指標說明。

四、實證結果

比較單一模型與集成學習模型的投資績效並進行分析，驗證運用離散小波轉換去除收盤價雜訊是否可增進模型的投資績效。

五、結論與建議

針對實證結果進行探討，探討實證過程未來可改善的部分與未來發展之方向。

研究流程圖如圖 1 所示：

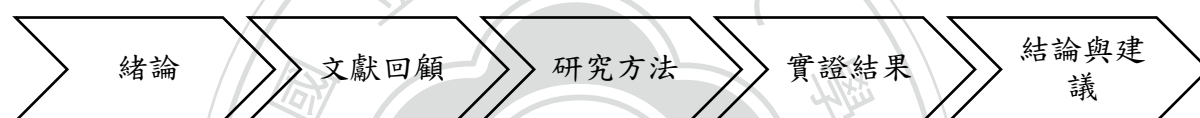


圖 1、研究流程

第二章 文獻回顧

本研究使用總體經濟指標以及技術指標，並運用離散小波轉換技術去除股價收盤價之雜訊，利用集成學習，將機器學習模型預測的股價上漲趨勢進行整合預測，本章節將會分成以下四個部分：第一節為離散小波轉換文獻回顧，第二節為選用指標與股價預測文獻回顧，第三節為股價預測與機器學習模型文獻回顧，第四節為集成學習用於投資市場預測文獻回顧。

第一節 離散小波轉換文獻回顧

由於金融數據資料中存在較多的雜訊，偶然因素往往導致股市的短期劇烈波動，偏離原來的正常趨勢，以至於不利於預測股價趨勢，為了解決這個問題，小波轉換降噪方法被廣泛應用於訊號降噪。Liang et al. (2019) 對訓練之數據進行預處理，實驗結果清楚地發現，進行過小波轉換預處理的方法，在預測準確度方面優於傳統方法。Wu et al. (2021) 使用離散小波轉換對股票進行數據降噪，因為對股票資訊進行降噪有助於消除短期隨機事件對於股票連續走勢的影響，結果顯示，降噪後的數據顯示出更穩定的股價趨勢特徵和平滑度。Tang et al. (2021) 表示為了進一步克服現有模型在處理非平穩和非線性特性方面的困難，針對高頻金融時間序列數據，提出了兩種數據降噪方法，包括小波轉換 (Wavelet Transform, WT) 以及奇異譜分析 (Singular Spectrum Analysis, SSA)，由於 WT 和 SSA 可以從原始序列中提取有用的訊息並避免過度擬合，實驗結果顯示，數據降噪可以大大提高數據的穩定性，有效提高預測模型的能力。

綜上所述，離散小波轉換在進行資料預處理上有顯著的貢獻，可消除短期隨機事件對股票整體趨勢的影響，提高股價資料的穩定性並使整體資料更為平滑，進而有效提高模型的預測能力。因此，本研究將收盤價進行離散小波轉換，將分解後的收盤價放入機器學習模型作為輸入特徵，期待能得到較良好的預測結果。

第二節 選用指標與股價預測文獻回顧

在建構模型時，模型的輸入特徵要包含哪些，是一個關鍵的問題。Weng et al. (2018) 提出了一種兩階段的方法，將 23 個總體經濟指標進行分析，探討隱藏在總體經濟指標背後的資訊，是否可用於準確預測未來一個月的美國主要股票和行業指數價格，結果顯示，三個評估指標 (RMSE、MAPE 和 MAE) 透過納入隱藏在總體經濟指標中的資訊，例如油價、黃金價格以及消費者物價指數等，通常可以改善 25~50%。Shynkevich et al. (2017) 表示技術指標分析在財務預測的應用已被許多研究人員成功使用，因此本篇研究使用技術指標，如簡單移動平均 (SMA)、順勢指標 (CCI)、相對強弱指數 (RSI)、威廉指標 (Williams %R) 等十個技術指標作為機器學習演算法的輸入特徵，以預測股票價格走勢的未來方向。Patel et al. (2015) 預測印度股票市場股票價格，該研究比較了一些預測模型，如人工神經網路 (ANN)、支持向量機 (SVM)、隨機森林，以及這些模型的兩種特徵輸入方法，輸入特徵的第一種方法使用股票交易數據 (開盤價、最高價、最低價和收盤價) 以及計算十個技術指標，如加權移動平均 (Weighted Moving Average)、動量指標 (Momentum)、平滑異同移動平均線 (Moving Average Convergence & Divergence) 等，而第二種方法著重於將這些技術指標表示為訊號指標，捕捉股價上漲或下跌的訊號，實驗結果顯

示，當這些技術指標表示為訊號指標時，所有預測模型的準確度都有所提高。

綜合以上文獻可發現，不僅是總體經濟指標有助於預測未來股價，技術指標以及技術指標所生成的訊號指標作為機器學習模型的輸入特徵，皆有助於預測未來的股價趨勢。因此，本文的模型輸入特徵除了基本的價量資料外，還會放入總經指標、技術指標以及訊號指標作為模型的訓練資料。

第三節 股價預測與機器學習模型文獻回顧

在進行研究時，機器學習模型的選用是一個重要的決策，因此本節將探討在股價預測領域的機器學習模型文獻。Chang (2011) 運用了三種模型，人工神經網路 (ANN)、決策樹 (Decision Tree) 以及人工神經網路和決策樹的混合模型 (Hybrid Model) 以預測股票價格，決策樹以樹的結構為模型，能夠提供股票價格預測結果的良好解釋，作者將模型應用於十種不同的股票並比較了股票三種方法衍生的價格預測模型。Hongjoong (2021) 使用機器學習模型進行股票收益預測，在這項研究中，提出了投資組合優化算法，利用極限梯度提升模型 (XGBoost) 進行股票收益預測，並使用平均-變異 (Mean Variance) 於投資組合選擇，作者使用韓國綜合股價指數成分股 2010 年至 2016 年的歷史數據進行實證檢驗，結果顯示提出的算法優於傳統方法。Chen et al. (2020) 將輕量化的梯度提升機 (LightGBM) 算法引入股票價格預測模型並利用最小方差投資組合建立交易策略系統，研究中提到在眾多機器學習算法中，LightGBM 算法具有高速、高精度、高穩定性和低內存空間等優點，可用於迴歸和分類模型。Chhajer et al. (2022) 概述了人工智慧和機器學習作為股票市場中的預測分析工具，並應用三種機器學習技術，包括人工神經網路，支持向量機 (SVM) 和長

短期記憶模型 (LSTM) 進行股市預測。其中，支持向量機 (SVM) 與人工神經網路相比，更快、更有效，且在高維度空間中的預測能力非常強大。

Hiransha et al. (2018) 使用了四種類型的深度學習模型架構，即多層感知器 (MLP)、遞迴神經網路 (RNN)、長短期記憶 (LSTM) 和卷積神經網路 (CNN) 用於預測股票價格，研究中提到，多層感知器是神經網路的一個特例，在此研究中與其他模型相比，具有良好的預測能力。

綜上所述，現今已有許多機器學習模型應用於股價預測，其中，決策樹 (Decision Tree)、極限梯度提升模型 (XGBoost)、輕量化的梯度提升機 (LightGBM)、支持向量機 (SVM) 與多層感知器 (MLP) 在股價預測上，皆有不錯的預測結果，因此本研究擬採用上述五個模型進行股價預測。

第四節 集成學習用於投資市場預測文獻回顧

使用機器學習技術進行股市預測可以提供更好的預測與更高的準確性，集成學習則是將許多機器學習模型結合，以此來獲得比單一模型更好的預測結果。Nti et al. (2020) 比較不同集成學習的方法：提升 (Boosting)、裝袋 (Bagging)、混合 (Blending) 和堆疊 (Stacking)，研究結果顯示股市方向預測領域應使用集成技術，可獲得比單一模型更好的結果，且在四個集成學習的方法中，堆疊法的預測準確度最高。Jiang et al. (2020) 將技術指標以及總體經濟指標放入各個機器學習模型中進行美國股價指數的漲跌預測，並使用集成學習的堆疊技巧 (Stacking) 進行整合，整合模型分別使用羅吉斯迴歸 (Logistic Regression) 以及正規化迴歸 (Regularized Regression) 中的 Lasso 迴歸 (The Least Absolute Shrinkage and Selection Operator)，研究結果顯示使用集成學習

和單一模型相比，可有效增加預測結果的準確度。Padhi et al. (2021) 研究了六種模型，即 XGBoost、AdaBoost、Gradient Boosting、LightGBM、CatBoost 和 Histogram-based Gradient Boosting，將技術指標作為輸入特徵並利用這些模型使用集成學習的堆疊 (Stacking) 技術來預測股票市場的方向，模型中使用了交叉驗證技術以減少過度擬合，研究結果顯示，元分類器 Meta-LightGBM 具有訓練和測試準確性。Ye et al. (2022) 提出了一種新穎的深度集成學習模型，利用價格數據、技術指標和情緒指數來預測比特幣未來 30 分鐘的價格，該模型利用集成學習的堆疊技術 (Stacking)，結合了長短期記憶模型 (LSTM) 和 Gated Recurrent Unit (GRU) 兩種神經網路模型以提高預測的準確性，實驗結果表示集成方法比其他傳統模型具有更好的預測能力，可以更好地幫助投資者做出正確的投資決策。

綜上所述，使用集成學習技術在預測投資市場上，可獲得比單一模型預測更好的結果，且在集成學習的四個方法中，堆疊技術 (Stacking) 的預測準確度最高，因此本研究將使用集成學習的堆疊技術 (Stacking)，且整合模型使用羅吉斯迴歸 (Logistic Regression)，期待能獲得比單一模型更良好的預測結果。

第三章 研究方法

研究方法共分為七節，第一節為研究架構，將對本論文之研究流程進行說明；第二節為變數選擇，將對使用之總體經濟指標以及技術指標進行介紹；第三節為離散小波轉換，介紹離散小波轉換拆解訊號概念；第四節為資料預處理，將說明資料期間與特徵放入模型前的處理過程；第五節為機器學習模型，本論文使用決策樹、極限梯度提升模型、輕量化的梯度提升機、支持向量機與多層感知器進行預測，本節將介紹上述模型訓練過程；第六節為集成學習選股，將說明多種模型訓練完成後如何使用堆疊技巧合併模型進而預測結果並進行選股；第七節將說明本論文衡量投資組合表現之績效指標。

第一節 研究架構

本研究之研究架構如圖 2 所示，以台灣上市股票作為研究標的並從資料庫中獲取交易日之開盤價、收盤價、最低價、最高價以及成交量，加入第二節介紹的技術指標以及總經指標作為模型訓練資料的特徵，並將收盤價運用第三節提到的離散小波轉換技術將股價中的雜訊去除，最後將上述參數進行標準化並放入模型訓練。

第四節說明本論文所採用的訓練期間與預測期間以及使用五個機器學習模型如決策樹 (Decision Tree)、極限梯度提升模型 (XGBoost)、輕量化的梯度提升機 (LightGBM)、支持向量機模型 (SVM) 以及多層感知器 (MLP) 分別進行個股漲跌預測並進行回測，探討各模型在台灣股市上的預測能力以及績效。

再來將五個機器學習模型的個股漲跌預測結果進行合併，合併方法如第五節介紹之集成學習（Ensemble Learning）中的堆疊技巧（Stacking），本研究在堆疊技巧中所使用的最終模型（Meta Model）為羅吉斯迴歸（Logistic Regression）以整合預測每檔股票下個月的最終上漲機率。

接著將上漲機率高的股票放進投資組合清單內，並採用 Markowitz（1952）提出之投資理論概念進行資產配置，本論文將使用切線投資組合（Tangency Portfolio）以及等權重的方法進行資產配置，最後在回測的部分，本研究採用多種不同之績效指標：累積報酬率、年化報酬率、平均報酬率、年化標準差、年化夏普比率以及最大回撤率，檢驗不同模型下之績效表現。

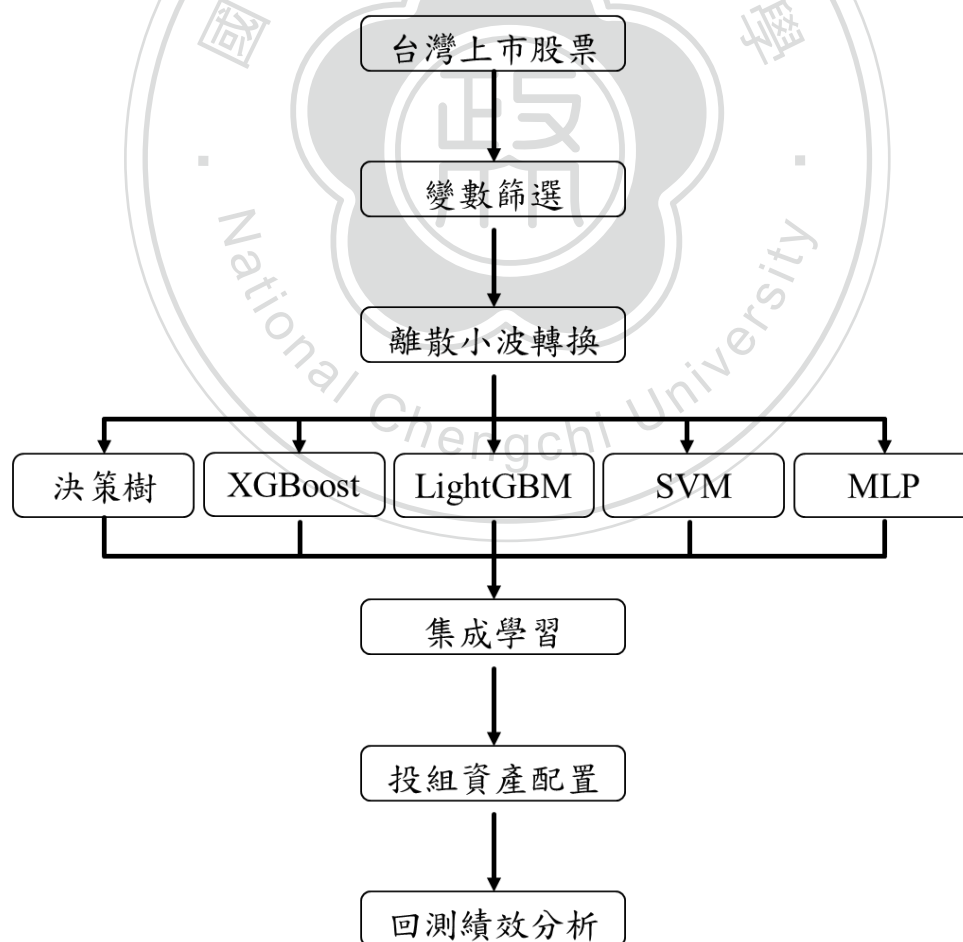


圖 2、研究架構

第二節 指標變數選擇

一、總體經濟指標變數

總體經濟指標是將一個國家的各項經濟活動，以統計數據反映出一個國家的經濟狀況。投資人透過觀察總體經濟指標，就能大致掌握整體大盤走勢，以及經濟活動循環狀態。本研究根據文獻回顧觀察到許多與股價走勢相關的總體經濟指標，因此從 TEJ 台灣經濟新報資料庫下載台灣加權量發行股價指數、美國道瓊工業指數、黃金價格、油價、消費者物價指數、美國公債一年期殖利率以及美國公債十年期殖利率共七項總經指標，做為放入模型之總經變數，以下分別介紹各指標。

I. 台灣加權量發行股價指數

台灣加權量發行股價指數的主要用途為衡量台灣所有上市公司個股的整體績效表現，更是證交所第一檔自行編製的指數，在實務上，台灣加權股價指數的走勢變化也會作為國內景氣的風向球，當股市快速成長時，代表投資人對上市公司的未來表現感到樂觀，通常處於國內景氣復甦或擴張期，反之亦然。

II. 美國道瓊工業指數

美國道瓊斯工業指數（DJI），由 Charles H. Dow 和 Edward Jones 創立於 1896 年，是歷史相當悠久的一項股價指數，也是我們對股市較為熟知的一種指數。道瓊指數在百年前最早時只有十二檔，在當年都是美國舉足輕重的企業，包含鐵路、棉花、天然氣、糖、煙草和石油等產業，由於這類工業是當時經濟的火車頭，再加上當時也沒有其他更好的指數來衡量整體市場表現，因此就一直被當成總體市場成績的指標。目前成分股則是包含了三十家上市公司。

III. 黃金價格

黃金一般被認為是貨幣避險機制，當經濟景氣良好時，投資者更願意從股市中投資獲利，所以往往在股市上升時黃金價格下跌，而景氣狀況較差，股市下跌時，金價一般呈上升趨勢。

IV. 油價

石油為世界上最重要的大宗商品，油價升跌會顯著影響經濟環境，亦會影響通貨膨脹壓力，因為油價反映商業經營成本，如機器燃料和交通運輸費用等，若油價高，代表經營成本高，商家便會相應提高產品或服務的價格。反之，若油價低，即經營成本降低，物價會相應降低。油價低可降低企業經營成本，理應有利股市表現，如航空業、汽車及運輸業可以受惠於燃油支出降低，而零售業、旅遊業等亦會受惠於消費意欲提升。

V. 消費者物價指數

消費者物價指數是各國重要的經濟指標，透過消費者物價指數的年增率可以用來衡量通貨膨脹率，一般來說，消費者物價指數上升代表通貨膨脹率提升，即物價、通貨膨脹率有升溫跡象，央行會採取貨幣緊縮政策，資金較易流出股市，導致股價下跌。當消費者物價指數下滑代表通貨緊縮，即物價、通貨膨脹率有下降跡象，央行可能就會採取貨幣寬鬆政策，熱錢較易流入股市，導致股價上漲，成為一個循環。

VI. 美國公債一年期殖利率

公債殖利率是美國政府債務的投資報酬率，以百分比表示，美國公債殖利率是美國政府在不同時間段內借錢而支付的有效利率。公債殖利率不僅影響政府支付多少借款以及投資者通過購買政府債券賺取多少，也影響個人和企業借錢購買房地產、車輛和設備的利率。

VII. 美國公債十年期殖利率

美國十年期公債殖利率是指從買入十年期公債至到期的投資報酬率，美國公債一般來說，被投資人視為相對安全的資產，一旦市場景氣有所改變，公債殖利率的變化也會影響投資人在股市的決策，當市場景氣復甦，投資人願意承擔更高風險的投資，此時債券價格會下跌、十年期公債殖利率上升，資金流向股市；反之，當市場恐慌，投資人會轉向投資穩定安全的商品，因此資金會流向債券，此時很多人要買債券，導致債券價格因此上漲。

二、技術指標變數

技術指標是指投資者用來做技術分析時，會使用到的指標資料，通常會是由價格、成交量或是其他公開資料組成指標，它能夠用簡化市場的資訊，包含過去價格、市場情緒、過去短期趨勢、平均成本等等訊息，反映在數值或圖面上，作為投資時的額外參考資訊，讓投資者做出投資判斷，本研究使用的變數除了上面介紹的總體經濟指標外，另採用之變數包含以下三類：價量指標、技術指標以及由技術指標衍生之訊號指標，以下分別說明。

- I. 價量資料：由 TEJ 台灣經濟新報資料庫下載開盤價、最高價、收盤價、最低價與當日成交量，為了去除短期投資市場雜訊，以取得長期股市趨勢，因此利用指數平滑 3 日之轉換，轉換公式為：

$$S'_t = a * S_t + (1 - a) * S'_{t-1}$$
$$a = \frac{2}{n+1}, n = 3$$

其中， S_t 為在時點 t 的原始價量， S'_t 為時點 t 指數平滑後之價量。

- II. 技術指標：透過指數平滑後的價量指標計算出技術指標，本論文共採十三種技術指標，如表 1 所示。

SMA	簡單移動平均 (Simple n-days Moving Average)
VSMA	成交量簡單移動平均 (Volume Simple n-days Moving Average)
WMA	加權移動平均 (Weighted Moving Average)
MTM	動量指標 (Momentum)
Aroon	阿隆指標 (Aroon Indicator)
BBands	布林通道指標 (Bollinger Bands)
MACD	平滑異同移動平均線 (Moving Average Convergence & Divergence)
ADX	平均趨向指標 (Average Directional Indicator)
CCI	順勢指標 (Commodity Channel Index)
MFI	資金流向指標 (Money Flow Index)
CHV	蔡金波動指標 (Chaikin Volatility)
RSI	相對強弱指數 (Relative Strength Index)
WPR	威廉指標 (Williams %R)
WAD	威廉多空力度線 (Williams Accumulation/Distribution)

表 1、技術指標

III. 訊號指標：為了捕捉股價趨勢轉折之處，本研究針對不同技術指標特性，設定相應門檻值，以計算買進賣出的訊號指標，使用的訊號指標如表 2 所示。

MTM_Signal	MTM 由下往上突破 0 時為買進訊號，MTM 由上往下突破 0 時為賣出訊號
Aroon_Signal	設定 AroonUp 高於 AroonDn 且 AroonUp 大於 70 時為買進訊號
BBands_Signal	當百分比指標大於 1 時，為買進訊號
MACD_Signal	計算長短期股價之指數平均差異為 DIF，將 DIF 扣除 MACD 為柱線 (OSC)，OSC 由負轉正為買進訊號，由正轉負則為賣出訊號
ADX_Signal	當上升動向指數向上穿越下降動向指數且高於 ADX 時，為買進訊號，當上升動向指數向下穿越下降動向指數且低於

	ADX 時，為買進訊號
CCI_Signal	當 CCI 小於 -100 為超賣訊號
MFI_Signal	當 MFI 大於 80 為超買訊號，MFI 小於 20 為超賣訊號
CHV_Signal	當 CHV 小於 0 時為買進訊號
RSI_Signal	當 RSI 大於 80 為超買訊號，RSI 小於 20 則為超賣訊號
WAD_Signal	當股價下跌且 WAD 上漲時為買進訊號，當股價上漲且 WAD 下跌時為賣出訊號

表 2、訊號指標

第三節 離散小波轉換

台灣股市股價易受經濟發展、政策變化和投資者情緒等大量因素的影響，因此股價趨勢通常包含大量雜訊且為非線性，為了提高機器學習模型的預測能力，應將資料中的雜訊去除。本研究將股價資料視為一段訊號，因此我們使用訊號分解的方法：離散小波轉換；此方法透過伸縮、位移等操作功能，可以對訊號進行多尺度精細化分析，有效消除數據中包含的雜訊並保留原始訊號的特徵。據此，本文擬採用離散小波分解與重構方法對金融時間序列進行數據預處理，並採用小波降噪消除高頻，即時間序列中的雜訊，從而削弱雜訊干擾對神經網路預測能力的影響，提高模型預測能力。

離散小波分解會將輸入訊號分解為低頻訊號和高頻訊號，每次分解僅分解低頻部分，如圖 3 所示。假設 C_0 為原始金融時間序列訊號， C_1 、 C_2 、...、 C_L 是第一個、第二個、...、第 L 層的低頻訊號， D_1 、 D_2 、...、 D_L 是第一個、第二個、...、第 L 層的高頻訊號。

用數學式表示如下：

$$C_0 = C_L + D_L + D_{L-1} + \dots + D_2 + D_1$$

為了降噪，本研究使用哈爾小波轉換（Haar wavelet）分解與重建，將收盤價分解兩次，即 $L=2$ ，並將離散小波分解後的高頻部分設為零再重建。因為小波低頻部分為財務序列數據的總體趨勢，高頻部分為該金融時間序列的短期隨機擾動。因此設置高頻部分歸零可以消除雜訊達到平滑訊號效果。

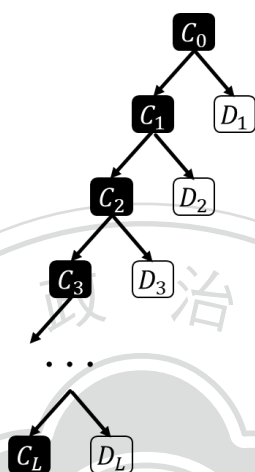


圖 3、離散小波分解示意圖

第四節 資料預處理

本節將說明機器學習模型所使用之特徵、資料放進模型前預處理的過程與資料期間、訓練集與測試集期間。

一、模型特徵與預處理

本研究使用的特徵包含第二節所介紹的價量資料、總經指標、技術指標、訊號指標以及離散小波轉換後的收盤價，機器學習時，不同的特徵值以及資料數值大小對模型訓練會造成影響，為了使特徵數值參考標準一致，在訓練模型之前會先將所使用之特徵進行標準化，標準化方式為將所有特徵數據縮放成平

均為 0、標準差為 1，標準化公式如下：

$$x'_i = \frac{x_i - \mu_x}{\sigma_x}$$

其中， x'_i 代表變數中第 i 個資料經過標準化後的值， x_i 代表原始變數， μ_x 為該變數的平均值， σ_x 為該變數的標準差。

二、資料期間

本論文使用之原始資料期間為西元 2006 年 7 月至 2021 年 9 月，因為計算技術指標時會運用到落後期，股價資料較前面的日期無法計算出所有的技術指標，因此 2006 年 7 月至 10 月的資料不參與模型訓練的過程，另外，因經濟景氣循環約為十年，故本研究使用前十年之資料預測下個月第一個交易日股價之漲跌，即訓練集長度為十年，由於模型是預測個股在下個月的第一個交易日的漲跌機率，因此取每個月的第一個交易日，作為模型的訓練資料。舉例來說，第一組訓練區間為 2006 年 11 月至 2016 年 10 月，第一個再平衡日為 2016 年 11 月的第一個交易日，第二組訓練區間為 2006 年 12 月至 2016 年 11 月，第二個再平衡日為 2016 年 12 月的第一個交易日，依此類推。

在將訓練資料放進模型前的最後一個步驟為流動性篩選，由於本研究會以模型的預測結果作為交易依據，為了確保投資者在股票市場可以購買到相對應張數的股票，因此本研究以交易量前 50% 為流動性篩選標準，透過此篩選標準能夠更有效的將流動性較差的股票刪去。值得注意的是，此交易量为經過簡單移動平均 120 日後的交易量，因此 2006 年 7 月至 2021 年 12 月的資料無法計算，以至於無法放入模型訓練，因此最終模型採用的資料期間為 2007 年 1 月至 2021 年 9 月，前兩組訓練區間皆調整為 2007 年 1 月開始，其餘訓練區間不變，詳細訓練集及與測試集如表 3 所示：

訓練集	測試集	訓練集	測試集
20070102~20161003	20161101	20090401~20190304	20190401
20070102~20161101	20161201	20090504~20190401	20190502
20070102~20161201	20170103	20090601~20190502	20190603
20070201~20170103	20170202	20090701~20190603	20190701
20070301~20170202	20170301	20090803~20190701	20190801
20070402~20170301	20170405	20090901~20190801	20190902
20070502~20170405	20170502	20091001~20190902	20191001
20070601~20170502	20170601	20091102~20101001	20191101
20070702~20170601	20170703	20091201~20191101	20191202
20070801~20170703	20170801	20100104~20191202	20200102
20070903~20170801	20170901	20100201~20200102	20200203
20071001~20170901	20171002	20100301~20200203	20200302
20071101~20171002	20171101	20100401~20200302	20200401
20071203~20171101	20171201	20100503~20200401	20200504
20080102~20171201	20180102	20100601~20200504	20200601
20080201~20180102	20180201	20100701~20200601	20200701
20080303~20180201	20180301	20100802~20200701	20200803
20080401~20180301	20180402	20100901~20200803	20200901
20080502~20180402	20180502	20101001~20200901	20201005
20080602~20180502	20180601	20101101~20201005	20201102
20080701~20180601	20180702	20101201~20201102	20201201
20080801~20180702	20180801	20110103~20201201	20210104
20080901~20180801	20180903	20110208~20210104	20210201
20081001~20180903	20181001	20110301~20210201	20210302
20081103~20181001	20181101	20110401~20210302	20210401
20081201~20181101	20181203	20110503~20210401	20210503
20090105~20181203	20190102	20110601~20210503	20210601
20090202~20190102	20190211	20110701~20210601	20210701
20090302~20190211	20190304	20110801~20210701	20210802

表 3、訓練集與測試集區間

第五節 機器學習模型

機器學習是人工智慧的一種，透過演算法並使用大量資料進行訓練，訓練完成後會產生模型，未來當有新的資料，可放入訓練產生的模型進行預測。機器學習模型可分為兩類：監督式學習（Supervised Learning）以及非監督式學習（Unsupervised Learning）；其差別在於監督式學習在訓練的過程中會告訴機器正確答案，非監督式學習在訓練資料時則沒有標準答案，故機器在學習時並不知道其分類結果是否正確。

本研究使用五個機器學習模型，皆為監督式學習模型，分別為決策樹（Decision Tree）、極限梯度提升模型（XGBoost）、輕量化的梯度提升機（LightGBM）、支持向量機（SVM）以及多層感知器（MLP）。以下將詳細介紹各個模型的特色以及建模流程。

一、模型介紹

I. 決策樹（Decision tree）

決策樹會根據訓練資料產生一棵樹，依據訓練出來的規則來對新樣本進行預測。決策樹屬於監督式學習，可用於解決迴歸和分類問題。使用決策樹的目的是創建一個訓練模型，該模型可以透過學習從先前數據（訓練數據）推斷出的簡單決策規則來預測目標變量的類別或值。一開始會先從根節點開始，然後依據各個特徵將資料作分割到左右兩邊。為了能在節點上，使用最具意義的特徵來做分割，需要透過訊息增益（Information Gain）來判斷。訊息增益是特徵選擇的一個重要指標，它定義為一個特徵能夠為分類系統帶來多少訊息，帶來的訊息越多，說明該特徵越重要，相應的訊息增益也就越大，除了訊息增益以外，研究中也常用獲利比率（Gain Ratio）或吉尼係數（Gini Index）作為分割的準則。在決策樹中，我們對新資料的預測方法如下：從根開始，將根所記錄

的特徵與新樣本的特徵進行比較，根據其結果跳至下一個節點，重複上述過程直到結束，即可獲得預測結果。

II. 極限梯度提升模型 (XGBoost)

極限梯度提升模型全名為 eXtreme Gradient Boosting，是目前 Kaggle 競賽中最常見到的算法，同時也是多數得獎者所使用的模型。極限梯度提升模型最早由華盛頓大學博士生陳天奇提出，是一種基於改進殘差的梯度提升方法，極限梯度提升模型首先會先訓練出第一個模型，這個最初的模型以訓練資料的實際值作為目標值，目標式是減少實際值和預測值的殘差。接著我們把前一個訓練好的模型的預測值和實際值的差做為第二個模型的目标值，第二個模型用來彌補第一個模型預測不足的部分。依序用剩下的殘差建立多個模型後，將這些模型組合起來，把輸入值在所有模型得到的輸出值加總作為最終預測值。

III. 輕量化的梯度提升機 (LightGBM)

前一段所提出的極限梯度提升模型是個常用在 Kaggle 競賽的方法，但仍有計算量大、空間複雜度高等缺點，造成在實際的應用上受到了一定限制，畢竟在現今的使用情境，所用的資料量很龐大。因此，為了改善極限梯度提升模型，Ke Guolin, et al.發表了一個新的演算法，稱為輕量化的梯度提升機，全名為 (Light Gradient Boosting Machine)。其優勢為更快的訓練效率、低記憶體使用、更好的準確度、支援平行運算以及可處理大規模數據。

IV. 支持向量機 (SVM)

支持向量機，全名 Support Vector Machine，是一種相對簡單的監督式學習，用於分類或迴歸問題。支持向量機的原理為，用統計風險最小化的原則來找到一個在數據類型之間創建邊界的超平面，目標是能夠將資料分得越開越好。在支持向量機中，我們將數據集中的每個數據點繪製在 N 維空間中，其中

N 是數據中特徵的數量，接著找到最適當的超平面來將數據完美分離，如此一來，即可透過此超平面對新數據進行預測。

V. 多層感知器 (MLP)

多層感知器，全名 Multilayer Perceptron，是一種深度人工神經網路。它由多個感知器組成，它們由接收訊號的輸入層、對輸入做出決策或預測的輸出層以及在這兩者之間的任意數量的隱藏層組成，這些隱藏層是多層感知器真正的計算引擎，接下來透過激活函數 (Activation Function) 進行轉換後再經過線性合成後進入輸出層。多層感知器通常應用於監督式學習問題，它們在一組輸入、輸出對上進行訓練，並學習對這些輸入和輸出之間的相關性進行建模。訓練涉及調整模型的參數、權重和偏差，以最大限度地減少誤差。神經網路的流程分為前向傳播和反向傳播，在前向傳播中，訊號流從輸入層通過隱藏層移動到輸出層，反向傳播用於對誤差進行權重和偏差調整，當多層感知器的參數更接近誤差最小值時，可以沿著該梯度調整參數，例如隨機梯度下降法，直到誤差不能再低。

二、建模流程

本研究使用各模型預測每檔股票下個交易日之股價漲跌，將報酬率分為三類，分別為上漲、不漲不跌、下跌，將報酬率分為三類可以減少模型對於上漲下跌的誤判，報酬率 5% 以上被分類為上漲，報酬率介於 -5% ~ 5% 之間被分類為不漲不跌，而下跌 5% 以上則被分類為下跌，模型經訓練後，可在交易日當天預測出每檔股票在未來一個月的上漲機率、不漲不跌機率以及下跌機率，因此可利用此預測結果，將上漲機率最高的前 10 檔、20 檔、30 檔股票建構成三個投資組合清單，並利用平均-變異 (Mean Variance) 最適投資組合將投資組合清單裡的股票配置權重進行交易及回測，平均-變異最適投資組合，又稱為相

切組合（Tangency Portfolio），為美國經濟學家 Markowitz（1952）提出的效率前緣（Efficient Frontier）與資本市場線相切位置之投資組合，為夏普比率最高之投資組合，模型訓練流程如圖 4 所示。

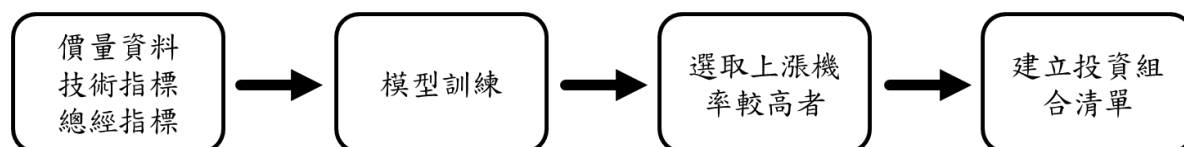


圖 4、模型訓練流程

第六節 集成學習選股

本節將介紹如何運用集成學習的方法將前一節介紹的機器學習模型結合，集成學習指的是以一個系統化的方式將好幾個監督式學習的模型結合在一起，目的是希望結合眾多的模型產生一個更強大的模型，集成學習在實務上對於提升模型預測準確率非常有效。在集成學習領域主要有三種方法，分別是裝袋（Bagging）、提升（Boosting）和堆疊（Stacking），以下分別介紹。

一、裝袋（Bagging）

裝袋，也稱為引導聚合（Bootstrap Aggregation），是一種集成學習方法，每次從訓練集中隨機抽取 n 個訓練樣本（在訓練集中，有些樣本可能被多次抽取到，而有些樣本可能一次都沒有被抽中），共進行 k 輪抽取，得到 k 個訓練集，利用每個訓練集獨立訓練模型，共得到 k 個模型，並根據預測的類型，例如迴歸或分類，再將這些不同的弱模型預測結果取平均值或投票，以產生更準確的預測結果，示意圖如圖 5。隨機森林（Random Forest）演算法被認為是裝

袋方法的例子，它使用裝袋和特徵隨機性來創建不相關的決策樹森林。

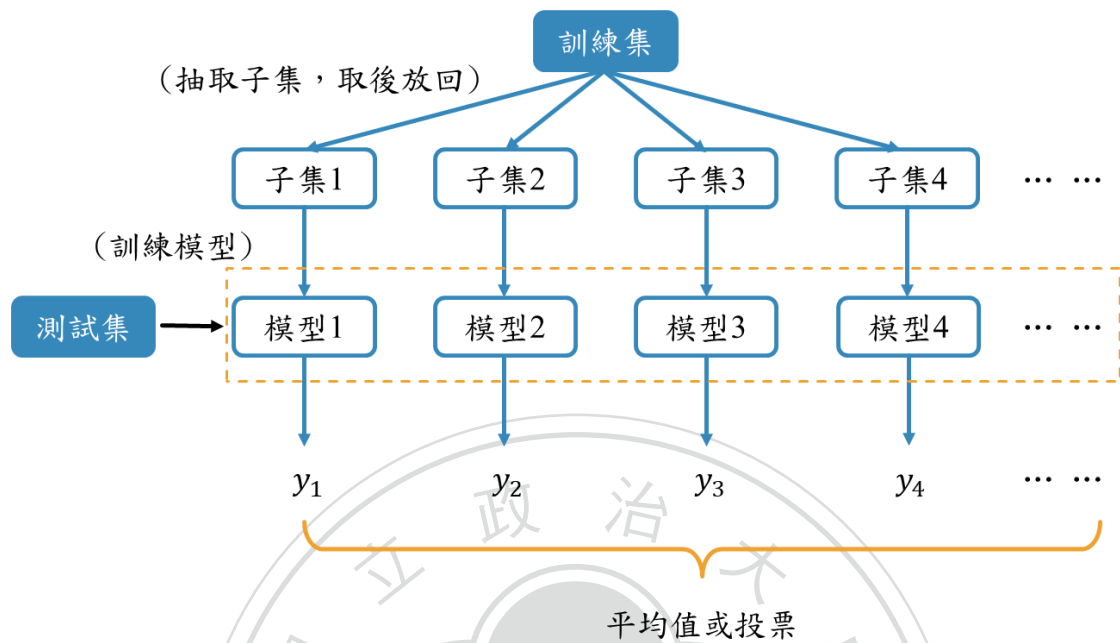


圖 5、裝袋法示意圖

二、提升 (Boosting)

提升是一種集成學習方法，它將一組弱學習器組合成一個強學習器，以最大限度地減少訓練錯誤。在提升法中，隨機選擇訓練樣本訓練模型，然後按順序訓練，和裝袋不同的是，子模型之間是有關聯性的，是透過將前一個子模型的錯誤資料權重提高，然後再訓練新的模型，每個模型都試圖彌補其前身的弱點進而提升分類結果，子模型表現錯誤率愈小，權重愈大，如圖 6 所示，常見的提升算法例子如 AdaBoost、XGBoost。

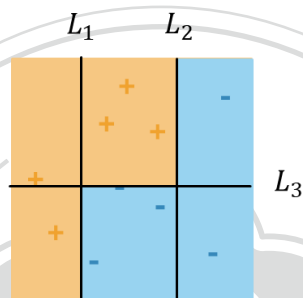
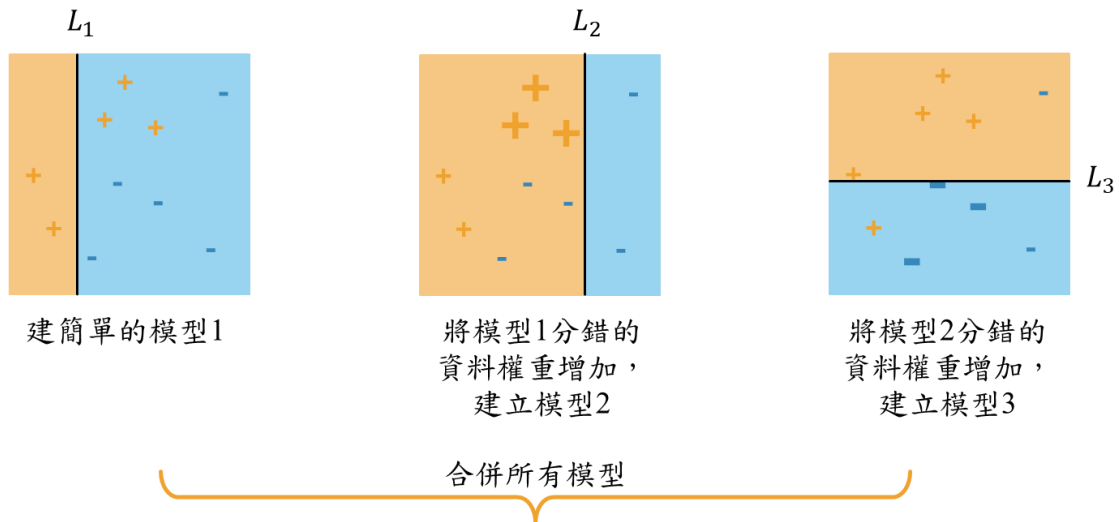


圖 6、提升法示意圖

三、堆疊 (Stacking)

堆疊法與裝袋法和提升法主要有兩個差異：

- I. 裝袋法和提升法通常是將同質的子模型組合，而堆疊法則是通常考慮異質的子模型組合。
- II. 堆疊法會利用第二階模型來組合基礎模型，而裝袋法和提升法則根據確定性算法組合子模型。

基礎模型的訓練如同前一節機器學習模型介紹所述，因此這裡主要介紹的為如何訓練第二階模型藉以得到最終預測結果，在訓練初始模型時，會搭配交叉驗證 (Cross-Validation) 技巧，這裡使用的為 K-fold 交叉驗證法，將訓練資料 (Train Data) 切割成 K 個子集資料，如圖 7 所示，假設將訓練資料分成三個

子集，藍色部分會用來訓練模型，橘色部分會拿來預測並取得預測值，稱為 Predict_X，或是 Meta-X，此外，再將原始的測試資料（Test Data）丟入模型預測後會得到 Predict（圖中橘框白底部分），但由於我們一開始把資料分成三份，所以會有三個模型，也會得到三組測試資料的預測結果，這時就直接用平均法或投票法，將這三組預測結果整合起來，變成 Predict_Y，又被稱作 Meta-Y，當擁有多個模型時，我們必須各自取得每個模型的 Meta-X 跟 Meta-Y，作為最終模型（Meta-Model）的訓練及測試資料集。

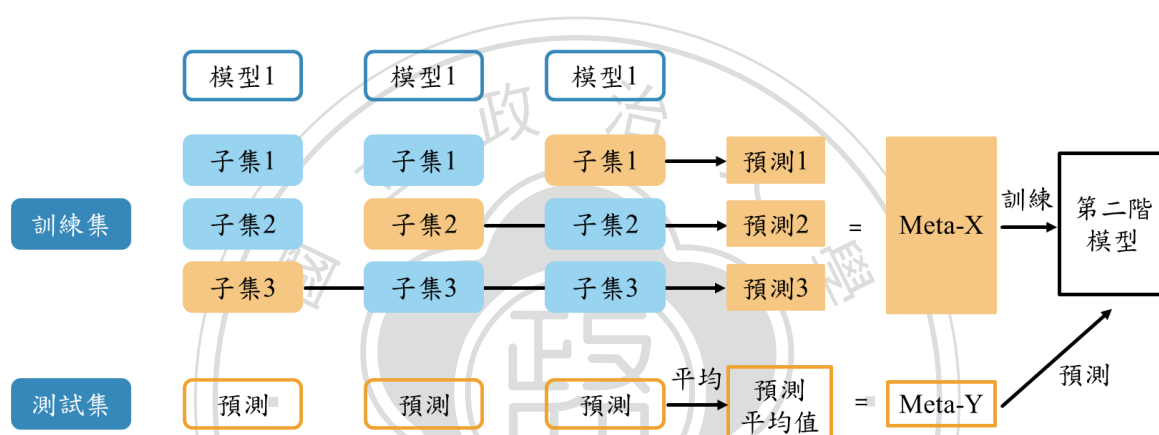


圖 7、堆疊法示意圖（第一層）

由於不同的模型可以學習資料中不同的資訊，每個模型都有其優點與弱點，因此使用堆疊法的好處在於，若我們將不同模型的結果結合再進行第二次的學習，可以弱化每個模型的弱點，將每個模型在表現好的部分保留，同時捨棄各自表現不好的部分，如此一來，將可有效地提高模型預測能力，最後得到一個更好的預測結果。

本論文中使用的集成學習方法為堆疊法（Stacking），在第一階段訓練單一模型時，使用交叉驗證技巧，將訓練資料切割成 3 個子集資料，並且分別訓練五種機器學習模型；決策樹（Decision Tree）、極限梯度提升模型（XGBoost）、輕量化的梯度提升機（LightGBM）、支持向量機（SVM）以及

多層感知器 (MLP)。依據上述介紹的訓練方法，即會得到每個機器學習模型的 Meta-X 跟 Meta-Y，作為最終模型 (Meta-Model) 的訓練及測試資料集，如圖 8 所示。在本研究中，第二階段選用的 Meta-Model 為羅吉斯迴歸 (Logistic Regression)，會預測出每檔股票的上漲機率、不漲不跌機率以及下跌機率，並依上漲機率最高的 10 檔、20 檔、30 檔股票組成三個投資組合清單，並利用平均-變異 (Mean Variance) 最適投資組合將投資組合清單裡的股票配置權重進行交易及回測。

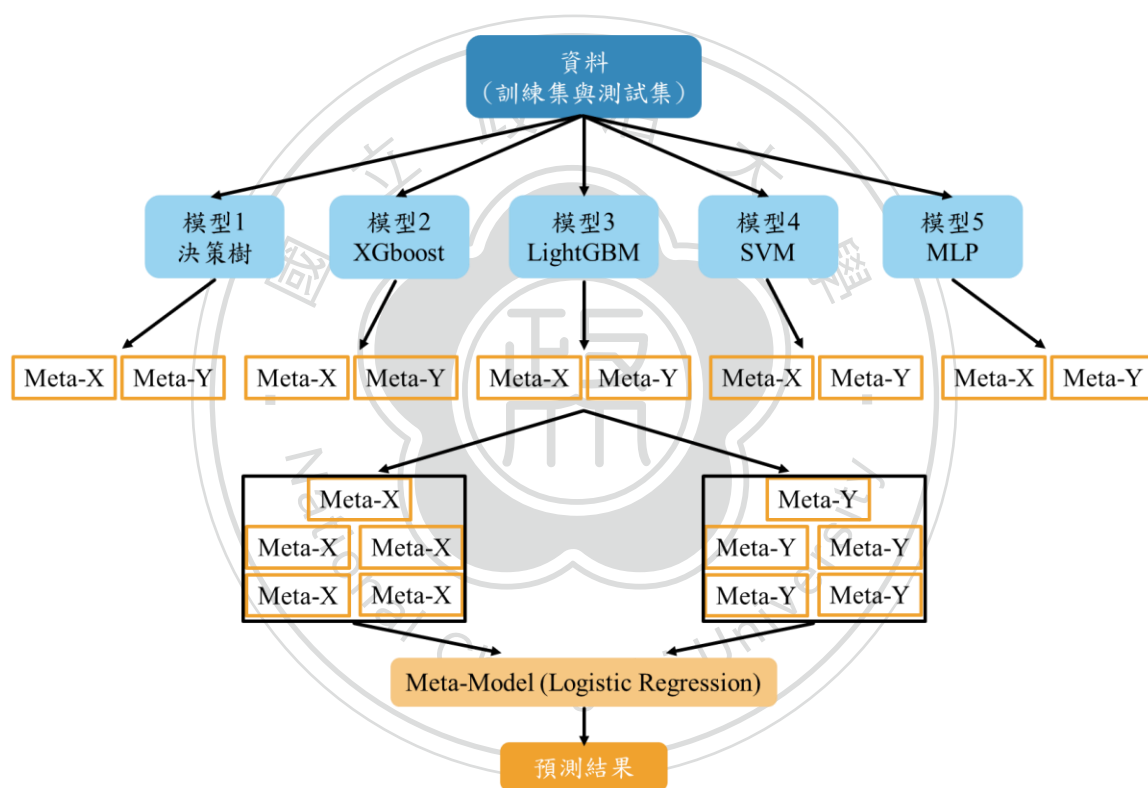


圖 8、堆疊法示意圖

第七節 績效指標說明

本研究中評價各模型的指標分別為累積報酬率、年化報酬率、平均年報酬率、年化標準差、年化夏普比率以及最大回撤率，在本節中，將詳細介紹各績

效指標的計算方式。

一、累積報酬率

$$\text{累積報酬率} = [\prod_{i=1}^k (1 + r_i)] - 1$$

其中 k 為回測期間總交易日數， r_i 為投資組合於第 i 個歷史交易日之日報酬率。

二、年化報酬率

$$\text{年化報酬率} = (1 + \text{累積報酬率})^{\frac{252}{k}} - 1$$

其中 k 為回測期間總交易日數。

三、平均報酬率

$$\text{平均報酬率} = E(R_p) = \bar{r} * 252$$

其中 \bar{r} 為投資組合平均日報酬率。

四、年化標準差

$$\text{年化標準差} = \sigma_p = \sqrt{\sum_{i=1}^k \frac{(r_i - \bar{r})^2}{n-1} * 252}$$

其中 k 為回測期間總交易日數， r_i 為投資組合於第 i 個歷史交易日之日報酬率，

\bar{r} 為投資組合平均日報酬率。

五、年化夏普比率

$$\text{年化夏普比率} = \frac{E(R_p) - R_f}{\sigma_p}$$

其中 $E(R_p)$ 為平均報酬率， R_f 為無風險利率， σ_p 為年化標準差。

六、最大回撤率

$$\text{最大回撤率} = \min\left(\frac{\text{第 } p \text{ 天資產淨值}}{\text{第 } q \text{ 天資產淨值}}\right) - 1$$

其中 p 為小於回測期間總交易日數之任意正整數， q 為小於 p 之任意正整數。



第四章 實證結果

本章將呈現使用第三章介紹之研究方法所得到的實證結果，本章共分為三節，第一節將呈現各個單一模型有無使用離散小波轉換技術，模型預測能力的差異；第二節將呈現在無離散小波轉換的情況下，單一模型與集成學習模型預測能力的比較；第三節將呈現同時使用離散小波轉換技術與集成學習的情況下，模型預測能力的提升。

第一節 離散小波轉換

一、決策樹 (Decision Tree)

圖 9 為決策樹模型在收盤價使用原始收盤價時，前 10 檔、20 檔、30 檔組合而成的投資組合清單累積報酬率表現。圖 10 為決策樹模型在原始收盤價替換成離散小波轉換收盤價時，前 10 檔、20 檔、30 檔組合而成的投資組合清單累積報酬率表現。表 4 為決策樹模型上述六個投資組合之績效。

投資策略回測績效_Ddecision tree
回測期間:2016-11-02 至 2021-09-30

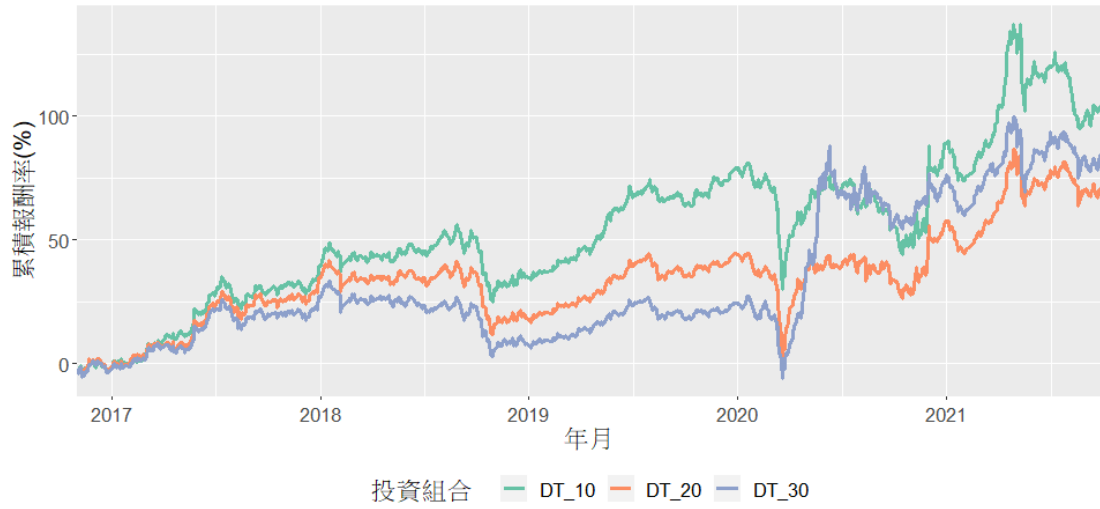


圖 9、決策樹回測績效

投資策略回測績效_DT_Wavelet
回測期間:2016-11-02 至 2021-09-30

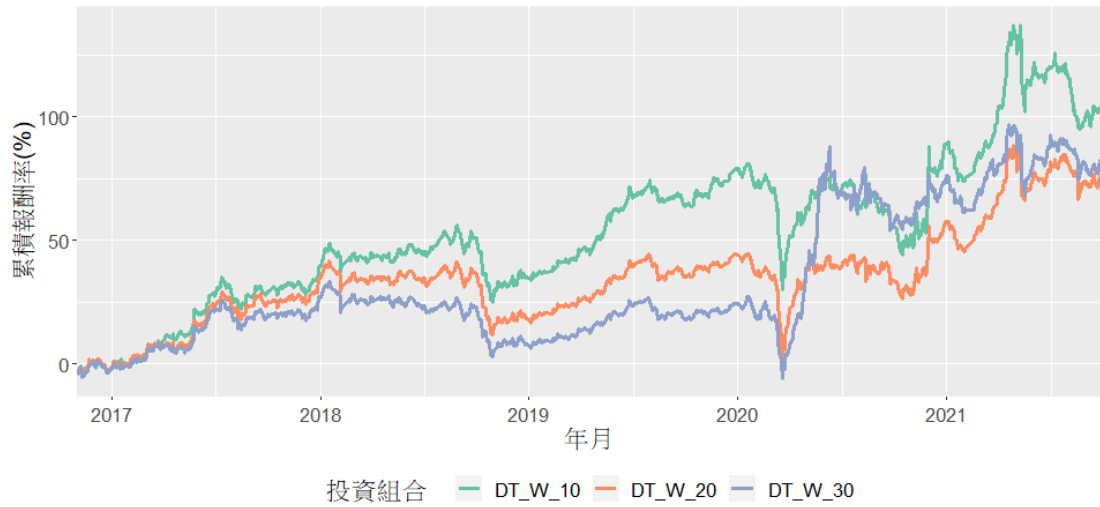


圖 10、決策樹_離散小波轉換回測績效

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
DT_10	101.28%	15.78%	16.54%	0.1971	0.8394	-28.21%
DT_20	69.61%	11.70%	12.69%	0.1805	0.7031	-29.83%
DT_30	82.65%	13.45%	14.30%	0.1828	0.7821	-29.67%
DT_W_10	101.28%	15.78%	16.54%	0.1971	0.8394	-28.21%

DT_W_20	74.32%	12.35%	13.28%	0.1815	0.7318	-29.83%
DT_W_30	80.88%	13.22%	14.08%	0.1819	0.7740	-29.67%

表 4、決策樹回測績效

表 4 中可發現，在決策樹模型中，將收盤價替換為離散小波轉換後的收盤價，在報酬率與波動度的部分皆沒有顯著的差異，挑選股票上漲機率前 20 名的投資組合報酬率有略微提升，而挑選股票上漲機率前 30 名的投資組合報酬率則略微下降，可能是離散小波轉換對於決策樹模型的預測能力沒有顯著的影響。

二、極限梯度提升模型 (XGBoost)

圖 11、圖 12 分別為極限梯度提升模型在收盤價使用原始收盤價、原始收盤價替換成離散小波轉換收盤價時，前 10 檔、20 檔、30 檔組合而成的投資組合清單累積報酬率表現。表 5 為極限梯度提升模型上述六個投資組合之績效。

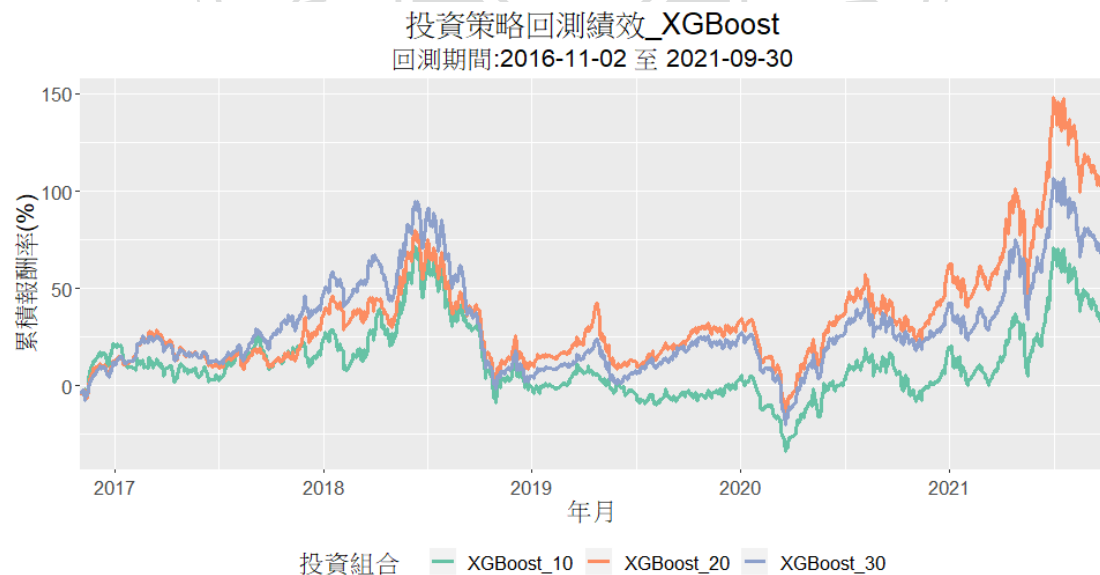


圖 11、極限梯度提升模型回測績效

投資策略回測績效_XGBoost_Wavelet
回測期間:2016-11-02 至 2021-09-30

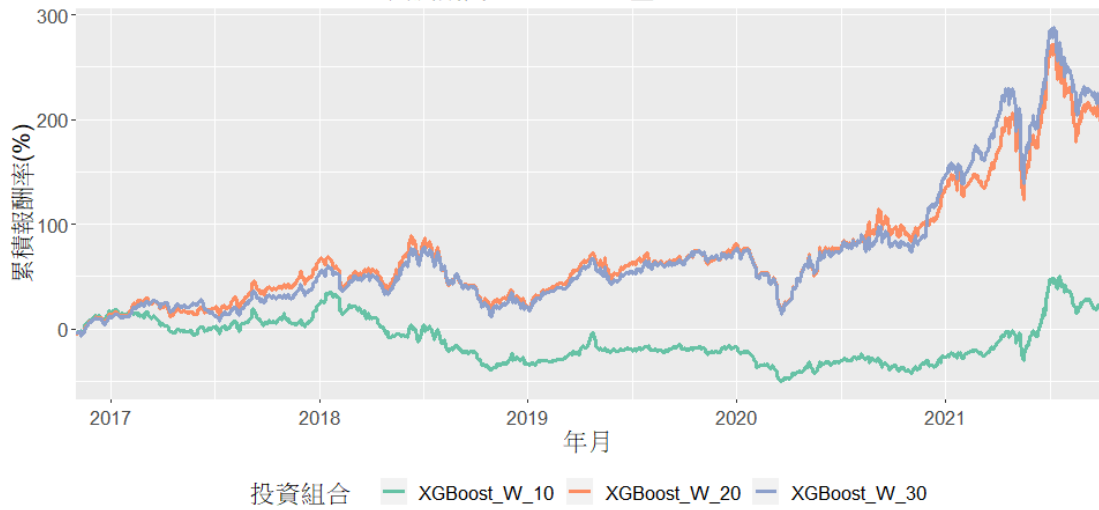


圖 12、極限梯度提升模型_離散小波轉換回測績效

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
XGB_10	31.24%	5.86%	11.28%	0.3340	0.3379	-61.54%
XGB_20	97.40%	15.31%	18.94%	0.3059	0.6192	-52.01%
XGB_30	64.14%	10.94%	14.55%	0.2880	0.5051	-58.95%
XGB_W_10	19.48%	3.80%	9.50%	0.3392	0.2801	-63.19%
XGB_W_20	196.50%	25.57%	27.35%	0.3017	0.9065	-38.93%
XGB_W_30	211.76%	26.89%	28.15%	0.2938	0.9583	-37.40%

表 5、極限梯度提升模型回測績效

表 5 中可發現，離散小波轉換對於極限梯度提升模型有顯著的影響力，使用此技術後，挑選股票上漲機率前 10 名投資組合的報酬率下降，但當投資組合為 20 支或 30 支股票時，投資組合的報酬率有非常顯著的提升，表示當股票支數增加的時候，可以幫助模型對於高報酬率的股票有更好的預測準確度。

三、輕量化的梯度提升機 (LightGBM)

圖 13、圖 14 分別為輕量化的梯度提升機收盤價使用原始收盤價、原始收盤價替換成離散小波轉換收盤價時，前 10 檔、20 檔、30 檔組合而成的投資組合表現。表 6 為輕量化的梯度提升機上述六個投資組合之績效。

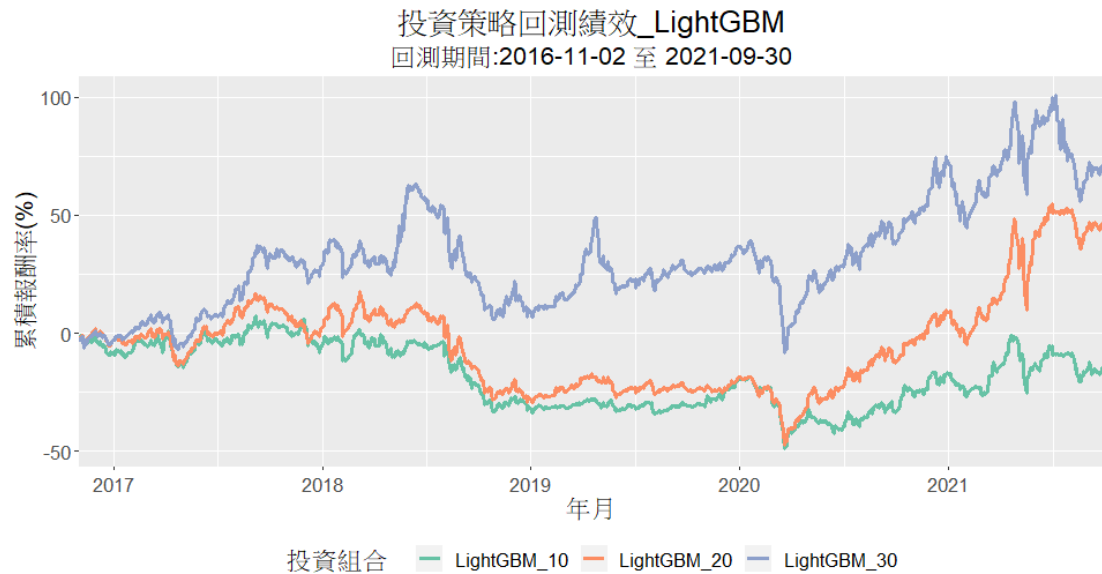


圖 13、輕量化的梯度提升機回測績效

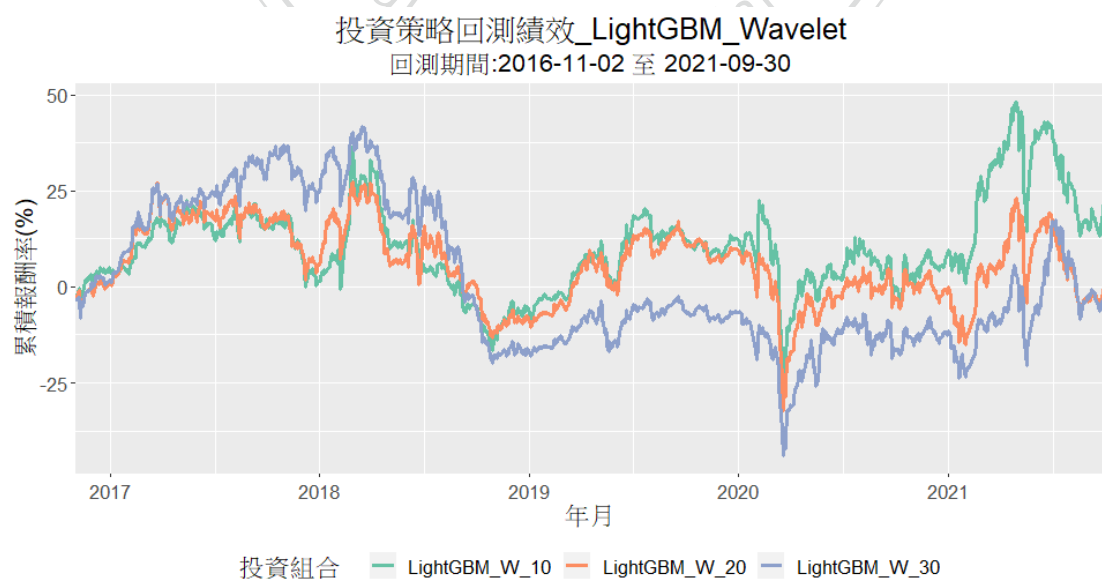


圖 14、輕量化的梯度提升機_離散小波轉換回測績效

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
LGB_10	-13.96%	-3.10%	0.01%	0.2510	0.0006	-52.49%
LGB_20	44.73%	8.05%	10.73%	0.2437	0.4404	-55.01%
LGB_30	68.22%	11.51%	14.03%	0.2497	0.5619	-43.94%
LGB_W_10	19.61%	3.82%	7.07%	0.2572	0.2748	-46.45%
LGB_W_20	-1.22%	-0.26%	2.87%	0.2494	0.1150	-46.84%
LGB_W_30	-3.86%	-0.82%	2.50%	0.2572	0.0973	-60.52%

表 6、輕量化的梯度提升機回測績效

表 6 中可發現，輕量化的梯度提升機模型使用離散小波轉換收盤價後，只有挑選股票上漲機率前 10 名的投資組合報酬率上升，其餘的投資組合報酬率皆顯著下降。相較前面的兩個預測模型，即決策樹以及極限梯度提升模型，輕量化的梯度提升機模型的預測能力相對較差。

四、支持向量機 (SVM)

圖 15、圖 16 為支持向量機有無使用離散小波轉換技術，前 10 檔、20 檔、30 檔組合而成的投資組合清單累積報酬率。表 7 為支持向量機上述六個投資組合之績效。

投資策略回測績效_SVM
回測期間:2016-11-02 至 2021-09-30

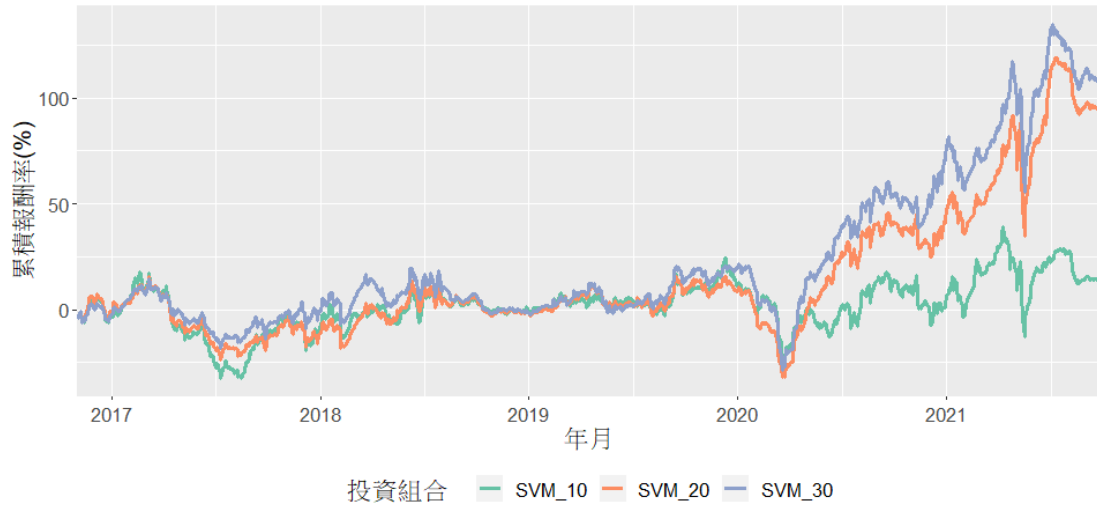


圖 15、支持向量機回測績效

投資策略回測績效_SVM_Wavelet
回測期間:2016-11-02 至 2021-09-30

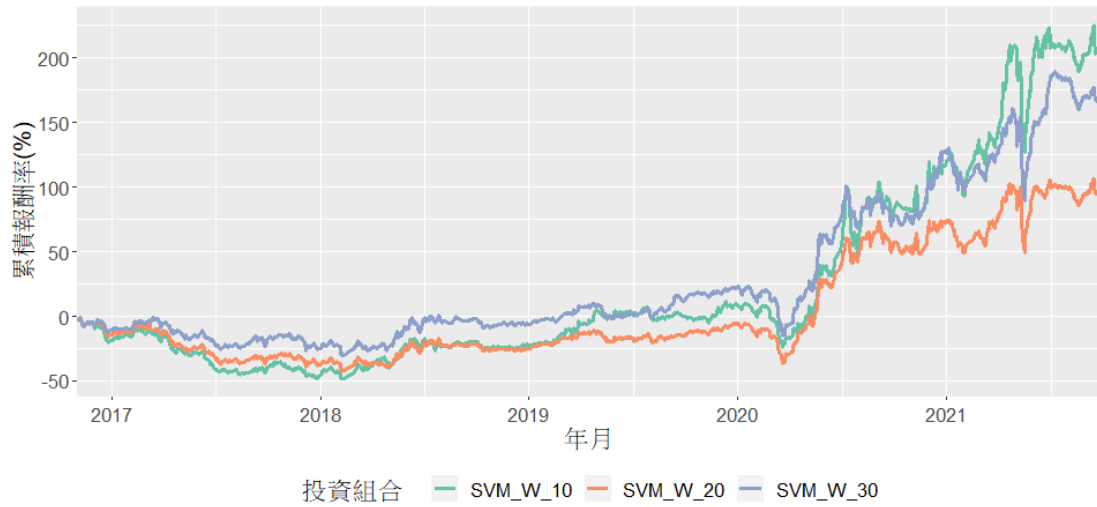


圖 16、支持向量機_離散小波轉換回測績效

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
SVM_10	13.06%	2.60%	7.51%	0.3133	0.2397	-42.76%
SVM_20	94.66%	14.97%	17.74%	0.2741	0.6472	-41.26%
SVM_30	108.37%	16.62%	18.82%	0.2615	0.7198	-41.46%
SVM_W_10	213.58%	27.05%	29.10%	0.3211	0.9062	-48.06%

SVM_W_20	100.26%	15.66%	18.43%	0.2784	0.6621	-41.40%
SVM_W_30	170.67%	23.19%	24.56%	0.2714	0.9051	-30.89%

表 7、支持向量機回測績效

觀察表 7 中數值變化可發現，將支持向量機模型的原始收盤價替換為離散小波轉換收盤價後，每個投資組合的報酬率皆增加，挑選股票上漲機率前 10 名的投資組合之年化報酬率甚至可增加接近 25%，而波動度的部分並無顯著增加，因此所有投資組合的夏普比率皆有所提升，上述結果表示離散小波轉換技術對於支持向量機模型的預測能力有非常顯著的貢獻。

五、多層感知器 (MLP)

圖 17、圖 18 為多層感知器有無使用離散小波轉換技術，前 10 檔、20 檔、30 檔組合而成的投資組合表現。表 8 為多層感知器上述六個投資組合之績效。

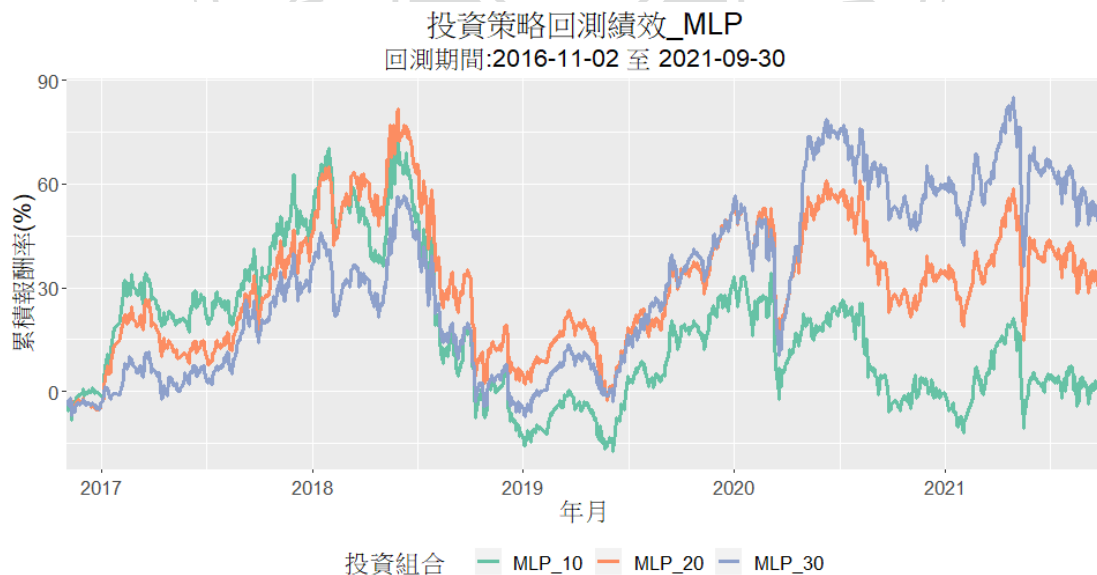


圖 17、多層感知器回測績效

投資策略回測績效_MLP_Wavelet
回測期間:2016-11-02 至 2021-09-30

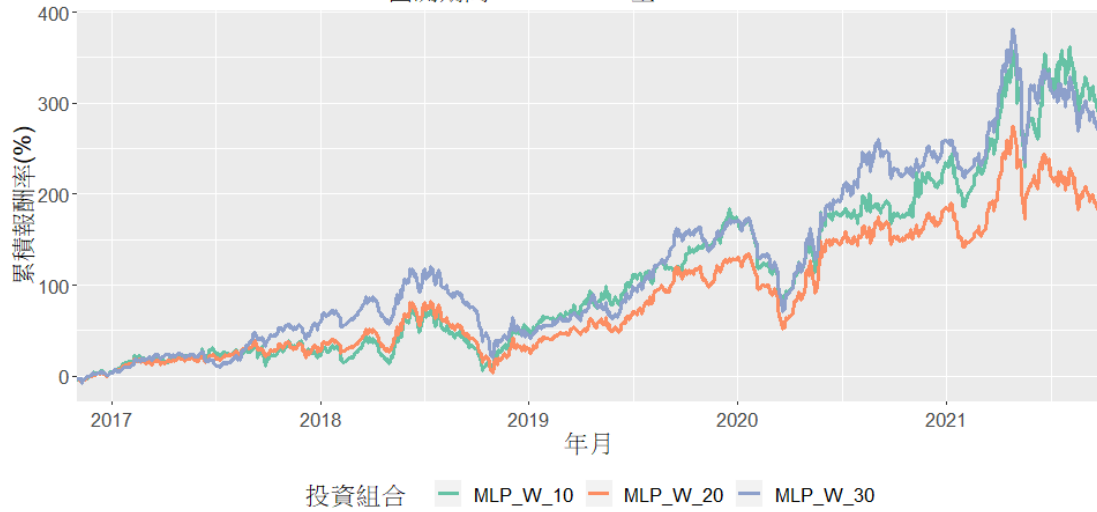


圖 18、多層感知器_離散小波轉換回測績效

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
MLP_10	1.73%	0.36%	-5.66%	0.3249	0.1741	-54.27%
MLP_20	31.83%	5.96%	10.21%	0.2966	0.3444	-46.48%
MLP_30	52.41%	9.23%	12.70%	0.2774	0.4577	-40.74%
MLP_W_10	296.59%	33.46%	35.10%	0.3526	0.9953	-39.82%
MLP_W_20	193.25%	25.28%	27.30%	0.3079	0.8865	-42.95%
MLP_W_30	282.23%	32.43%	32.58%	0.2986	1.0910	-45.50%

表 8、多層感知器回測績效

表 8 中可以發現，多層感知器模型使用離散小波轉換收盤價後，每個投資組合的報酬率皆有大幅度的增加，預期上漲機率最高的 10 支股票組成的投資組合，其年化報酬率甚至增加約 33%，雖然使用離散小波轉換後，年化標準差有些微上升，但夏普比率仍有非常大福的增長，上述結果表示，離散小波轉換技術對於多層感知器模型的預測能力有非常顯著的貢獻。

整體而言，在尚未使用離散小波轉換收盤價前，挑選股票上漲機率前 10 名

的投資組合中，報酬率最高的為決策樹模型；挑選股票上漲機率前 20 名的投資組合中，報酬率最高的模型為極限梯度提升模型；挑選股票上漲機率前 30 名的投資組合中，報酬率最高的為支持向量機模型。將離散小波轉換收盤價作為模型輸入特徵後，多層感知器模型以及支持向量機模型的模型預測能力有非常顯著的上升。挑選股票上漲機率前 10 名的投資組合中，報酬率最高的為多層感知器模型；挑選股票上漲機率前 20 名的投資組合中，報酬率最高的模型為極限梯度提升模型；挑選股票上漲機率前 30 名的投資組合中，報酬率最高的為多層感知器模型。

第二節 集成學習

為了觀察集成學習的效果，本節呈現在未使用離散小波轉換技術的情況下，使用集成學習堆疊法，將五個機器學習模型結合後的回測表現，以觀察模型預測能力是否提升，以及與台灣股市大盤績效比較。圖 19 為透過集成學習選股，依上漲機率最高的 10 檔、20 檔、30 檔股票組成三個投資組合清單以及台股大盤的累積報酬率圖，表 9、表 10、表 11 為單一模型與集成學習模型的比較表，表 12 為集成學習與台股大盤的績效比較表。

集成學習回測績效
回測期間:2016-11-01 至 2021-09-30

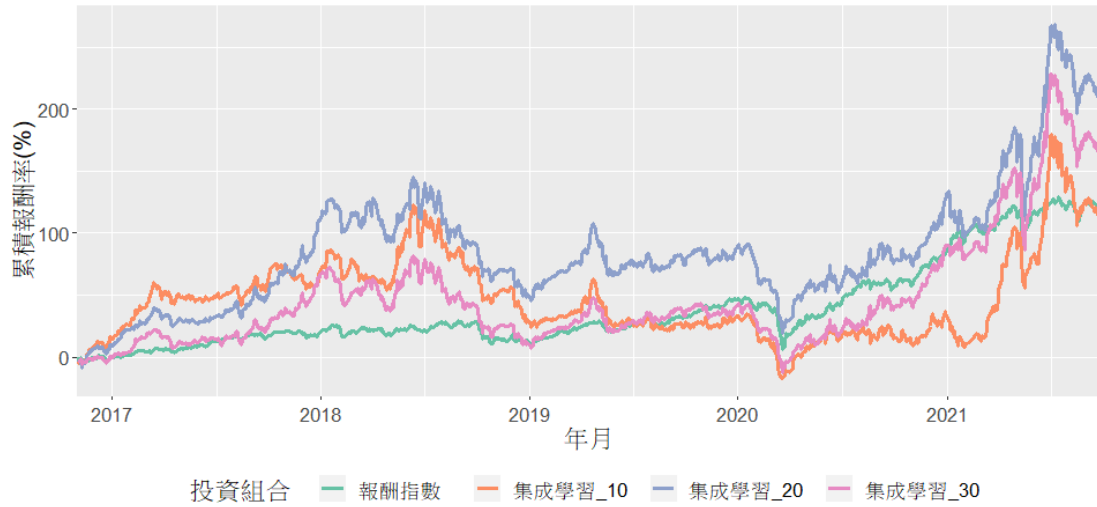


圖 19、集成學習回測績效

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
集成_10	109.92%	16.81%	21.14%	0.3342	0.6324	-62.94%
DT_10	101.28%	15.78%	16.54%	0.1971	0.8394	-28.21%
XGB_10	31.24%	5.86%	11.28%	0.3340	0.3379	-61.54%
LGB_10	-13.96%	-3.10%	0.01%	0.2510	0.0006	-52.49%
SVM_10	13.06%	2.60%	7.51%	0.3133	0.2397	-42.76%
MLP_10	1.73%	0.36%	5.66%	0.3249	0.1741	-54.27%

表 9、10 檔股票投組回測績效

表 9 中可清楚發現，透過集成學習的堆疊技巧將五個模型結合，利用 Meta-Model 的預測結果挑選上漲機率最高的 10 支股票組成投資組合，報酬率比單一模型預測結果最佳的決策樹還高，雖然其他單一模型的預測結果不佳，但是在集成學習的模型中仍可保留其表現好的部分，對於整個模型有所貢獻，讓集成學習模型的預測能力提高。

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
集成_20	202.75%	26.12%	27.83%	0.3030	0.9184	-51.25%
DT_20	69.61%	11.70%	12.69%	0.1805	0.7031	-29.83%
XGB_20	97.40%	15.31%	18.94%	0.3059	0.6192	-52.01%
LGB_20	44.73%	8.05%	10.73%	0.2437	0.4404	-55.01%
SVM_20	94.66%	14.97%	17.74%	0.2741	0.6472	-41.26%
MLP_20	31.83%	5.96%	10.21%	0.2966	0.3444	-46.48%

表 10、20 檔股票投組回測績效

表 10 的結果為，集成學習模型與單一模型挑選預測上漲機率最高的 20 支股票組成投資組合之績效比較，相較於 10 支股票的投資組合，20 支股票的投資組合在每個模型的預測表現上較為平均，其利用集成學習預測的結果也有顯著的表現，報酬率遠高於單一模型，表示利用集成學習技術確實可有效提升模型的預測能力。

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
集成_30	158.12%	21.97%	24.20%	0.2935	0.8246	-51.71%
DT_30	82.65%	13.45%	14.30%	0.1828	0.7821	-29.67%
XGB_30	64.14%	10.94%	14.55%	0.2880	0.5051	-58.95%
LGB_30	68.22%	11.51%	14.03%	0.2497	0.5619	-43.94%
SVM_30	108.37%	16.62%	18.82%	0.2615	0.7198	-41.46%
MLP_30	52.41%	9.23%	12.70%	0.2774	0.4577	-40.74%

表 11、30 檔股票投組回測績效

表 11 的結果顯示，在投資組合為 30 檔股票的情況下，集成學習模型的報酬率與單一模型相比，有明顯的提升，與 10 支股票的投資組合相比，30 支股票的投資組合在每個模型的預測表現上較為平均，其利用集成學習技術的效果也較為顯著。

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
報酬指數	119.01%	17.83%	17.61%	0.1542	1.1418	-28.55%
集成_10	109.92%	16.81%	21.14%	0.3342	0.6324	-62.94%
集成_20	202.75%	26.12%	27.83%	0.3030	0.9184	-51.25%
集成_30	158.12%	21.97%	24.20%	0.2935	0.8246	-51.71%

表 12、集成學習回測績效

表 12 的結果為集成學習模型的三個投資組合與大盤的績效比較，可觀察出使用集成學習堆疊法技巧後，將上漲機率最高的 20 檔與 30 檔股票組成的投資組合，其報酬率表現優於大盤，但標準差的部分和大盤相比略高，以至於夏普比率仍低於大盤，對於積極追求報酬率的民眾而言，若以此模型創建之理財機器人作為投資方式仍不失一個好選擇。

整體而言，在尚未使用離散小波轉換技術時，集成學習模型可有效保留每個單一模型的優勢，並將其弱點去除，結合成一個預測能力更好的模型，不論投資組合包含 10 檔、20 檔、30 檔股票，皆比原本的單一模型能夠創造更高的報酬率，年化報酬率甚至能超過大盤 8% 以上，顯示出集成學習在本研究中具有關鍵的影響力。

第三節 最終模型

前面兩節的結果中發現使用離散小波轉換技術後，能使大部分單一模型預測能力有顯著的提升，以及使用集成學習後，能使集成學習模型比單一模型有更好的預測能力，因此本節將呈現離散小波轉換技術結合集成學習模型，模型回測的結果，看是否同時使用兩種策略，可使模型的預測能力更好。將離散小

波轉換結合集成學習模型的預測結果，依上漲機率最高的 10 檔、20 檔、30 檔股票組成三個投資組合清單，圖 20 為累積報酬率，表 13、表 14、表 15 為在使用離散小波技術的情況下，單一模型與集成學習模型的比較表，表 16 為集成學習與台股大盤的績效比較表。

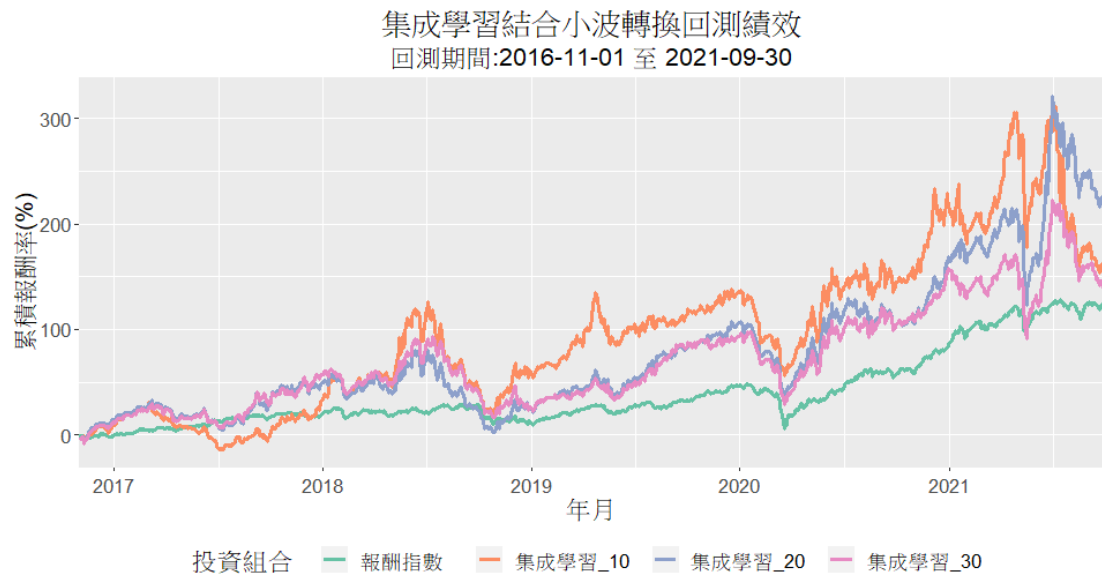


圖 20、集成學習_離散小波轉換回測績效

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
集成_W_10	152.15%	21.38%	26.62%	0.3798	0.7011	-46.71%
DT_W_10	101.28%	15.78%	16.54%	0.1971	0.8394	-28.21%
XGB_W_10	19.48%	3.80%	9.50%	0.3392	0.2801	-63.19%
LGB_W_10	19.61%	3.82%	7.07%	0.2572	0.2748	-46.45%
SVM_W_10	213.58%	27.05%	29.10%	0.3211	0.9062	-48.06%
MLP_W_10	296.59%	33.46%	35.10%	0.3526	0.9953	-39.82%

表 13、離散小波轉換_10 檔股票投組回測績效

表 13 的結果為，將上漲機率最高的 10 支股票設為投資組合時，單一模型與集成學習模型的績效比較，可發現當一般收盤價替換為離散小波收盤價時，

集成學習模型的表現並不會比單一模型好，有可能是因為支持向量機以及多層感知器的單一模型績效已有非常優異的表現，五個集成學習的模型彼此的模型預測能力差異太大，以至於無法讓集成學習發揮其截長補短的功用。

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
集成_W_20	217.24%	27.36%	29.47%	0.3246	0.9079	-42.89%
DT_W_20	74.32%	12.35%	13.28%	0.1815	0.7318	-29.83%
XGB_W_20	196.50%	25.57%	27.35%	0.3017	0.9065	-38.93%
LGB_W_20	-1.22%	-0.26%	2.87%	0.2494	0.1150	-46.84%
SVM_W_20	100.26%	15.66%	18.43%	0.2784	0.6621	-41.40%
MLP_W_20	193.25%	25.28%	27.30%	0.3079	0.8865	-42.95%

表 14、離散小波轉換_20 檔股票投組回測績效

表 14 的回測結果可發現，當一般收盤價替換為離散小波收盤價時，集成學習模型有達到其效果，獲得超過單一模型的報酬率，相較於 10 檔股票的投資組合，20 檔股票的投資組合中，集成學習模型的報酬率有所提升，其原因可能為單一模型之間的預測能力落差較小。

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
集成_W_30	139.50%	20.08%	22.82%	0.3000	0.7607	-39.60%
DT_W_30	80.88%	13.22%	14.08%	0.1819	0.7740	-29.67%
XGB_W_30	211.76%	26.89%	28.15%	0.2938	0.9583	-37.40%
LGB_W_30	-3.86%	-0.82%	2.50%	0.2572	0.0973	-60.52%
SVM_W_30	170.67%	23.19%	24.56%	0.2714	0.9051	-30.89%
MLP_W_30	282.23%	32.43%	32.58%	0.2986	1.0910	-45.50%

表 15、離散小波轉換_30 檔股票投組回測績效

表 15 的結果可發現和表 13 有相同的現象，在使用離散小波轉換技術後，多層感知器與極限梯度提升模型預測結果十分優異，五個模型彼此的預測能力

差異太大，可能導致集成學習無法讓模型更為優化，得到更好的預測結果。

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
報酬指數	119.01%	17.83%	17.61%	0.1542	1.1418	-28.55%
集成_10	109.92%	16.81%	21.14%	0.3342	0.6324	-62.94%
集成_20	202.75%	26.12%	27.83%	0.3030	0.9184	-51.25%
集成_30	158.12%	21.97%	24.20%	0.2935	0.8246	-51.71%
集成_W_10	152.15%	21.38%	26.62%	0.3798	0.7011	-46.71%
集成_W_20	217.24%	27.36%	29.47%	0.3246	0.9079	-42.89%
集成_W_30	139.50%	20.08%	22.82%	0.3000	0.7607	-39.60%

表 16、集成學習回測績效比較表

表 16 的結果為集成學習模型的三個投資組合、集成學習結合離散小波轉換的三個投資組合以及大盤的績效比較表，可發現使用離散小波轉換後，三個投資組合的報酬率皆有超過大盤，其中，10 檔與 20 檔股票的集成學習投資組合在使用離散小波轉換技術後，績效也有所提升。

由於在表 13、表 14 以及表 15 中，發現在使用離散小波轉換技術後，所有單一模型中，績效較差的模型為決策樹以及輕量化的梯度提升機，因此本研究中試著將這兩個單一模型從集成學習模型中去除，也就是說集成學習模型僅包含極限梯度提升模型、支持向量機以及多層感知器，觀察是否集成學習模型能有更好的預測能力，圖 21 為不同檔數股票投資組合下的績效圖，表 17 為減少模型與原集成學習模型的績效比較表。結果顯示，減少模型後的回測績效大幅下降，表示雖然決策樹以及輕量化的梯度提升機單一模型的預測效果不好，但是用在集成學習上，仍可保留其優點，若集成學習少了這兩個模型，績效仍會受到很大的影響。

集成學習結合小波轉換回測績效
回測期間:2016-11-02 至 2021-09-30

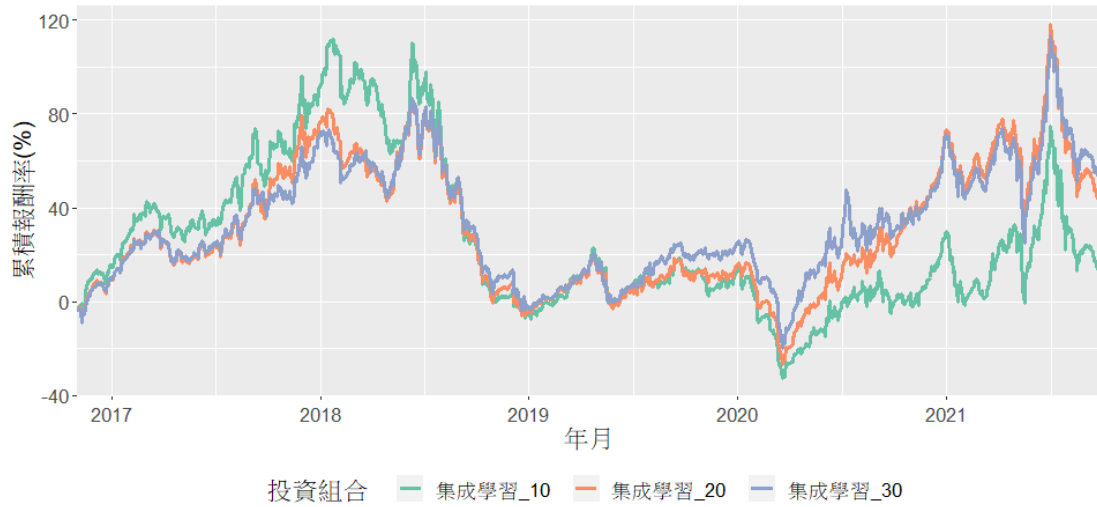


圖 21、集成學習_減少模型_離散小波轉換回測績效

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
報酬指數	119.01%	17.83%	17.61%	0.1542	1.1418	-28.55%
集成_W_10 (減少模型)	10.42%	2.10%	7.75%	0.3359	0.2009	-68.30%
集成_W_20 (減少模型)	39.07%	7.15%	11.40%	0.2991	0.3479	-60.78%
集成_W_30 (減少模型)	48.18%	8.59%	12.37%	0.2868	0.3965	-56.94%
集成_W_10	152.15%	21.38%	26.62%	0.3798	0.7011	-46.71%
集成_W_20	217.24%	27.36%	29.47%	0.3246	0.9079	-42.89%
集成_W_30	139.50%	20.08%	22.82%	0.3000	0.7607	-39.60%

表 17、集成學習回測績效比較表

綜合以上結果發現，離散小波轉換技術以及集成學習堆疊法在提升模型績效上，都有顯著的表現，離散小波轉換對於提升支持向量機以及多層感知器的模型績效最為顯著，其次為極限梯度提升模型，集成學習在未使用離散小波轉換技術時，集成學習皆能有效提升模型的預測能力，使用離散小波轉換技術

後，因為有些單一模型的績效會有非常大的提升，所以再使用集成學習技術後就不一定能再進一步提升模型的預測能力。在本研究的最後，將離散小波轉換結合集成學習模型的投資組合配置權重方式由平均-變異最適投資組合改為等權重配置，觀察是否報酬率會顯著降低，圖 22 為等權重配置的累積報酬率，表 18 為兩種權重配置方式與台股大盤的績效比較表，可發現 10 檔與 30 檔投資組合的報酬率有明顯提升，且三個投資組合回測績效仍皆優於大盤。

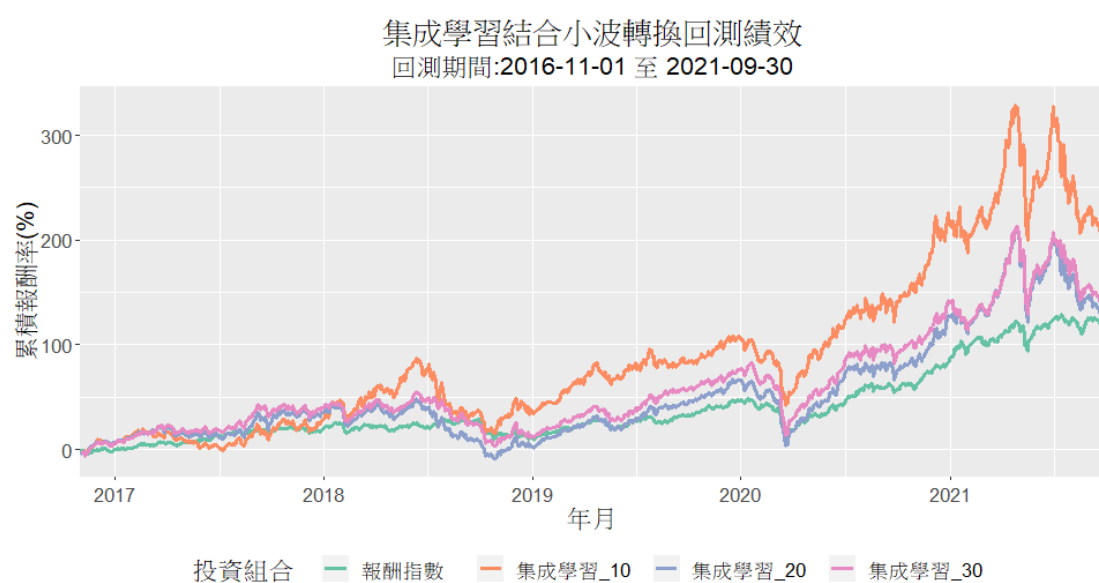


圖 22、集成學習_離散小波轉換回測績效

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
報酬指數	119.01%	17.83%	17.61%	0.1542	1.1418	-28.55%
集成_W_10 (等權重)	209.07%	26.66%	28.34%	0.3053	0.9281	-38.35%
集成_W_20 (等權重)	134.28%	19.52%	21.33%	0.2629	0.8110	-38.65%
集成_W_30 (等權重)	145.43%	20.69%	21.85%	0.2453	0.8907	-38.17%
集成_W_10 (平均變異)	152.15%	21.38%	26.62%	0.3798	0.7011	-46.71%

集成_W_20 (平均變異)	217.24%	27.36%	29.47%	0.3246	0.9079	-42.89%
集成_W_30 (平均變異)	139.50%	20.08%	22.82%	0.3000	0.7607	-39.60%

表 18、集成學習_離散小波轉換回測績效

將在各情況下的最佳投資組合綜合比較的績效結果如表 19，在投資組合設定為 10 檔股票時，報酬率最高的為離散小波轉換結合集成學習，並使用等權重方法進行投資組合權重配置；投資組合設定為 20 檔股票時，報酬率最高的為離散小波轉換結合集成學習，並使用平均-變異 (Mean Variance) 最適投資組合進行投資組合權重配置；在投資組合設定為 30 檔股票時，報酬率最高的為使用集成學習但不使用離散小波轉換，且使用平均-變異 (Mean Variance) 最適投資組合進行投資組合權重配置，以上的投資組合報酬率皆明顯大於大盤報酬，若以此三個投資組合建立的理財機器人，將能有效滿足投資人達到自行投資且報酬率優於大盤之目的。

投資組合	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回撤率
報酬指數	119.01%	17.83%	17.61%	0.1542	1.1418	-28.55%
集成_W_10 (等權重)	209.07%	26.66%	28.34%	0.3053	0.9281	-38.35%
集成_W_20 (平均變異)	217.24%	27.36%	29.47%	0.3246	0.9079	-42.89%
集成_30 (平均變異)	158.12%	21.97%	24.20%	0.2935	0.8246	-51.71%

表 19、最佳績效投資組合比較表

第五章 結論與建議

本研究使用台灣上市公司股價資料、其計算出之技術指標以及總經指標，利用離散小波轉換技術將收盤價分解，去除雜訊干擾部分，並將以上資料作為模型輸入特徵放入機器學習模型進行股價預測，利用集成學習的堆疊法技巧將決策樹（Decision Tree）、極限梯度提升模型（XGBoost）、輕量化的梯度提升機（LightGBM）、支持向量機（SVM）以及多層感知器（MLP）整合，並在第二階段使用羅吉斯迴歸（Logistic Regression）做出股價預測，依個股上漲機率預測的結果將前 10 檔、20 檔以及 30 檔股票分別組成投資組合，並以平均-變異（Mean Variance）方法配置投資組合的個股權重進行回測。結果顯示使用離散小波轉換可顯著提升支持向量機以及多層感知器的模型預測能力，使用集成學習在大部分情況也能夠獲得比單一模型更好的投資績效，但若原本的模型預測能力已經非常良好，則集成學習可能就無法再進一步提升模型的績效。

本研究使用非常多技術指標以及一些總經指標作為特徵輸入模型，在技術指標部分，未來可以嘗試先將特徵篩選，挑選對模型預測較有幫助的特徵再放入模型；總經指標部分，未來可以嘗試其他的總經指標，或許對模型的預測會有更大的幫助。離散小波轉換的部分，在未來可以嘗試使用其他的基底函數，或是將其他輸入特徵也一併分解。而在模型的訓練上，除了本研究使用的五個模型外，可以增加更多的異質模型看是否可以增加模型預測的準確度，集成學習的 Meta-Model 也可以嘗試使用其他模型。

參考文獻

- Chang, T. S. (2011) . A comparative study of artificial neural networks, and decision trees for digital game content stocks price prediction. *Expert systems with applications*, 38 (12) , 14846-14851.
- Chen, Y., Liu, K., Xie, Y., & Hu, M. (2020) . Financial trading strategy system based on machine learning. *Mathematical Problems in Engineering*, 2020.
- Chhajer, P., Shah, M., & Kshirsagar, A. (2022) . The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction. *Decision Analytics Journal*, 2, 100015.
- Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018) . NSE stock market prediction using deep-learning models. *Procedia computer science*, 132, 1351-1362.
- Hongjoong, K. I. M. (2021) . MEAN-VARIANCE PORTFOLIO OPTIMIZATION WITH STOCK RETURN PREDICTION USING XGBOOST. *Economic Computation & Economic Cybernetics Studies & Research*, 55 (4) .
- Jiang, M., Liu, J., Zhang, L., & Liu, C. (2020) . An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. *Physica A: Statistical Mechanics and its Applications*, 541, 122272.
- Liang, X., Ge, Z., Sun, L., He, M., & Chen, H. (2019) . LSTM with wavelet transform based data preprocessing for stock price prediction. *Mathematical Problems in Engineering*, 2019.

- Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data*, 7 (1), 1-40.
- Padhi, D. K., Padhy, N., Bhoi, A. K., Shafi, J., & Ijaz, M. F. (2021). A Fusion Framework for Forecasting Financial Market Direction Using Enhanced Ensemble Models and Technical Indicators. *Mathematics*, 9 (21), 2646.
- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert systems with applications*, 42 (1), 259-268.
- Shynkevich, Y., McGinnity, T. M., Coleman, S. A., Belatreche, A., & Li, Y. (2017). Forecasting price movements using technical indicators: Investigating the impact of varying input window length. *Neurocomputing*, 264, 71-88.
- Tang, Q., Shi, R., Fan, T., Ma, Y., & Huang, J. (2021). Prediction of Financial Time Series Based on LSTM Using Wavelet Transform and Singular Spectrum Analysis. *Mathematical Problems in Engineering*, 2021.
- Weng, B., Martinez, W., Tsai, Y. T., Li, C., Lu, L., Barth, J. R., & Megahed, F. M. (2018). Macroeconomic indicators alone can predict the monthly closing price of major US indices: Insights from artificial intelligence, time-series analysis and hybrid models. *Applied Soft Computing*, 71, 685-697.
- Wu, D., Wang, X., & Wu, S. (2021). A hybrid method based on extreme learning machine and wavelet transform denoising for stock prediction. *Entropy*, 23 (4), 440.
- Ye, Z., Wu, Y., Chen, H., Pan, Y., & Jiang, Q. (2022). A Stacking Ensemble Deep Learning Model for Bitcoin Price Prediction Using Twitter Comments on Bitcoin. *Mathematics*, 10 (8), 1307.