國立政治大學風險管理與保險學系

碩士學位論文

結合基本面分析及集成學習模型建構最適投資組合

Combining Fundamental Analysis and Ensemble Learning to Construct the Optimal Portfolio

指導教授：黃泓智　博士

研究生：許育愷　撰

中華民國一一〇年六月

DOI:10.6814/NCCU202100808

摘要

本研究透過結合基本面分析以及集成學習的方式，在上市個股當中建立最適的投資組合，首先藉由兩個基本面的因子：本益比（PE Ratio）與股價淨值比（PB Ratio）篩選個股，接著，藉由包含支持向量迴歸模型（SVR）及多層感知器（MLP）的集成學習模型預測個股的數日後報酬，藉此挖掘出有潛力的個股進行投資組合的配置，並藉由歷史回測分析其績效。實驗結果顯示，在基本面篩選下的投資組合績效高於大盤指數績效，而加入集成學習模型後可以進一步提升其報酬績效。

國立政治大學
National Chengchi University

關鍵字：基本面分析、集成學習、支持向量迴歸模型、多層感知器、報酬率預測、投資組合

DOI:10.6814/NCCU202100808

Abstract

The purpose of this study is trying to combine fundamental analysis and ensemble learning model to construct an optimal portfolio with listed stocks. First, use two fundamental factor, price-to-equity ratio and price-to-book ratio, to select potential stocks, and then predict the future return of each stock through ensemble learning model which including SVR and MLP to construct an optimal portfolio. Finally, analysis the performance according history stock data. The result shows that the portfolio constructed by only fundamental analysis outperforms Taiwan Capitalization Weighted Stock Index (TAIEX), and the portfolio constructed by fundamental analysis and ensemble learning outperforms the portfolio constructed by only fundamental analysis.

國立政治大學 National Chengchi University

Keywords: Fundamental Analysis, Ensemble Learning, SVR, MLP, Return Predict, Portfolio

DOI:10.6814/NCCU202100808

目次

第一章　緒論.....6
第一節　研究動機.....6
第二節　研究目的.....7
第三節　研究流程.....7
第二章　文獻回顧.....9
第一節　投資組合文獻探討.....9
第二節　基本面分析文獻探討.....10
第三節　集成學習文獻探討.....10
第三章　研究方法.....12
第一節　研究架構.....12
第二節　基本面篩選.....13
第三節　集成學習模型.....16
第四節　權重配置方法.....23
第五節　績效分析指標.....24
第四章　實證分析.....26
第一節　SVR_MLP 模型實證結果.....26
第二節　5MLP 模型實證結果.....28
第三節　SMX 模型實證結果.....30
第五章　結論與建議.....32
參考文獻.....33

DOI:10.6814/NCCU202100808

表次

表格 1、SVR_MLP 模型實證績效.....27

表格 2、5MLP 模型實證績效.....29

表格 3、SMX 模型實證績效.....30

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202100808

圖次

圖 1、研究流程圖 8
圖 2、研究架構 13
圖 3、本益比資料處理方式 15
圖 4、集成學習架構 21
圖 5、SVR_MLP 模型架構 22
圖 6、5MLP 模型架構 23
圖 7、SMX 模型架構 23
圖 8、效率前緣與權重配置 24
圖 9、SVR_MLP 歷史累積報酬率走勢圖 27
圖 10、5MLP 歷史累積報酬率走勢圖 29
圖 11、SMX 歷史累積報酬率走勢圖 31

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202100808

第一章　緒論

第一節　研究動機

在 2019 年 COVID-19 疫情爆發時，世界各國的經濟皆受到強烈衝擊，引發了金融危機，而台灣的大盤指數在疫情爆發初期雖一度重跌至 9234 點，但由於當時臺灣疫情控制穩定，台指隨後便快速攀升至四月底的 17595 點，可以窺得金融危機過後股市活絡之現象，再加上台灣的高科技產業世界不景氣下逆風高飛，因此有大量的散戶決定在這時加入股市投資的行列。

綜觀民眾的理財方式，除了最普遍的儲蓄外，大多人會選擇購買一些理財商品，諸如儲蓄型或利變型保險商品、共同基金，然而這些商品雖然利潤相對穩定卻必須被收取不少手續費用或是管理費用；而有一些人會選擇透過理專的專業理財能力來代為管理資產，除了必須支付管理費用之外，自今年初以來陸續發生的理專掏空案也不免讓投資民眾人心惶惶，難以將自己畢生積蓄交由他人代管。因此，許多民眾決定搭上這波股市熱潮，隻身投入股票買賣的行列，然而民眾在選擇投資標的時，往往沒有時間及能力去針對上市股票一一進行分析，通常是透過親朋好友或是理財節目「報明牌」的方式來決定，且常常投資於固定一檔或少數幾檔個股，造成投資風險高、獲利不穩定等情形。

所幸近年來科技迅速發展，在資訊流通速度及運算處理能力方面有大幅度的進步，再加上人工智慧（Artificial Intelligence）的發展，許多金融業也陸續推出新穎的金融科技（FinTech）服務。而其中一大特色便是理財機器人，透過機器學習的方式，理財機器人能夠提供投資人一些投資標的選擇，而且相關的手續費用或成本較其他投資方式低廉，更進一步我們希望能透過理財機器人建構長期而穩定的投資組合。

DOI:10.6814/NCCU202100808

第二節　研究目的

綜觀機器學習應用於股價學習的相關文獻，可以發現大多數研究都著重於大盤指數或少數個股的未來股價、未來報酬率、上漲機率等方面進行預測，鮮少有研究將這些技術延伸應用於長期投資組合建構。因此本研究希望能夠透過機器學習建構出一種理財機器人，並提供投資組合而非單一個股的投資推薦，試圖透過較為低廉的營運成本提供投資人一個在追求投資效益最大化的同時也能在投資組合分散風險效益下兼顧降低風險的參考資訊，並透過臺灣上市企業的歷史資料進行回測

第三節　研究流程

確立研究目的後，本研究依照研究動機蒐集相關文獻並做整理，並透過台灣經濟新報資料庫（TEJ）蒐集臺灣上市公司歷史股價資訊進行回測並進行績效分析。本論文主要分為五個部分：

一、　緒論

闡述本論文將機器學習運用於建構投資機器人的動機與目的。

二、　文獻回顧

探討與本研究相關之文獻資料，其中包含投資組合建構、基本面、集成學習等相關文獻回顧，並從中汲取部分觀點提升本研究素質。

三、　研究方法

本章節將會敘述本研究主要的研究架構，說明使用變數的產生及處理方式，以及集成學習的組成及運用模式，並提供績效分析的方法。

四、　實證分析

透過臺灣上市公司歷史的股價資料進行長期投資組合模擬，並分析不

DOI:10.6814/NCCU202100808

同集成學習模組下實證分析的結果。

五、 結論與建議

針對本研究做總結，並提出建議及未來方向參考。

本研究架構如圖 1 所示：

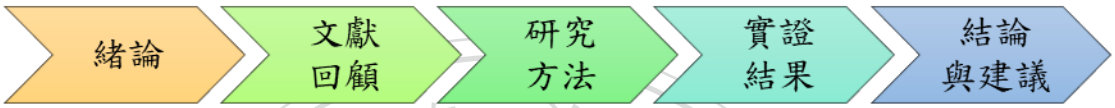

圖 1、研究流程圖

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202100808

第二章　文獻回顧

本研究欲透過集成學習的方式建立一套能夠透過分析基本面及技術面指標的方式提出投資組合的組成，並透過歷史文獻及研究的探討，尋找出合適的資料進行參考。以下將分為三個部分討論：第一節是針對投資組合建構的文獻探討，第二節以基本面分析相關的文獻回顧為主，第三節則以集成學習的相關文獻進行整理。

第一節　投資組合文獻探討

美國經濟學家 Harry Markowitz(1927)曾提出投資組合理論（Portfolio Theory），採用平均值變異數模型（Mean-Variance Model）分析投資組合，得出可以透過投資組合在固定預期報酬下達到降低投資風險的效益，並提出效率前緣（Efficient Frontier）等觀念，提供一個有效率的方式幫助投資人在風險與報酬率當中做取捨。

在衡量風險及報酬方面，美國還有一位經濟學家 Sharpe W.F. (1966) 曾提出夏普比率（Sharpe ratio）可以用來衡量一個投資組合在承擔每一單位風險下獲得溢額報酬（excess return）的效率。

在實務建構方面，張家偉（2017）以臺灣上市股票資料進行實證分析，將所有上市個股先經過基本面分析的方式進行第一次篩選，隨後透過籌碼面分析結合關聯性分析演算法再篩選一次，進而找出最佳投資組合組成。

而本研究參考張家偉（2017）的研究架構，先將所有臺灣上市個股先透過基本面的篩選，隨後再進行集成學習，除了有助於挑選出有較有獲利能力的股票外，也有助於縮減整理運算處理的時間成本。

DOI:10.6814/NCCU202100808

第二節　基本面分析文獻探討

基本面分析法式目前股市投資人當中，在面對中長期投資時最廣泛運用的方法之一，主要是透過分析一家企業定期發布的財報資訊分析其獲利能力，藉此在股票投資中獲利。

Lev and Thiagarajan (1993)經過實證分析得出財報與股價報酬有著顯著的關係。Baus (1977) 以紐約證交所（NYSE）的上市股票做實證研究，根據本益比（Price-to-Equity Ratio）大小將股票進行分群，而實證結果顯示擁有較低本益比之股票有較高的報酬率。Chan(1991)等人以東京證交所（TSE）的上市股票做研究，實證結果顯示淨值股價比（Book-to-Market Ratio）和股票報酬有正向關係。

蕭鈞銓（2016）提出比例交集法，透過將數個不同的基本面因子做結合並調整篩選閾值，將個股進行階層式篩選，並使用臺灣上市股票資訊進行實證分析，實驗結果證實，在採用自由現金流收益率（Free cash flow yield）前 10%、資產報酬率（ROA）前 20%做為兩階段的基本面篩選條件，並搭配一些權重配置的方法可以獲得最大效益。

國立政治大學 National Chengchi University

第三節　集成學習文獻探討

近年來，機器學習被廣泛地運用在股價預測的方面，諸如 Hegazy, Soliman and Salam(2014)使用支持向量迴歸模型（SVR）和類神經網路模型（ANN）分別結合五種技術指標對未來股價進行學習並預測。Huang and Liu(2019)則採用 ANN、SVR、隨機森林（Random Forest）及單純貝氏分類器（Naive-Bays Classifier）分別使用籌碼面及技術面的指標對大盤指數（TWSE Index）隔日漲跌方向進行學習及預測。

DOI:10.6814/NCCU202100808

隨著機器學習的發展，集成學習的概念也被提出來，藉由將數個相同或不同的分類器結合起來，除了能提高準確度還能減少過度學習（Overfitting）的情況產生。最早由 Clemen(1989)提出集成學習的概念。Mangalampalli 等人（2020）提出藉由集成線性及非線性的模型可以得到較好的學習效果，並研究將 ARIMA-GRU 模型運用於單一個股的開盤價預測上。

Zhang, Li, and Chen (2020)提出集成 SVR 及多層感知器（MLP）兩個學習器來學習，首先針對個股的技術面指標先用 SVR 學習器進行預測，其後再將結果輸入 MLP 進而預測出個股未來的股價，該研究將模型運用在上海及深圳交易所中四檔個股，並針對不同天期進行預測，均得出不錯的結果。因此本文主要參考此一集成模型並運用於臺灣股票市場實證，並進一步建構出自動化投資組合組成。

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202100808

第三章　研究方法

本章節將說明本研究的進行方式，主要分為五個小節，第一節為研究架構，說明本研究的執行流程；第二節為基本面篩選，主要介紹基本面指數以及如何透過基本面交集法篩選個股；第三節為集成學習模型，說明本研究所採用的三種集成學習模型，並針對其中使用的技術指標以及使用單元進行介紹以及學習方式說明；第四節為各種權重配置法的介紹；第五節提供數個分析整體回測績效的衡量指標。

第一節　研究架構

本研究架構如圖 2 所示，每一個交易日開始時，便會先計算上一個交易日的投資組合報酬，由於缺乏盤中資訊，因此本研究採用兩日間收盤價作為報酬率衡量，唯有回測起始時會跳過此步驟。為避免頻繁交易造成交易成本過重的問題，本研究設定一個再平衡期間，在回測時每經過一定時間才會更新一次投資組合，若回測當日非再平衡日，則直接前往下一個交易日，並且預設回測起始日為第一個再平衡日（方能進行投資組合組成）。

若該交易日屬於再平衡日將會透過一系列方法組成投資組合。首先將該交易日所有個股先經過基本面交集法進行第一次篩選，此一步驟除了可以先篩選出具潛力個股外，也可以減少後續集成學習所需要耗費的計算量，詳細說明將會於第二節提出。

在經過基本面篩選過後，接著會再進行一次集成學習模型的篩選，本研究在這部分採用三種集成學習模型進行比較，分別是 SVR-MLP、5MLP、SMX，詳細說明將會在第三節提出，而績效比較將會在第四章呈現。

經過前兩步驟篩選出的股票最後再加上權重比例計算後便會形成一個投資

DOI:10.6814/NCCU202100808

組合，並隨著回測計算報酬，直到下一個再平衡日才會再更新投資組合，詳細說明將會呈現於第四節。

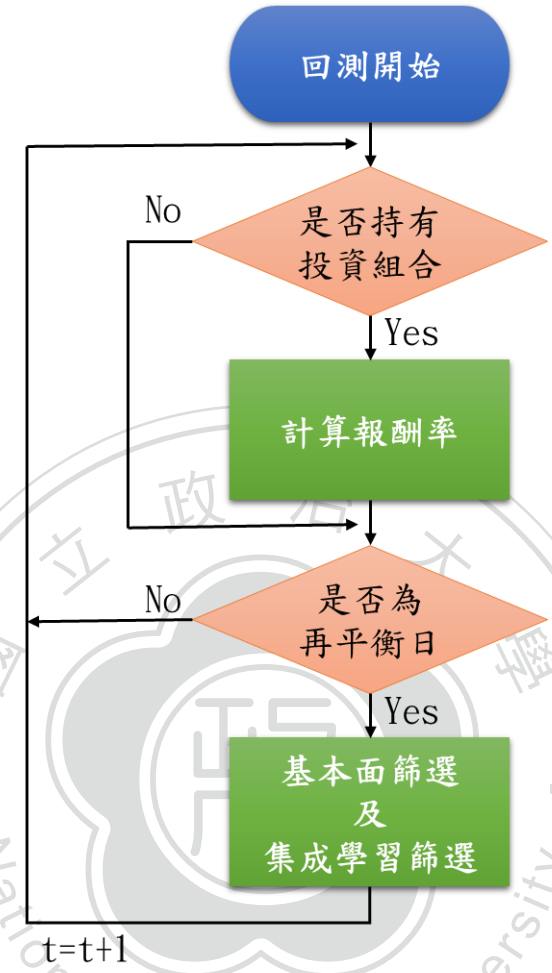

圖 2、研究架構

第二節　基本面篩選

當回測期間進入再平衡日時，本研究將進行一系列的步驟重新調整投資組合組成。首先將再平衡當日所有個股挑選出來，先進行一次基本面篩選，本研究參考蕭鈞銓（2008）的基本面雙因子比例交集法，其概念十分簡單，所有個股必須先經過第一個基本面因子篩選才能進到第二個基本面篩選，舉例來說，若採用 ROA(0.5)-EPS(0.2)篩選，則所有股票當中 ROA 前 50%的股票會通過第

DOI:10.6814/NCCU202100808

一階段篩選並進到第二階段，在滿足 ROA 前 50%之股票當中 EPS 前 20%的股票作為最終結果被篩選出來。

本研究會採用此一篩選步驟主要有兩大因素，首先，經過此一篩選方法可以將具有獲利潛能的個股篩選出來，同時將表現不佳之個股排除；另外，透過減少進入集成學習篩選法階段之個股數，可以大幅度減輕整體的運算時間，進而提高整理運行效率。

結合其他文獻參考，本研究最終選出採用本益比（PE Ratio）以及市價淨值比（PB Ratio）做為基本面篩選的雙因子，以下將針對兩個因子進行介紹及資料處理方式的說明。

一、 本益比（Price to Equity Ratio）

$$\text{本益比} = \frac{\textit{每股市值（Price）}}{\textit{每股盈餘（EPS）}}$$

本益比是透過將該企業之股價除上每股盈餘所產生的指數，可以視為購買該企業每一單位資本盈餘所花費的成本。根據本益比效益指出，當該企業本益比越低時，表示你用相對低廉的成本來購買該企業的資本盈餘，反之，本益比越高，表示你必須用相對高昂的價格來購買該企業相同單位的獲益。因此，本研究在採用這項基本面因子時，傾向採用本益比為正（由於股價恆非負，故同理於挑選 EPS 為正之企業）且數值越小的企業為佳。

在資料處理方面，由於 EPS 揭露於每期財務報表，是為季資料，而股價是日資料，因此，當每一個再平衡日要篩選本益比指數時，本研究採用當時已公布財報之 EPS 值及前一日之收盤價計算出當日本益比指數，詳細方式如圖 3 所示。另外，為了方便篩選出越小且為正值的本益比指數，本研究將該指數進行倒數處理（換言之就是益本比），方便篩選時可以直接取

DOI:10.6814/NCCU202100808

值較大者。

再平衡交易日	(A):EPS	(B):close price	(C): PE Ratio = (B)/(A)
2012/01/02	2011Q3 EPS	2011/12/30 close	2012/01/02 PE Ration
2012/01/16	2011Q3 EPS	2012/01/13 close	2012/01/16 PE Ration
2012/02/07	2011Q3 EPS	2012/02/06 close	2012/02/07 PE Ration
2012/02/21	2011Q3 EPS	2012/02/20 close	2012/02/21 PE Ration
2012/03/07	2011Q3 EPS	2012/03/06 close	2012/03/07 PE Ration

圖 3、本益比資料處理方式

二、 市價淨值比（Price to Book Ratio）

$$\text{市價淨值比} = \frac{\text{每股市值（Price）}}{\text{每股淨值（Book Value）}}$$

市價淨值比是透過將該企業之股價除上每股淨值所產生的指數，通常用於衡量該企業在市場上的價值以及在會計報表上淨值之間的差距。當該企業的市價淨值比大於 1 時，表示在市場上的投資人對於該企業有過熱交易之情形；反之，當市價淨值比小於 1 時，表示該企業在市場行情上有被低估之情形，易有併購行為的產生。根據淨值市價比效益（留意與本研究指標互為倒數）指出，淨值市價比越高的企業，其未來報酬率會越好，因此本研究傾向採用市價淨值比為正（由於股價恆非負，固同理於挑選公司淨值為正之企業）且數值越小之企業為佳。

在資料處理方面同本益比，由於每股淨值揭露於每期財務報表，屬於季資料，而股價是日資料，因此，當每一個再平衡日要篩選市價淨值比指數時，本研究採用當時已公布財報之每股淨值及前一日之收盤價計算出當日市價淨值比指數，詳細方式可以參考圖 3。另外，為了方便篩選出越小且正值之市價淨值比指數，本研究將該指數進行倒數處理（同理於淨值市價比），方便篩選時可以直接取值較大者。

DOI:10.6814/NCCU202100808

第三節　集成學習模型

所有個股在經過上一階段的雙因子基本面篩選後，具潛力個股將會進到本章節的集成學習模型進行第二階段篩選，本研究建構三種集成學習模型，分別為 SVR_MLP、5MLP、SMX 三種，其目的皆相同，跟大部分研究相似，透過每一檔個股之技術指標對其未來的報酬率進行預測。最後根據模型對每檔個股的預測報酬率進行排序篩選為最終進入投資組合之成分股。

以下將分成三個部分說明，首先是針對本研究所使用的技術指標進行介紹及資料處理的說明；接者針對本研究運用於集成學習模型中的機器學習模型單元進行介紹，分別為支持向量迴歸模型（SVR）、多層感知器（MLP）及極限梯度提升模型（XGBoost）；最後說明 SVR_MLP、5MLP、SMX 三種集成學習模型的建構方式。

一、　技術指標介紹

技術面分析是最受投資人歡迎的一種方式，由於技術指標大多指標都簡單易懂，並能簡化市場資訊提供投資人對於短期投資上的參考，此外市面上大多投資軟體皆有提供這些指標，因此大眾接受度較高。

在機器學習方面，採用技術指標的研究也多如繁星，本文主要參考 Zhang, Li, and Chen (2020)這篇文獻所採用的技術指標，挑選移動平均線（MA）、指數移動平均線（EMA）、乖離率（Bias）、平滑異同移動平均線指標（MACD）與相對強弱指標（RSI）作為預測個股未來報酬率的指標，均為日資料，以下針對這些技術指標一一進行介紹：

DOI:10.6814/NCCU202100808

1. 移動平均線（MA）

$$MA_t = \frac{1}{n}\sum_{i=1}^{n} Close_{t-i+1}$$

移動平均線又名均線，其公式簡單易懂，將近幾個交易日的交易價格進行平均，表示過去一段時間內該股票的平均成本，主要用來判斷目前股票的價格趨勢。在均線策略上，當價格或短均線向上突破長均線，表示過去的股價到現在有上漲之趨勢，未來有多頭空間，是為進場訊號；但也有人認為，當價格低於均線表示現在購入該股票的成本較低，是為進場訊號。

由於本研究預計建立一個中期的投資組合，因此採用 5MA、10MA 及 20MA 三個指數。

2. 指數移動平均線（EMA）

$$EMA_t = \frac{Close_t + (n-1) \times EMA_{t+1}}{n}$$

指數移動平均線在均線上多加入權重的概念，在指數計算時，越接近今天的價格會獲得較高的權重，越往過去的價格權重會越稀釋。由於 EMA 在近期的權重較大，因此對價格波動及反轉更為敏感，相較於均線能更快速反映市場情勢，多用於短期投資。本研究採用 12EMA 及 26EMA 兩個指數。

3. 乖離率（Bias）

$$Bias_t = \frac{Close_t - MA_t}{MA_t}$$

如果說均線代表投資人一定期間內購入股票的成本，那乖離率便表示投資人於一定期間內的平均報酬率，也是目前股價偏離

DOI:10.6814/NCCU202100808

均線的比例。當乖離率過高時表示是目前股價在高點，未來可能會下跌，反之亦然。本研究採用 5Bias 和 10Bias 兩個指數。

4. 平滑異同移動平均線指標（MACD）

$$\mathrm{MACD}_t = \frac{(\alpha - 1) \times MACD_{t-1} + 2 \times Diff_t}{\alpha + 1}, where$$

$$Diff_t = n\,EMA_t - m\,EMA_t, m > n$$

MACD 主要是透過短天期的 EMA 指數減去長天期的 EMA 產生差值（Diff_t），在對差值進行指數平滑均線。由於在股市強烈震盪時可能難以判斷股市趨勢，正如同均線是要消除股價震盪雜訊，MACD 可以進一步減少 EMA 的雜訊，提供投資人更進一步的資訊。

一般而言，MACD 會分別採用 26EMA 和 12EMA 作為長短 EMA，並用 9 日的差值做 MACD 的計算，因此本研究採用 MACD(12,26,9)來代表該指標。

5. 相對強弱指標（RSI）

$$RSI_t = \frac{\overline{U}}{\overline{U} + \overline{D}} \times 100, where$$

$$\overline{U} = \frac{1}{n}\sum_{i=1}^{n} \max(Close_{t-i+1} - Close_{t-i}, 0)$$

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} \max(Close_{t-i} - Close_{t-i+1}, 0)$$

相對強弱指標是透過一段期間內平均漲幅及平均跌幅來計算，其分別代表市場上買賣盤雙方力道的強弱，根據公式，RSI 計算出來會落在 0~100 的區間內，一般而言，若 RSI 高於 70 則表示在過去幾天內市場快速上揚，未來恐有反轉機會；反之 RSI 小於 30 表示市場迅速下跌。本研究採用 6RSI 和 12RSI 兩個指標。

DOI:10.6814/NCCU202100808

二、 機器學習模型單元

有了個股的技術指標之後，接著便是希望可以透過機器學習模型來達到預測未來報酬率的目的。本研究將在後續介紹的集成學習模型當中用到以下三種常見的機器學習模型：

1. 支持向量迴歸模型（SVR）

Support Vector Regression，簡稱 SVR，是由 SVM（Support Vector Machine）的技術延伸而來。SVM 是一種二元分類器，可用於監督式學習，其方式在於，透過在空間中尋找一個超平面（Hyper plane）使得所以資料點到這個超平面的距離總和最短且超平面兩側之分類誤差最小。SVM 的特色在於可以透過核心函數（Kernal Function）的方式先將空間上的資料點進行座標轉換，使得非線性超平面解變得可能，進而提高分類正確性。

而 SVR 便是透過同樣的技術，在資料空間中找到一個最適的迴歸平面用於預測，並搭配核心函數達到非線性解並提高準確率的目的。

2. 多層感知器（MLP）

Multilayer perceptron，簡稱 MLP，屬於深度神經網路（deep neural network）的一種。結構上通常包含至少三個層級，分別是輸入層（Input layer）、隱藏層（hidden layer）及輸出層（Output layer），並運用倒傳遞（Backward propagation）的技術達到監督式學習的效果。

在學習上主要是透過將資料送入輸入層，接著便會將輸入層的資料經過線性組合及激活函數（Active function）的轉換將資料

DOI:10.6814/NCCU202100808

送到隱藏層當中，再將隱藏層的資料經過線性組合送到輸出層作為預測值和實際值做比較，最後再透過損失函數及倒傳遞的演算法將層與層之間的權重進行修正，進而提升學習效益。

3. 極限梯度提升模型（XGBoost）

eXtreme Gradient Boosting，簡稱 XGBoost，是近年來科技比賽最為廣泛運用的模型，由於其在無論是訓練速度亦或是精準度方面都表現卓越，因此廣受好評。

XGBoosting 這項技術本身就是集成學習的一種，透過將數個弱學習器—分類和迴歸樹（Classification and Regression Trees）集合而成一個強學習器，主要用於監督式學習上，可作為分類器亦可用於迴歸問題上。

三、 集成學習模型介紹

個股在經過基本面交集法篩選過後，剩餘有潛力的每檔個股，均分別透過集成學習模型藉由其技術指標預測其未來報酬率，其後便會針對每檔股票的未來報酬率進行排序篩選，挑選未來預測報酬率最高的前幾檔個股作為投資組合的成分股，其架構如圖 4 所示。

本文採用的集成學習模型有三種，第一種是讓 SVR 及 MLP 先後預測的 SVR_MLP 模型，第二種是藉由讓 MLP 重複學習五次並取其平均的 5MLP 模型，第三種是藉由讓 SVR、MLP、XGB 分別進行一次學習再取其平均結果的 SMX 模型，以下將一一介紹：

一、 SVR_MLP 模型

根據 Zhang, Li, and Chen (2020)所提出的架構，此一集成學習

DOI:10.6814/NCCU202100808

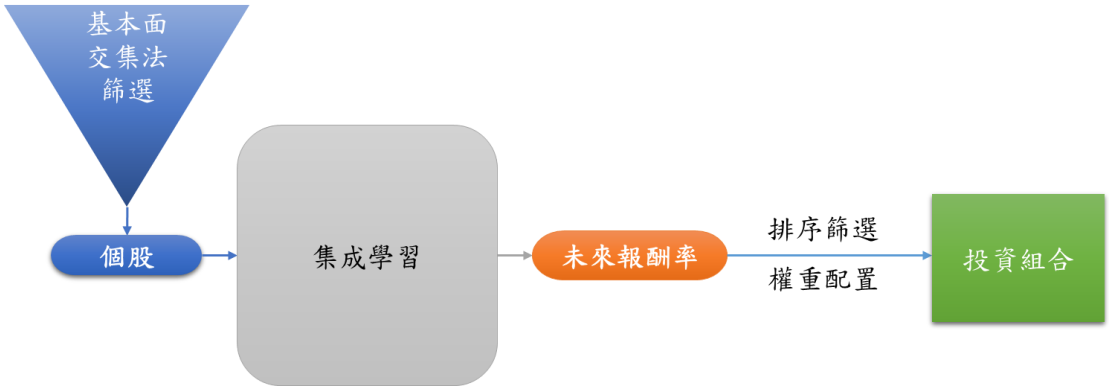

圖 4、集成學習架構

模型是針對單一個股進行未來股價之預測，而本研究稍作修改，改為針對未來報酬率的預測。

在整體架構方面，首先，產生個股的技術面指標，透過分類樹（Decision Tree）進行一次技術指標重要性篩選，將對預測未來報酬率較有效率之技術指標用於後續的集成學習當中。接著，針對重要度前五高的每個技術指標都會獨立學習一個 SVR 模型並對其於下一個再平衡日（假設是 t 日後）的數值進行預測，也就是預測 t 日後之技術指標。最後再透過 MLP 模型，輸入這些經過預測的技術指標，最終預測出該股票在下一個再平衡日的預期報酬率。整體流程如圖 5 所示。

在整個模型學習上，本研究將所有在交易當日之前的所有資料都作為訓練集，但由於 SVR 模型在訓練上需要 t 日後的技術指標，為避免使用到未來資料，因此交易日前 t 日的資料不會被列入訓練集。

二、 5MLP 模型

不同於在執行上有先後順之分的 SVR_MLP 模型，5MLP 模型顧名思義是透過 5 個 MLP 平行執行所集成的結果。在計算出

DOI:10.6814/NCCU202100808

圖 5、SVR_MLP 模型架構

個股的技術指標後，首先進行技術指標的重要度篩選，隨後再透過 MLP 模型，藉由輸入重要性前五高的技術指標資訊，預測出該股票到下一個再平衡日的預期報酬率。由於在研究時發現，當用相同條件訓練 MLP 模型時，其結果會有可能不同，時而影響甚微、時而影響甚鉅，為求其在預測結果尚能有更高的準確度及穩定性，因此本研究採用相同技術指標與訓練集同時訓練五個 MLP 模型，並對五個預測結果取平均，作為對未來報酬率的最終預測結果。整體流程如圖 6 所示。

三、 SMX 模型

SMX 模型和 5MLP 模型較為相似，相較於 5MLP 是採用數個相同的機器學習器單元進行組合，SMX 模型則是透過結合 SVR、MLP 和 XGBoost 三個不同的機器學習單元進行集成。大致上的架構模式與 5MLP 雷同，首先進行技術指標的重要度篩選，隨後再透過 SVR、MLP 或 XGBoost 模型，藉由輸入重要性前五高的技術指標資訊，分別預測出該股票到下一個再平衡日的預期報酬率，最後再取其三者的平均作為對未來報酬率之最終預測結果。整體流程如圖 7 所示。。

DOI:10.6814/NCCU202100808

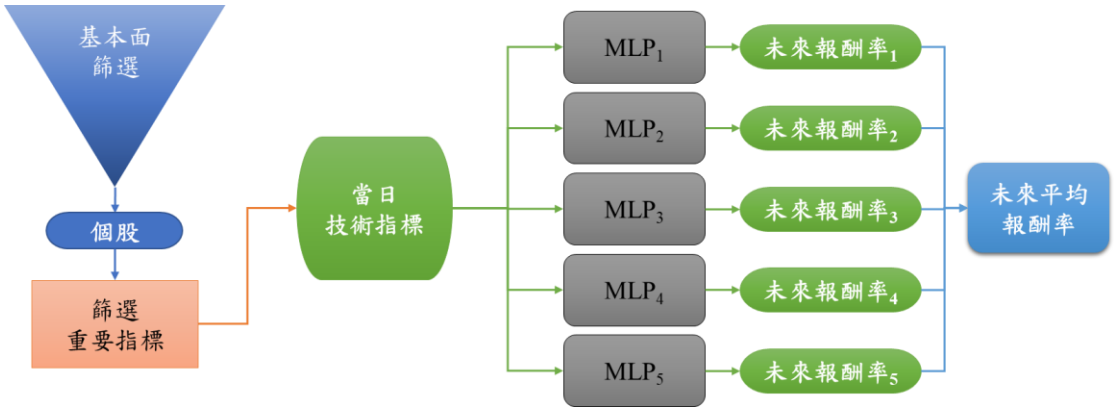

圖 6、5MLP 模型架構

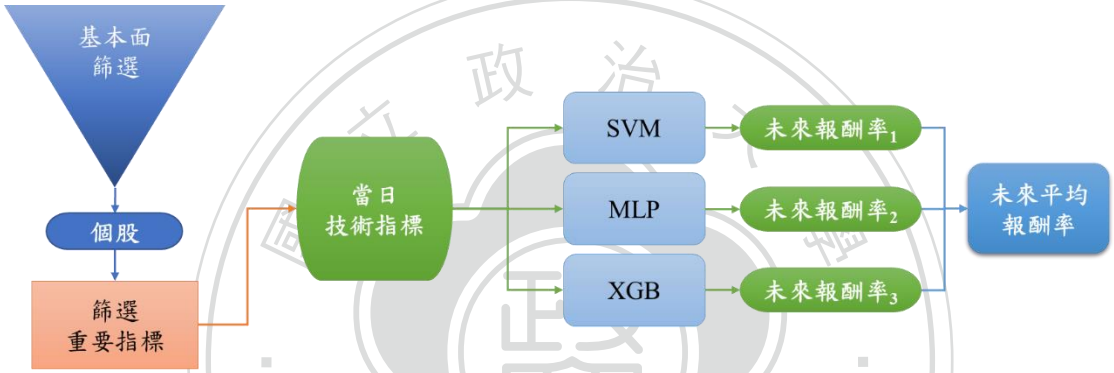

圖 7、SMX 模型架構

第四節　權重配置方法

在經過基本面篩選以及集成學習篩選後，已經挑出具有獲利潛力之個股，本研究除了比較不同集成學習模型對於回測績效的貢獻外，也探討不同權重配置下對於回測績效的影響，包含等權重配置（Equal Weight）、最小風險組合（Global Minimum Risk）及最佳效率組合（Tangency Portfolio）。

根據美國經濟學家 Markowitz（1952）所提出的效率前緣（Efficient Frontier）理論，透過形成投資組合並搭配不同權重的配置可以達到在相同風險下可獲得最高的預期報酬率或在相同預期報酬下可減少總風險到最低的效益。最小風險組合是採用效率前緣當中風險最小的投資組合，但同時也是效率前緣

DOI:10.6814/NCCU202100808

當中預期報酬率最小的投資組合；最佳效率組合是效率前緣當中夏普比率最高的投資組合，如圖 8 所示，也就是圖形上一直線經過無風險資產並與效率前緣相切之該點的投資組合；等權重則是最為直覺的權重配置，對於所有股票進行相同比例的權重配置。

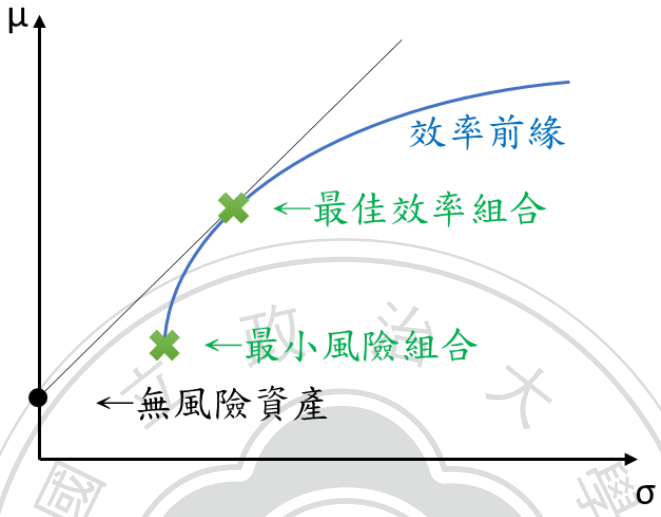

圖 8、效率前緣與權重配置

第五節　績效分析指標

為了有效分析整個回測的績效，針對本研究採用的數個績效指標進行說明。

一、　累積報酬率

$$累積報酬率 = \prod_{i=1}^{n}(1 + r_i) - 1$$

其中 r_i 為回測期間投資組合之日報酬率，n 為總回測日數。

二、　年化報酬率

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

DOI:10.6814/NCCU202100808

其中 n 為總回測日數。採用幾何平均數的方式計算年化報酬率。

三、 平均報酬率

$$\text{平均報酬率} = \left(\frac{1}{n}\sum_{i=1}^{n} r_i\right) * 252$$

其中 r_i 為回測期間投資組合之日報酬率，n 為總回測日數。採用算術平均數的方式計算平均報酬率。

四、 年化標準差

$$\text{年化標準差} = \sqrt{Var(r_i) \times 252}$$

其中 r_i 為回測期間投資組合之日報酬率。

五、 夏普比率

$$\text{夏普比率} = \frac{\textit{平均報酬率} - \textit{無風險利率}}{\textit{年畫標準差}}$$

六、 最大回撤率

$$\text{最大回撤率} = \min_{m<n}\left(\frac{\textit{第}n\textit{天累積資產}}{\textit{第}m\textit{天累積資產}}\right) - 1$$

其中 m, n 為回測期間內任意天數，表示在回測期間可能產生的最大虧損率。

DOI:10.6814/NCCU202100808

第四章　實證分析

本章節將透過第三章建構投資組合以及回測方式進行實證績效分析。本研究的資料來源為台灣經濟新報資料庫（TEJ），採用所有上市公司的資料，資料期間為 2007 年 1 月 2 日到 2021 年 4 月 22 日，回測期間從 2012 年 1 月 2 日開始，由於集成學習對於長天期的預測準確率較低，且過於頻繁交易會受到交易成本的懲罰，因此本研究選擇 10 天作為再平衡期間，也就是每過 10 個交易日就會調整一次投資組合。在基本面交集法的權重比例上，本研究目標將個股數篩選至約 100 檔，最終採用前 25%的本益比及前 25%的市價淨值比做為篩選基準，可以將所有個股篩選至 80~100 檔。最終在進行投資組合配置時，本研究挑選未來預期報酬率最高的前 10 檔作為投資組合之成分股，隨後進行三種不同的權重配置觀其績效。

以下將依據三個模型分別呈現其實證結果。

第一節　SVR_MLP 模型實證結果

採用 SVR_MLP 模型的實證結果列於表格 1，其歷史累積報酬率如圖 9 所示。其中報酬指數為大盤指數（TWSE）的歷史累積報酬率，基本面代表僅經過基本面交集法後，針對篩選完的 80~100 檔股票直接用等權重組成投資組合進行回測。隨後尚有三個 SVR_MLP 模型的回測績效搭配不同權重配置之結果，EW 代表等權重配置、GMR 為最小風險組合、TGC 則代表使用最佳效率組合。

在回測績效可以發現，當採用純基本面交集法篩選的回測結果便有這超越大盤指數的成果，顯示出基本面交集法之績效所在。當進一步運用到 SVR_MLP 模型時，無論是在何種權重配置下，其績效均勝過大盤指數及純基本面交集法。比較三種不同的權重配置方法，在累積報酬率方面，可以發現到最佳效率組合的報酬率比最小風險組合的報酬率高，而等權重的報酬率為最

DOI:10.6814/NCCU202100808

投資組合名稱	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回徹率
報酬指數	234.49%	14.34%	14.42%	0.1429	1.0097	-28.55%
基本面	419.57%	20.06%	19.06%	0.1236	1.5418	-30.23%
SVR_MLP (EW)	461.91%	21.11%	20.46%	0.1604	1.2757	-28.65%
SVR_MLP (GMR)	713.41%	26.06%	24.36%	0.1541	1.5801	-27.21%
SVR_MLP (TGC)	725.17%	26.39%	24.95%	0.1743	1.4320	-28.09%

表格 1、SVR_MLP 模型實證績效

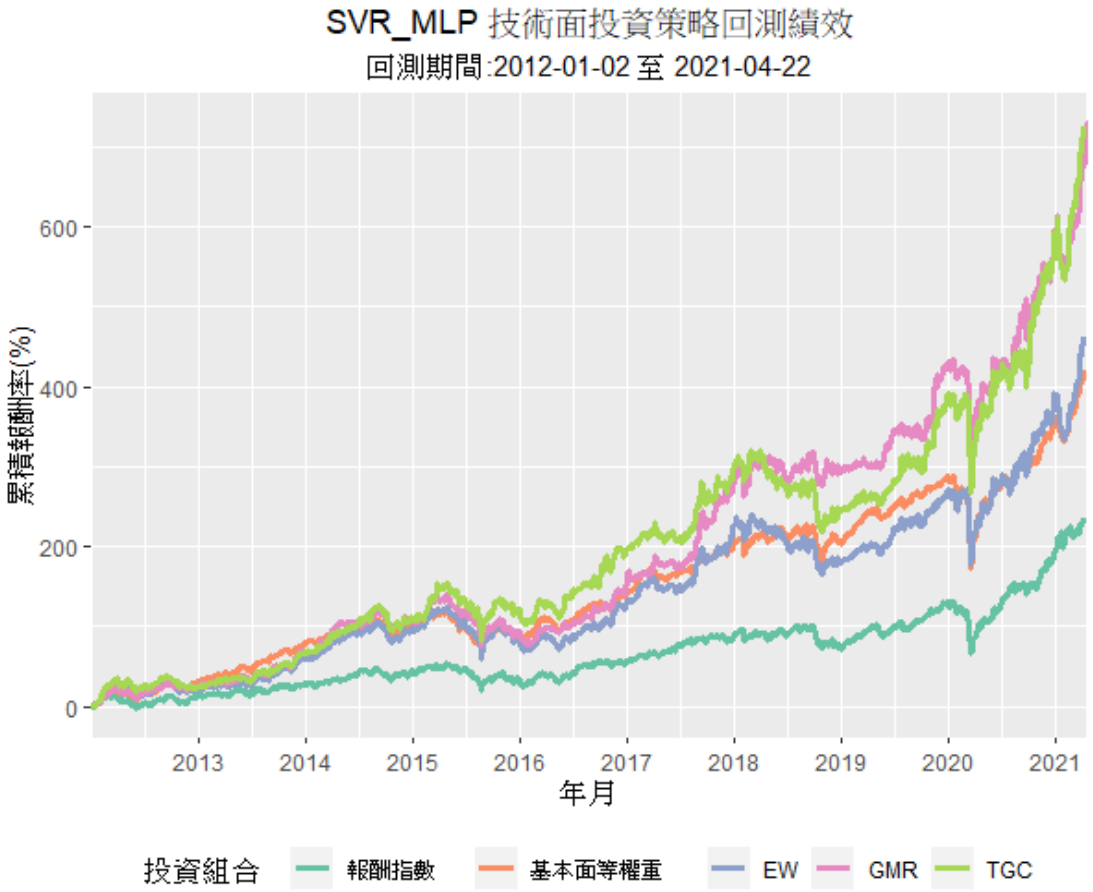

圖 9、SVR_MLP 歷史累積報酬率走勢圖

DOI:10.6814/NCCU202100808

低；但在風險控管方面，無論是從年化標準差或是最大回測率，亦或是透過圖形觀察皆可發現最小風險組合的風險最低，表示採用這種權重配置的方法有助於降低整體的投資風險。然而在夏普比率表現的部分，原則上應該是最佳效率組合的夏普比率較高，但實證結果卻是最小風險組合最高，由於兩者的累計報酬相去不遠，但最小風險組合的年化標準差較低，以至於在最終結果呈現與理論相反的結果。

第二節　5MLP 模型實證結果

採用 5MLP 模型的實證結果列於表格 2，其歷史累積報酬率如圖 10 所示。同樣包含大盤指數、基本面交集法以及 5MLP 模型搭配三種權重配置下的結果。

在績效表現方面，可以看到最大效益組合的累積報酬率最高，高過大盤指數及純基本面，但最小風險組合的報酬率卻是全部的報酬率當中最低的，甚至低於大盤指數，而等權重配置則是略遜於基本面的績效；在風險控管方面 GMR 是三種權重配置方法當中標準差最小的，而由於最佳效率組合追求高報酬，以至於風險最高；最後在夏普比率的方面，最佳效率組合雖是三種全種組合當中最高的但卻低於純基本面的夏普比率。

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202100808

投資組合名稱	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回徹率
報酬指數	234.49%	14.34%	14.42%	0.1429	1.0097	-28.55%
基本面	419.57%	20.06%	19.06%	0.1236	1.5418	-30.23%
5MLP (EW)	388.09%	19.14%	18.79%	0.1589	1.1826	-33.85%
5MLP (GMR)	218.81%	13.67%	13.93%	0.1490	0.9348	-32.48%
5MLP (TGC)	560.78%	23.20%	22.62%	0.1868	1.2109	-33.71%

表格 2、5MLP 模型實證績效

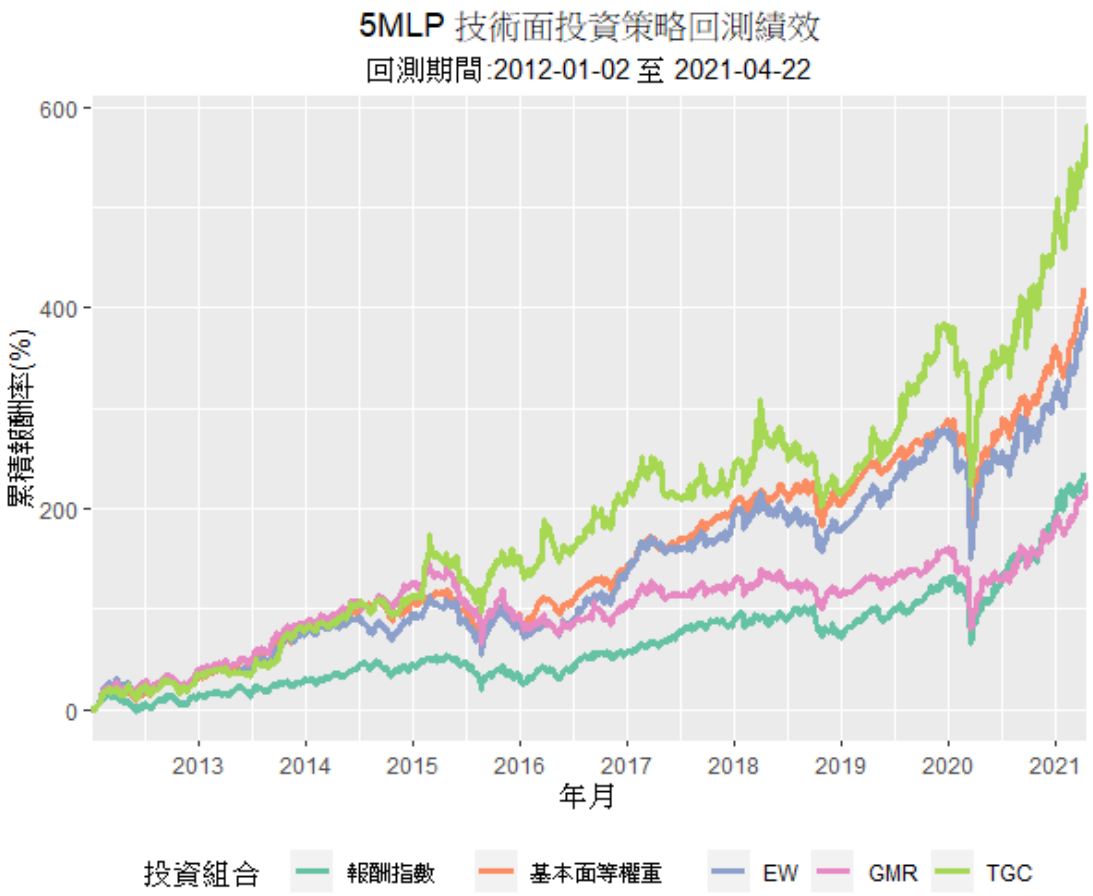

圖 10、5MLP 歷史累積報酬率走勢圖

DOI:10.6814/NCCU202100808

第三節　SMX 模型實證結果

採用 SMX 模型的實證結果列於表格 3，其歷史累積報酬率如圖 11 所示。同樣包含大盤指數、基本面交集法以及 5MLP 模型搭配三種權重配置下的結果。

在累積報酬率表現方面，採用 SMX 模型無論在哪一種權重配置當中，結落後於大盤指數；然而在三個權重配置當中，依舊是最佳效率組合的累積報酬率大於最小風險組合，最後是等權重；在風險控制方面依舊是最小風險組合的年化標準差最小；在夏普比率的部分則是最小風險組合最高，其原由應與 SVR_MLP 模型的情況相同。

投資組合名稱	累積報酬率	年化報酬率	平均報酬率	年化標準差	夏普比率	最大回徹率
報酬指數	234.49%	14.34%	14.42%	0.1429	1.0097	-28.55%
基本面	419.57%	20.06%	19.06%	0.1236	1.5418	-30.23%
SMX (EW)	146.46%	10.48%	11.42%	0.1698	0.6723	-36.05%
SMX (GMR)	200.34%	12.92%	13.46%	0.1614	0.8341	-36.80%
SMX (TGC)	210.93%	13.35%	14.27%	0.1859	0.7677	-38.71%

表格 3、SMX 模型實證績效

DOI:10.6814/NCCU202100808

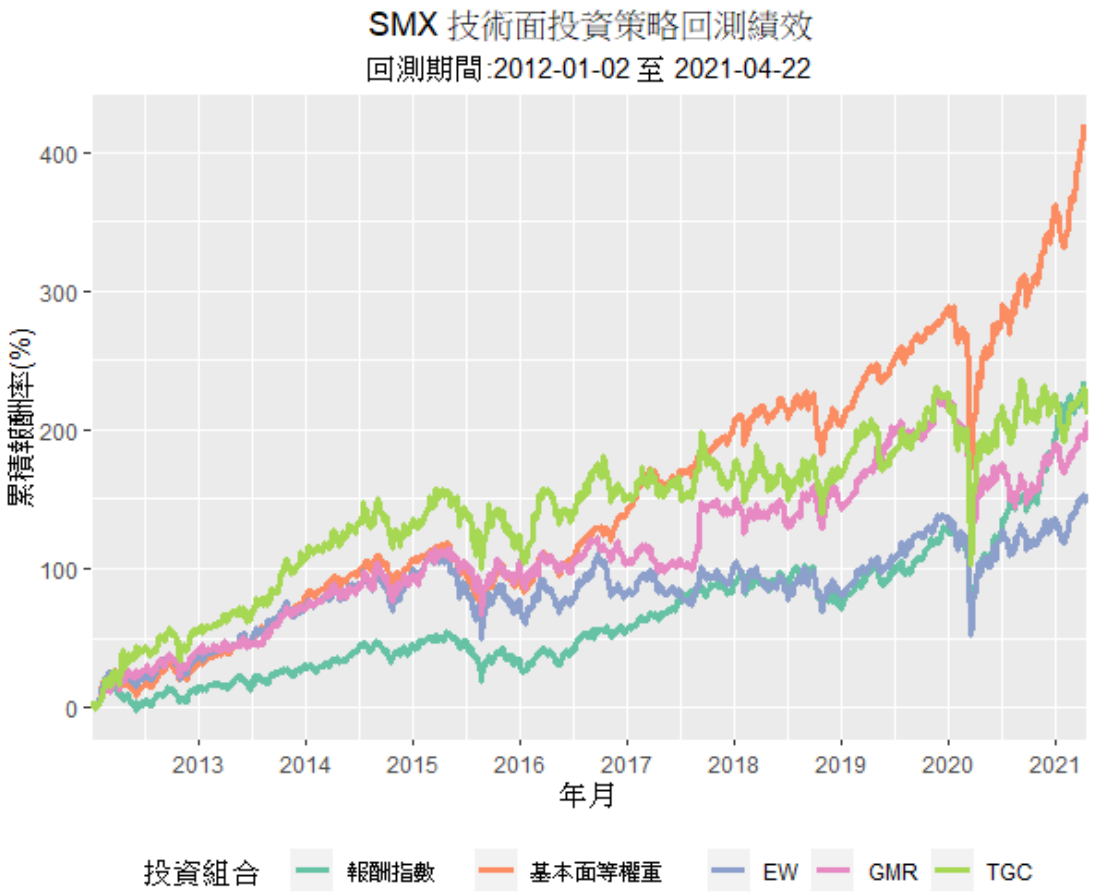

圖 11、SMX 歷史累積報酬率走勢圖

最後比較三種模型間的績效差異，可以發現在報酬率的部分，在相同權重配置下，SVR_MLP 模型的績效皆優於 5MLP 模型，而 SMX 模型均為墊底。可得出採用不同機器學習單元先後集成學習的效果優於相同機器學習單元平行集成學習的效果，而在相異機器學習單元平行集成學習的方面則須審慎考慮，留意彼此間的契合度，並探討是否有助於提升學習績效。

DOI:10.6814/NCCU202100808

第五章　結論與建議

本研究從台灣經濟新報資料庫（TEJ）蒐集臺灣上市公司的歷史股價資訊進行研究，首先藉由兩因子基本面交集法的方式先針對所有上市個股進行一次篩選，挑出具有潛力的個股，接著透過集成學習模型對股票的未來報酬率進行預測，並依據結果挑選出具有較高預期報酬率的個股組成投資組合，並透過不同的權重配置方式，針對投資組合做進一步的優化。

實驗結果證實，當採用 SVR_MLP 模型搭配最佳效率組合的權重配置，可以達到最高的報酬率回測績效。此外在三種權重配置的比較中，最佳效率組合有最高的報酬率表現，最小風險組合能夠降低整體投資風險，在夏普比率則是沒有一定趨勢。在三種集成學習模型的比較中，採用不同機器學習單元先後集成學習的效果優於相同機器學習單元平行集成學習的效果，而相異機器學習單元平行集成學習的效果最差。

本研究在進行雙因子基本面交集法時僅採用一種基本面組合，未來可以嘗試不同的基本面組合或是採用技術面或籌碼面指標作為篩選標的，或是針對本研究的 SVR_MLP 模型之架構，採用不同的機器學習單元進行替換，以提升整體的回測績效。

國立政治大學 National Chengchi University

DOI:10.6814/NCCU202100808

參考文獻

1. 張家瑋（2017）。利用籌碼面分析與關聯規則建構最適投資組合。國立政治大學碩士論文。
2. 蕭鈞銓（2016）。以基本面分析建構最適資產配置流程。國立政治大學碩士論文。
3. Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. The journal of Finance, 32(3), 663-682.
4. Chan, L. K., Hamao, Y., & Lakonishok, J. (1991). Fundamentals and stock returns in Japan. The journal of finance, 46(5), 1739-1764.
5. Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. International journal of forecasting, 5(4), 559-583.
6. Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). A machine learning model for stock market prediction. International Journal of Computer Science and Telecommunications, 4(12), 17-23
7. Huang, C. S., & Liu, 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.
8. Lev, B., & Thiagarajan, S. R. (1993). Fundamental Information Analysis. Journal of Accounting Research, 31(2), 190–215.
9. Markowitz H., (1952). Portfolio Selection. The Journal of Finance, 7(1), 77–91.
10. Mangalampalli, R., Pandey, V., Khetre, P., & Malviya, V. (2020). Stock Prediction using Hybrid ARIMA and GRU Models. International Journal of

DOI:10.6814/NCCU202100808

Engineering Research & Technology, 9(5), 737-743.

11. Sharpe, W. F. (1966). Mutual fund performance. The Journal of business, 39(1), 119-138.
12. Zhang, J., Li, L., & Chen, W. (2020). Predicting stock price using two-stage machine learning techniques. Computational Economics, 1-25.

國立政治大學
National Chengchi University

DOI:10.6814/NCCU202100808