

國立政治大學風險管理與保險學系

碩士學位論文

利用集成學習建構股市最適投資組合

Using Ensemble Learning to Construct

The Optimal Portfolio in Stock Market

指導教授：黃泓智 博士

研究生：林晏緯 撰

中華民國一一〇年六月

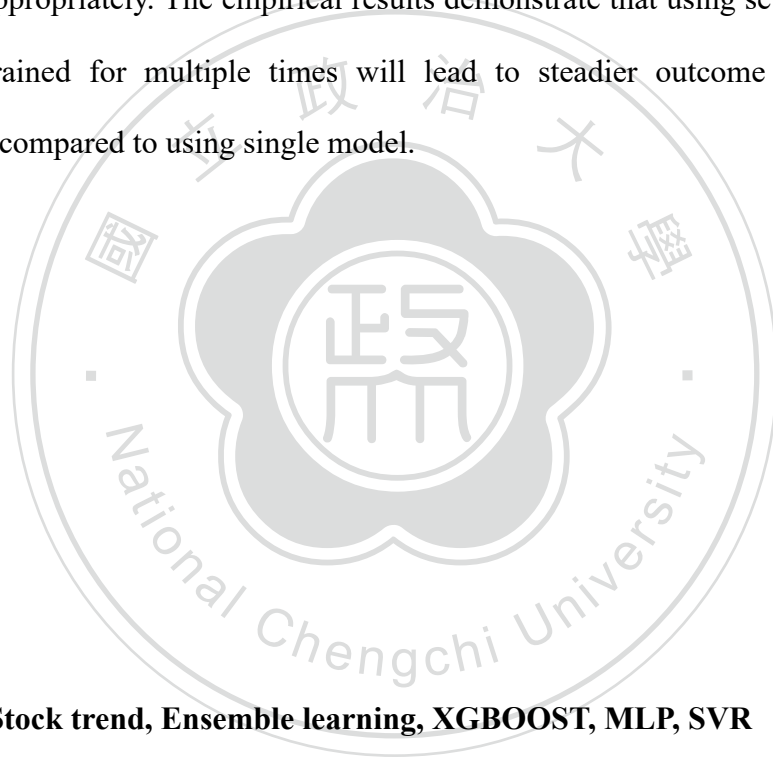
## 摘要

本研究使用台灣上市公司之財報資料以集成學習概念進行台灣股市個股漲跌預測，並建立最適投資組合。本研究使用多個不同的機器學習模型如極限梯度提升模型 (XGBOOST)、多層感知器 (MLP)、支持向量迴歸模型 (SVR) 等模型進行建模。為了使模型訓練結果更為穩定與準確，本研究使用上述模型進行多次訓練，選出各模型中上漲機率高的股票並對其進行綜合評分，接著組成股票投資清單，將評分高的股票進行權重配置建立投資組合。實證結果發現，相較於使用單一模型做一次的訓練，使用多種模型進行多次訓練後建立的投資組合能夠有更穩定的結果，且整體績效也優於單一模型。

**關鍵字：**股市漲跌、集成學習、極限梯度提升、多層感知器、支持向量迴歸

## Abstract

This dissertation aims to use ensemble learning to predict the trend of stocks in Taiwan stock market and build an optimal portfolio. The machine learning models used in the study include XGBOOST, MLP, and SVR. To make the model training results more stable and accurate, this study uses the above models for multiple trainings, and selects the stocks with high rising probability in each model and rate each of them comprehensively. Consequently, the optimal portfolio is built by allocating stocks with high rating appropriately. The empirical results demonstrate that using several models which are trained for multiple times will lead to steadier outcome and greater performance compared to using single model.



**Keywords: Stock trend, Ensemble learning, XGBOOST, MLP, SVR**

# 目次

<b>第一章 緒論</b> .....	<b>1</b>
第一節 研究動機 .....	1
第二節 研究目的 .....	2
第三節 研究流程 .....	3
<b>第二章 文獻探討</b> .....	<b>4</b>
第一節 機器學習模型文獻探討 .....	4
第二節 股價預測與機器學習模型文獻探討 .....	5
第三節 資產配置文獻探討 .....	6
<b>第三章 研究方法</b> .....	<b>8</b>
第一節 研究架構 .....	8
第二節 財報變數篩選 .....	10
第三節 機器學習模型 .....	13
第四節 集成學習選股 .....	22
第五節 資產配置策略 .....	25
第六節 績效指標說明 .....	27
<b>第四章 實證結果</b> .....	<b>28</b>
第一節 流動性篩選選股 .....	28
第二節 市值篩選選股 .....	33
<b>第五章 結論與未來方向建議</b> .....	<b>38</b>
<b>參考文獻</b> .....	<b>39</b>

## 表次

表 1、財務參數.....	11
表 2、財報發布截止日.....	12
表 3、模型訓練集期間與再平衡日.....	14
表 4、報酬率分群.....	15
表 5、XGBoost 參數設定.....	16
表 6、XGBoost 投組上漲比率.....	17
表 7、MLP 上漲比率.....	19
表 8、SVR 上漲比率.....	21
表 9、流動性篩選投資組合績效-1.....	30
表 10、流動性篩選投資組合績效-2.....	32
表 11、市值前 100 投資組合績效.....	35
表 12、市值前 100 各模型投資組合績效.....	35
表 13、市值前 100 名 XGB 與 MLP 綜合評分投資績效.....	37

## 圖次

圖 1、研究流程.....	3
圖 2、研究架構.....	9
圖 3、XGBoost 模型訓練流程圖 .....	16
圖 4、XGBoost 回測績效 .....	17
圖 5、MLP 模型訓練流程圖 .....	18
圖 6、MLP 回測績效 .....	19
圖 7、SVR 模型訓練流程.....	20
圖 8、SVR 回測績效.....	21
圖 9、單一種模型評分流程示意圖.....	22
圖 10、XGB 與 MLP 綜合評分示意圖.....	23
圖 11、綜合評分流程示意圖.....	24
圖 12、最適投資組合與最小風險投資組合.....	26
圖 13、流動性篩選建模與資產配置流程.....	28
圖 14、流動性篩選最適投資組合績效-1 .....	29
圖 15、流動性篩選最小風險投資組合績效-1 .....	29
圖 16、流動性篩選最適投資組合績效-2 .....	31
圖 17、流動性篩選最小風險投資組合績效-2 .....	31
圖 18、市值前 100 之建模與資產配置流程.....	33
圖 19、市值前 100 最適投資組合績效.....	34
圖 20、市值前 100 最小風險投資組合績效.....	34
圖 21、XGB 與 MLP 綜合評分最適投資組合績效.....	36
圖 22、XGB 與 MLP 綜合評分最小風險投資組合績效.....	36

# 第一章 緒論

## 第一節 研究動機

近年來由於科技的進步電腦算力提升，使得人工智慧 (Artificial Intelligence) 在各領域有著不錯的發展，如語音辨識、人臉辨識等應用，都已實現相當水平的精準度。而在金融業中，受惠於金融科技 (FinTech) 蓬勃發展，除了網路銀行、行動支付日益普及外，理財機器人在國際上也有許多不錯的發展案例，如美國的 Betterment 與 Vanguard 公司等推出理財機器人，為客戶進行資產管理。

相較於傳統投資人需要仰賴專業理財顧問進行資產配置，理財機器人能夠以更低的手續費為一般金融消費者進行資產配置規劃。而理財機器人之所以能夠為進行資產配置與投資規劃則是因為仰賴人工智慧在金融業的發展。由於金融業擁有許多金融資產的歷史資料，透過機器學習的方式使用金融資產的歷史資料建立模型預測金融資產未來走勢。

而在機器學習中，集成學習透過多種不同的學習方式來達成模型的訓練，集成學習能夠獲得比使用單一訓練模型還要好的預測效果。由於使用不同學習方式能夠截長補短，使訓練出來的模型更加完善。

Harry Markowitz 於 1952 年提出投資組合理論，認為可以透過投資不同的金融資產能夠有效降低非系統性風險，並提出效率前緣概念，透過不同權重配置找出最適的投資權重配置。

綜上所述，本研究欲使用集成學習概念對台灣上市公司之財報資料進行機器學習並預測公司下一季股價漲跌與報酬率，也就是使用多種不同的機器學習模型分別對台灣上市公司財報進行學習並預測股價趨勢，最後進行綜合評分建立投資清單。其概念就如同模擬多位專業投資人分析財報預測下一季各公司股價趨勢，並由該多位專業投資人最後進行投票選出下一季上漲股票之名單，並使用效率前緣的概念建立投資組合，期望能建立良好收益的股票投資組合。

## 第二節 研究目的

台灣目前的一般民眾在進行投資時，大部分透過手續費較高的理財專員或是自行投資。但從過去文獻與案例中可以發現大多數自行投資的民眾並無法有良好的投資績效甚至還會出現賠錢的狀況。而提供理財機器人給一般民眾進行投資理財規劃，則能夠解決高手續費與自行投資績效不理想之問題。故本研究希望透過機器學習的方式建立投資台灣股市之理財機器人，期望能夠達到優於大盤的投資績效，協助一般民眾進行投資規劃。

本研究欲達成以下研究目的：

- (一) 探討財報基本面選股，利用每一季財報參數如資產報酬率、股東權益報酬率等進行研究分析，找出財報中對模型預測較有幫助之參數。
- (二) 利用極限梯度提升模型 (XGBoost)、多層感知器 (MLP)、支持向量迴歸模型 (SVR) 三種模型分別對利用公司財報參數進行下一季股價趨勢預測，挑選出有上漲潛力之股票並進行回測，了解模型對股價的預測力。
- (三) 由上述三種模型對具上漲潛力之股票進行綜合評分，驗證多模型評分之績效與三種模型分開預測之績效差別。
- (四) 藉由上述方法挑選出綜合評分高之股票，透過效率前緣概念進行權重配置並組成投資組合，建置最適投資組合。



### 第三節 研究流程

本論文依據上述研究動機與研究目的，並透過台灣上市公司之資訊進行資料處理。本研究將分為 5 個章節，詳細敘述如下：

#### 一、緒論

針對本論文利用集成學習概念建構台灣上市股票最適投資組合之研究動機與研究目的進行闡述。

#### 二、文獻探討

回顧與本研究相關之文獻資料，包含集成學習與各種機器學習模型論文以及資產配置相關之論文，並從上述文獻提取對本研究有幫助之部分進行研究。

#### 三、研究方法

介紹本論文機器學習模型使用的變數與後續資料處理以及如何使用集成學習概念進行股票篩選與資產配置方式。

#### 四、實證結果

根據台灣上市股票進行投資策略分析，並比較單個機器學習模型與集成學習下的實證結果。

#### 五、結論與未來方向

針對上述實證結果進行探討，並探討本研究可進步之方向與未來建議。

研究架構圖如圖 1 所示：

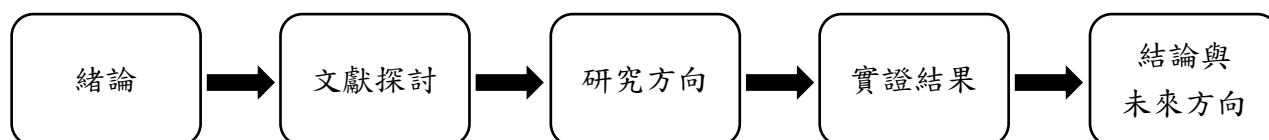


圖 1、研究流程

## 第二章文獻探討

本論文以集成學習概念，使用多個機器學習模型包含極限梯度提升模型、多層感知器、支持向量迴歸等機器學習模型進行股市預測，並建立最適投資組合。本章節一共會分為三個部分：第一節為本論文使用的機器學習模型與方法之相關文獻探討，第二節為機器學習與股價預測相關之文獻探討，第三節為資產配置相關之文獻探討。

### 第一節 機器學習模型文獻探討

Dietterich (2002) 提出集成學習概念，以多種基礎的演算法訓練模型，藉此提升預測結果準確度。集成學習解決了訓練模型常碰到的問題，如過度擬合 (Overfitting) 問題，當訓練集較大時常會使訓練的模型發生過度擬合，導致模型無法在測試集有好的發揮。演算法問題，在訓練模型時可能會因為演算法方式導致無法求得最佳模型，如梯度下降法可能會收斂在損失函數的局部極小值，而不是全域最小值。預測集問題，在機器學習演算法中，模型預測集中可能並沒有包含訓練集的最佳答案。

Chen et al. (2016) 提出極限梯度提升模型 (XGBoost)，使用決策樹 (Decision Tree) 的增強 (Boosting) 方法，透過疊層的方式改變樣本的權重，由前一次訓練來進行疊加，藉此改變樣本權重。並且將模型的複雜度加入至正則項中減少過度擬合問題。極限梯度提升模型自從被提出後，在數據分析競賽 Kaggle 成為常勝軍的機器學習模型，也因此極限梯度提升模型成為近年數據分析被廣泛使用的模型。

Gardner et al. (1998) 對多層感知器 (Multilayer Perceptron, MLP) 進行完整的介紹，多層感知器為向前傳導的人工神經網路。透過多個節點層進行學習，每個節點層都由一個激活函數 (Activation Function) 進行樣本學習的調整。多層感知器裡基本由三個層組成，分別為輸入層、隱藏層、輸出層。透過權重的分配的

調整，藉此找到損失函數的最小值。多層感知器在處理非線性系統上相較於傳統的方法能夠有較佳的表現。

Smola et al. (2004) 對支持向量迴歸模型 (Support Vector Regression, SVR) 對支持向量迴歸模型以及支持向量模型在當時的發展狀況進行了系統性的介紹。支持向量模型為支持向量機 (Support Vector Machine, SVM) 之延伸，其概念為透過將資料投影到高維度並建立超平面將資料進行分割。支持向量迴歸模型則是透過建置超平面進行迴歸，也就是說透過在高維度透過迴歸的方式找到最佳平面，藉此透過建置的超平面進行預測。而支持向量模型不論是分類或是迴歸預測在現今都有不錯的表現。

## 第二節 股價預測與機器學習模型文獻探討

Emerson et al. (2019) 對近期機器學習在計量金融領域的應用與趨勢進行整理，文中將應用領域分為三大領域，分別為投資組合配置、報酬率預測與風險相關應用，其中風險相關應用包含預測金融資產波動度、破產機率等。另外，文中在對預測報酬率部分進行的預測模型的統整，最常被用來進行報酬率預測的前三名模型分別為多層感知器、支持向量機與長多期記憶模型 (Long Short-Term Memory, LSTM)。而在三大金融應用領域中又以多層感知器為最廣泛使用。機器學習的出現給了複雜的金融相關問題一個可行的解決方法。

Khaidem et al. (2016) 使用隨機森林模型預測股價漲跌，透過對股價以及技術指標進行指數平滑 (Exponential Smoothing) 轉換，將此作為訓練資料來預測報酬率。使用此方法分別對三支科技類股進行預測一個月到三個月的漲跌機率。實證結果發現預測一個月後的股價漲跌平均有 80% 以上的準確率，而預測時間越長則有更高的準確率。表示長期而言，股價的漲跌趨勢能夠被技術指標捕捉。另外，作者在結論提到在股價漲跌預測中，與其使用人工神經網路或支持向量機等模型，訓練出一個最佳模型，使用類似投票機制的集成學習類型機器學習模型能夠獲得更好的訓練成效。

Alberg et al. (2017) 使用財報資訊進行深度學習藉此預測股市，作者認為長期而言財報基本面所提供的資訊如營業淨利、收入等參數，與股價呈現一定相關，因此能夠使用財報基本面的資訊進行標準化後進行深度學習的訓練。其結果表明使用多層感知機所訓練出來的模型其績效為年報酬率 17.1%，相比傳統因子模型年報酬率 14.4% 來的好。

Basak et al. (2019) 使用決策樹類型的模型進行股市漲跌的預測，其決策樹模型有隨機森林模型以及極限梯度提升模型。訓練資料採用技術面指標，如相對強弱指標、平滑異同移動平均線 (Moving Average Convergence Divergence, MACD) 等常用指標進行模型的學習。作者對其進行了短期至長期的預測，從預測三天、五天後漲跌到預測六十天、九十天後漲跌不等的時間長度。結果表明時間越長，準確度也越高，預測三天後漲跌的準確率約在 55% 至 60% 之間，而預測九十天後的漲跌則可以來到約 90%，表示對於機器學習模型而言長期的預測將可以帶來較好的預測效果。

### 第三節 資產配置文獻探討

Markowitz (1952) 提出投資理論，裡面闡述了平均投資報酬率與投資報酬率波動度之間的關係。人們在金融市場所追求的是較高的投資報酬率與較低的報酬率波動度。在此延伸出所謂的效率前緣 (Efficient Frontier) 投資組合，指在相同的報酬率下報酬率波動度最低的投資組合，又或是在相同的波動度下最高報酬率的投資組合。該理論闡明分散投資能夠降低報酬率的波動度，也就是說通過投資組合能夠有效的降低風險。

Clarke (2006) 使用 1968 年至 2005 年的美國股市資料進行實證研究，發現使用最小方差投資組合 (Minimum Variance Portfolio) 比市場投資組合擁有更低的標準差與更高的平均報酬率。也就是說實證研究表明效率前緣上面投資報酬率標準差的最小值能夠擁有較高的夏普比率 (Sharpe Ratio)，而在近四十年的回測中，最小方差的投資組合也擁有較高的累積報酬率。

Tokat et al. (2007) 使用 1960 年至 2003 年美國市場資料進行投資組合再平衡的實證研究，其結果顯示資產定期的再平衡是有其必要性的。而再平衡日數則是要視資產屬性與風險屬性而定。其中作者以六成的股票與四成的債券組成投資組合進行每月、每季、每年再平衡以及不曾再平衡的模擬，發現每季再平衡能夠獲得最高的超額報酬，而每個月再平衡一次則會有最低的超額報酬。





## 第三章研究方法

本章共分為六節，第一節為研究架構，說明本論文之研究流程。第二節為財報變數篩選，本論文使用台灣公司之財報進行股價漲跌預測，本節將對財報變數篩選進行說明。第三節為機器學習模型，本論文使用極限梯度提升模型、多層感知器與支持向量迴歸進行預測，本節將說明上述模型訓練過程。第四節為集成學習選股，本節將討論多種模型訓練完成後如何合併模型預測結果並進行選股。第五節為資產配置策略，本節將說明選股完後各檔股票其投資配置權重方法。第六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

本論文之研究架構圖如圖 2 所示，首先在第二節的部分，先對財務報表進行簡單介紹，其中包含台灣公司財報發佈日期以及財報資訊。並從中篩選出對機器學習模型有用之變數進行討論。

第三節介紹本研究使用的機器學習模型以及所採用的訓練期間與預測期間，本論文機器學習模型總共使用了極限梯度提升模型 (XGBoost)、多層感知器 (MLP) 與支持向量迴歸模型 (SVR)。本論文將會介紹上述三種機器模型並且分別使用此三種模型對台灣上市公司股價進行多次漲跌與報酬率預測並進行回測，探討三種模型對台灣股市之預測力。

在使用上述三種模型分別預測後，第四節將討論如何使用集成學習概念選股。本節將使用上述以訓練完成之模型對台灣上市股票進行綜合評分。透過上述模型之預測結果進行投票，藉此來為個股票進行評分選出下一季上漲潛力較高之股票。

在選完上漲潛力較高之股票後，第五節將討論如何使用上一節所選出上漲潛力高之股票進行資產配置。本研究使用 Markowitz (1952) 提出之投資理論概念進行資產配置。其中會使用最小風險投資組合 (Global Minimum Risk Portfolio) 進行以及最適投資組合 (Optimal Portfolio) 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如累積報酬率、年化報酬率、年化標準差、夏普比率、最大回撤率等績效指標，檢驗不同模型下之績效表現。

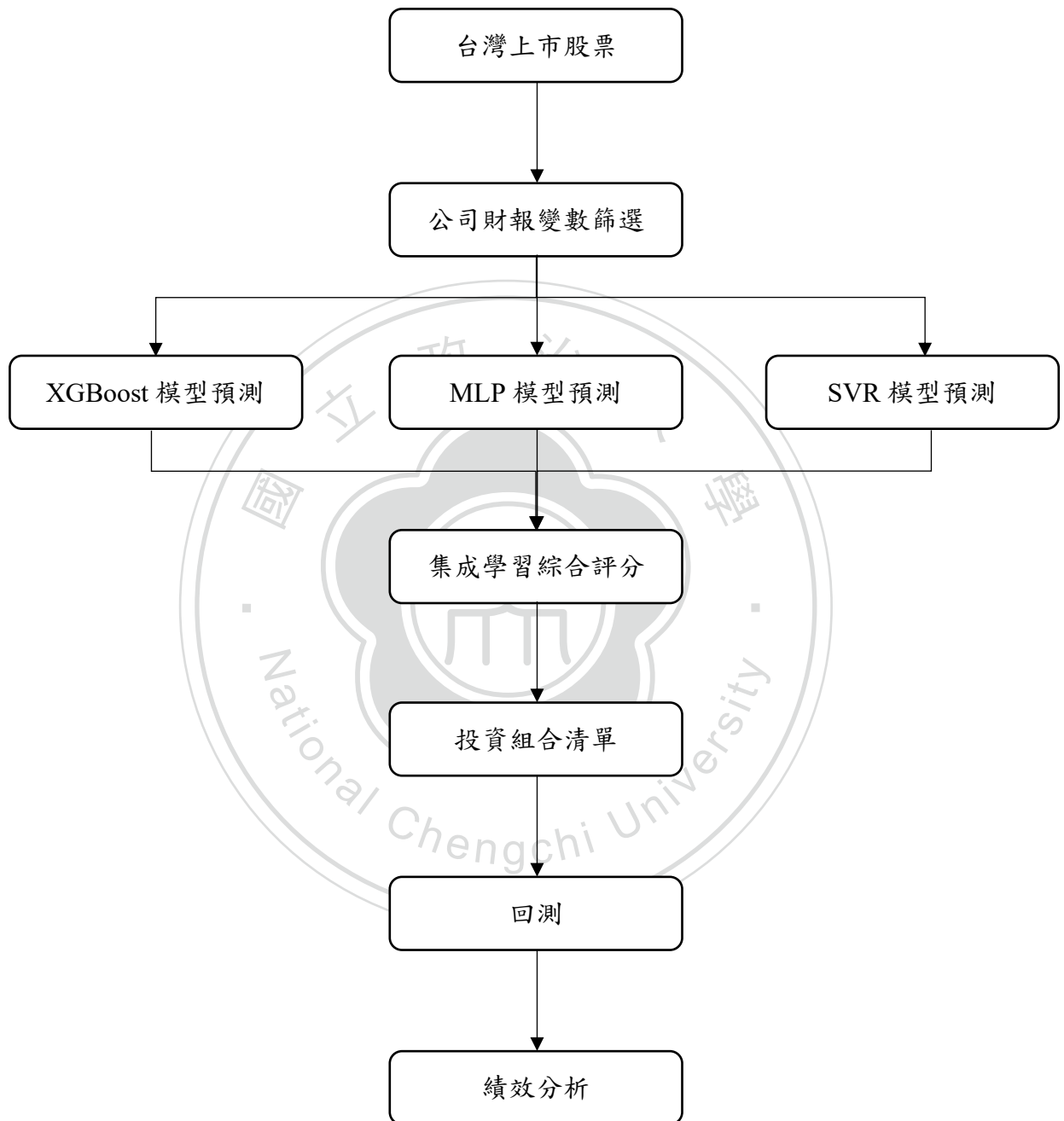


圖 2、研究架構

## 第二節 財報變數篩選

台灣上市公司接必須每一季相對應的時間公布財報，透過揭露財報相關數值能夠使投資人了解公司目前營運狀況以及公司體質。另外，也有許多文獻指出股票價值與公司營運狀況密切關係。因此財務報表所提供的資訊是判斷股價未來漲跌的重要依據。本節將討論本研究使用的財報變數，並分為財務結構指標、損益指標以及其他財務比率進行討論，而本研究詳細所使用之財務指標將在表 1 當中呈現。

### 一、 財務結構指標

公司之財務結構指標能夠反映公司營運狀況以及公司財務狀況，當市場投資人想了解公司發展現況時可以從財務結構指標了解。而常見的財務結構指標有現金比率、存貨比率、應收帳款比率等。

### 二、 損益指標

公司之損益指標能夠反映出公司的獲利能力以及公司每年給配息的狀況，此指標對於股價之影響力較大，因市場投資人對於公司是否有在未來有更好的獲利能力相當看重，常見的損益指標有股東權益報酬率、資產報酬率等。

### 三、 其他財務比率

其他財務比率如流動比率、速動比率、研發費用率、存貨周轉率，其代表的是產業上的特性以及公司發展現況。投資人可以透過檢視上述財務比率了解公司屬性以及公司資金流動狀況。

而在財報的發布日期上因為台灣在民國 102 年採用國際財務報導準則(IFRSs)規範，故每季財報發佈時間有些許不同這將影響到下一節訓練集期間，表 2 呈現在民國 102 年前後公司財報發佈截止日之差異。



表 1、財務參數

財務結構指標	損益指標	其他財務指標
現金比率	累計每股營收	流動比率
應收帳款比率	單季每股營收	速動比
存貨比率	近四季每股營收	長期資金適合率
固定資產比率	累計毛利率	利息保障倍數
流動負債比率	單季毛利率	利息支出比率
總負債比率	近四季平均毛利率	研發費用率
	累計營業利益率	總資產周轉率
	單季稅後淨利率	存貨周轉率
	近四季平均營業利益率	固定資產周轉率
	累計稅後淨利率	權益周轉率
	單季稅後淨利率	應付帳款付現天數
	近四季稅後淨利率	
	累計每股盈餘	
	單季每股盈餘	
	近四季每股盈餘	
	累計 ROE	
	單季 ROE	
	近四季平均 ROE	
	累計 ROA	
	單季 ROA	
	近四季平均 ROA	

表 2、財報發布截止日

	第一季截止日	第二季截止日	第三季截止日	第四季截止日
102 年前	4 月 30 日	8 月 31 日	10 月 31 日	隔年 3 月 31 日
102(含)年後	5 月 15 日	8 月 14 日	11 月 14 日	隔年 3 月 31 日



### 第三節 機器學習模型

本節將討論本研究所使用之機器學習模型以及模型所使用之參數處理與建模過程。本節將分為兩個部分，第一部分將介紹模型所使用之參數以及資料前處理過程與訓練集與測試集期間。第二部分則會分別介紹本論文所使用的機器學習模型以及建模過程，本論文所使用的機器學習模型有極限梯度提升模型、多層感知器、支持向量迴歸。另外，在討論完每個模型後也會探討每個模型對台灣股市之預測力。

#### 一、 模型參數與資料期間

本研究機器學習模型所選用之參數如表 1 所呈現，使用的是每一季財報發布所公布之資訊以及當作訓練參數。其中除了當季所使用到的參數，本論文還考慮到了產業有淡旺季之分別。為了使模型不要產生誤判，訓練參數中還加入了前三季財報之參數進行學習如近四季 ROA 以及近四季 ROE 等等參數。

另外，為了使參數數值參考標準一致，在訓練模型之前會先將所使用之參數進行標準化。標準化方式如下：

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

其中， $x_i$  為該變數第  $i$  個資料之原始數值， $\mu_x$  為該變數之平均值， $\sigma_x$  為該變數之標準差， $x'_i$  代表該變數第  $i$  個資料之原始值經過標準化。

資料期間本論文使用之資料期間為 2008 年五月至 2021 年四月，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一季財報發布日股價之漲跌，也就是說訓練集長度為十年，但因為在 2008 年以前財報多為半年發布一次並不是每季發布所以有某一部份之訓練集長度會不足十年。本論文使用滾動視窗（Moving Window）法進行學習，表 3 為本研究訓練模型之訓練期間與再平衡日對照表。

表 3、模型訓練集期間與再平衡日

訓練起始日	訓練結束日	再平衡日
20080501	20150401	20150518
20080501	20150516	20150817
20080501	20150816	20151116
20080501	20151116	20160401
20080501	20160401	20160516
20080501	20160516	20160816
20080501	20160816	20161116
20080501	20161116	20170405
20080501	20170401	20170516
20080501	20170516	20170816
20080501	20170816	20171116
20080501	20171116	20180402
20080501	20180401	20180516
20080901	20180516	20180816
20081101	20180816	20181116
20090401	20181116	20190401
20090501	20190401	20190516
20090901	20190516	20190816
20091101	20190816	20191118
20100401	20191116	20200401
20100501	20200401	20200518
20100901	20200516	20200817
20101101	20200816	20201116
20110401	20201116	20210401

## 二、機器學習模型與建模流程

本論文使用三個模型極限梯度提升模型 (XGBoost)、多層感知器 (MLP)、支持向量迴歸 (SVR)。以下將詳細介紹各個模型以及建模過程與討論模型對於台灣股市之預測能力。

### (一) 極限梯度提升模型 (XGBoost)

#### ● 模型簡介

XGBoost 模型，全名 Extreme Gradient Boosting，為 Boosting 計算方法的其中一種，Boosting 概念為將許多弱分類器集成為一個強分類器。而 XGBoost 模型則是含有 Boosting 概念提升樹模型，將許多樹模型集合在一起行成一個強分類器。其概念為透過不斷地增加樹進行學習，每新增一棵樹就相當於是學習了一個新的函數，藉此降低前一次學習後的殘差。以下式子為 XGBoost 模型的預測函數：

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)$$

$\hat{y}_i^{(t)}$  代表第  $t$  次迭代後第  $i$  個樣本的預測結果， $f_t(x_i)$  為第  $t$  棵樹的模型函數。

#### ● 建模流程

本研究使用 XGBoost 模型預測股價在再平衡日到下一次財報發布日時之股價漲跌，但並非以報酬率正負當作漲跌標準，而是將報酬率分為三群，分別為報酬率分為上漲、不漲不跌、下跌。透過將報酬率分為三群則可以減少模型對於上漲下跌的誤判，只有在上漲 5% 以上才算是上漲，而下跌也是如此。表 4 為報酬率分群之狀況：

表 4、報酬率分群

	下跌	不漲不跌	上漲
報酬率門檻	小於 -5%	-5% ~ 5%	大於 5%
訓練集佔比	約 25%	約 50%	約 25%

為了避免過度擬合 (Overfitting) 問題，在建模過程中使用交叉驗證 (Cross-Validation) 的方式調整參數降低過度擬合問題。交叉驗證其概念為輪流將不同部分的訓練集當作測試集，藉此進行參數調整。本模型所使用之交叉驗證方法為較常見的 K-fold 交叉驗證法。其概念為將訓練集平均分成 k 個集合，將其中一個集合當作測試集並把剩下的 k-1 個集合當作是測試集，如此重複進行直到 k 個集合都當作測試集過。而本論文將 k 設定為 4，也就是說將訓練集平均分成 4 個集合並且分別把這 4 個集合當作是測試集進行參數調整。

參數調整的部分主要是調整 max depth，max depth 為 XGBoost 模型樹的最大深度，深度越深則代表將樣本拆分的更細。另外，本模型還有調整其他的參數，表 5 呈現的為本研究 XGBoost 模型所使用參數的數值，沒有呈現在表中的參數則是使用模型預設值。

表 5、XGBoost 參數設定

eta	max depth	min child weight	gamma	subsample	alpha
0.1、0.2	3~10	0.5、1	0	1	0.05

最後，因為在每次進行交叉驗證時可能會出現模型選取之參數不同導致模型訓練結果不同，為了使模型更加穩定，所以透過重複訓練模型十次後並選取每個模型預測上漲機率最高之三十檔作為投資清單，最後再從該投資清單選取重複率高者當作投資標的進行回測。圖 3 為本模型訓練流程圖：

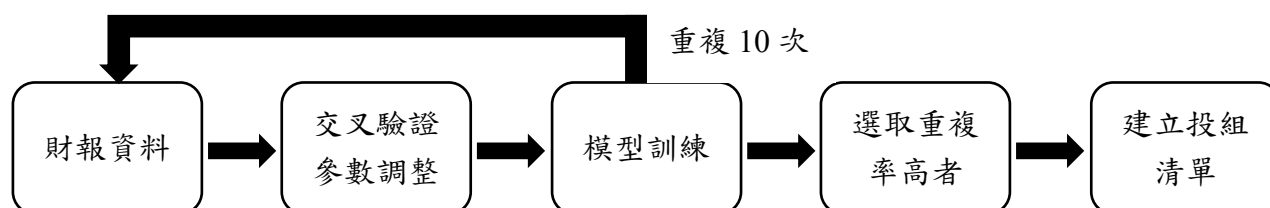


圖 3、XGBoost 模型訓練流程圖

● 模型預測力

由上述建模流程所得到的投資組合清單後，將投資組合清單進行回測後，在圖 4 的累積報酬率中可以看到重複率最高的前 10 檔、20 檔、30 檔所組合而成的投資組合清單累積報酬率表現，表 6 則是在回測期間每個投資組合內部各檔股票上漲的比率。

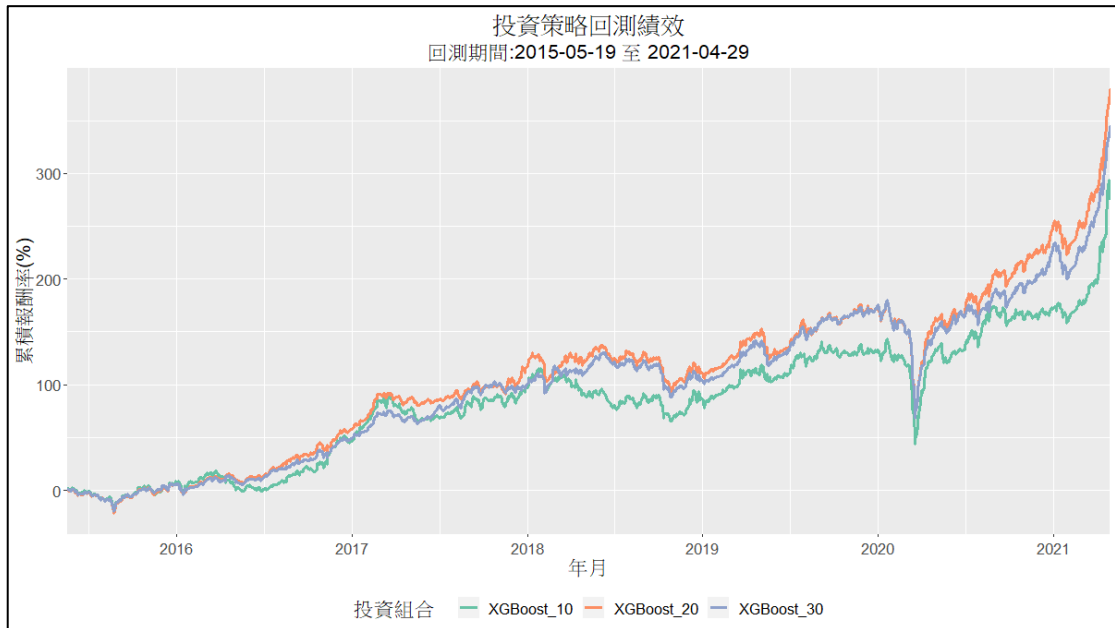


圖 4、XGBoost 回測績效

表 6、XGBoost 投組上漲比率

投資組合名稱	上漲比率
XGBoost_10	58.70%
XGBoost_20	59.48%
XGBoost_30	59.80%

## (二) 多層感知器 (MLP)

### ● 模型簡介

多層感知器 (MLP)，全名 Multilayer Perceptron，屬於深度學習模型。MLP 模型由輸入層、隱藏層、輸出層構成，將輸入層的資料經過線性合成後進入隱藏層，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，最後再透過反向傳遞進行參數調整。而反向傳遞就是利用損失函數進行參數的調整，損失函數通常都是使用均方差 (Mean square error)，如下面方程式所示：

$$E^{(i)} = \frac{1}{2} \sum_{j=0}^m (\hat{y}_j^{(i)} - y_j^{(i)})^2$$

其中  $E^{(i)}$  代表第  $i$  筆資料之誤差， $\hat{y}_j^{(i)}$  代表第  $i$  筆資料之第  $j$  個參數預測值， $y_j^{(i)}$  代表第  $i$  筆資料之第  $j$  個參數實際值。而 MLP 模型模標則是將誤差降到最低。

### ● 建模流程

本模型建模流程與 XGBoost 模型方式類似，皆是透過將報酬率分為上漲、不漲不跌、下跌三群，而分類方式則與表 4 相同都是以正負 5% 當作上漲與下跌的門檻。而在模型訓練上因為 MLP 參數調整時所使用的是梯度下降法 (Gradient Descent)，其概念為透過將損失函數透過偏微分找尋參數調整知方向藉此找到損失函數的局部極小值。但並不是每次找到的局部極小值都是全域極小值，為求模型更加穩定，MLP 模型訓練如同 XGBoost 模型訓練一樣皆是以訓練十次後找尋投組清單中重複率較高者當作投資標的。圖 5 為本模型之建模流程：

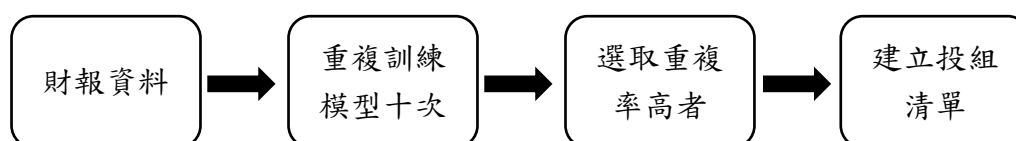


圖 5、MLP 模型訓練流程圖



● 模型預測力

如同 XGBoost 模型，在這裡也從每個 MLP 模型所選取具有上漲潛力之股票取出後找出重複率較高之股票，並以這些股票作為投資組合清單。在圖 6 的累積報酬率中可以看到重複率最高的前 10 檔、20 檔、30 檔所組合而成的投資組合清單累積報酬率表現，表 7 則是在回測期間每個投資組合內部各檔股票上漲的比率。

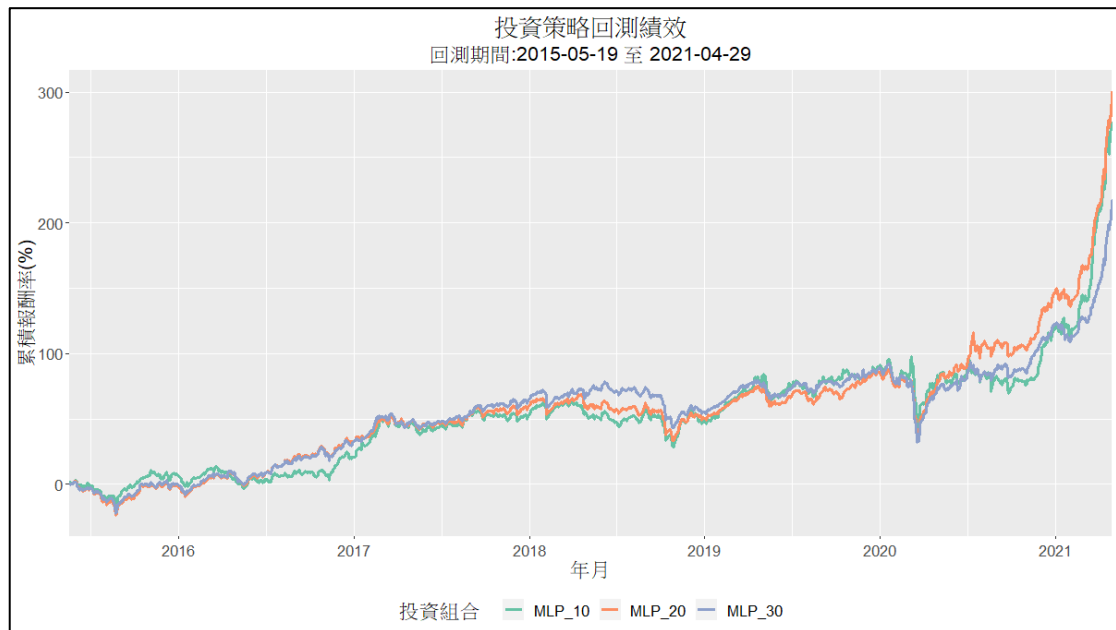


圖 6、MLP 回測績效

表 7、MLP 上漲比率

投資組合名稱	上漲比率
MLP_10	57.64%
MLP_20	57.52%
MLP_30	58.66%

### (三) 支持向量迴歸 (SVR)

#### ● 模型簡介

支持向量迴歸模型 (SVR)，全名 Support Vector Regression，為 SVM 模型 (Support Vector Machine) 的延伸，其概念為透過建置一個最適的超平面將資料進行分群。也就是說透過將資料透過核函數 (Kernel Function) 的轉換後找到一個超平面能夠將資料進行最適的分割，目標是能夠將資料分得越開越好。而 SVR 模型則與 SVM 相當類似，也是透過將資料經過核函數轉換後找到一個超平面。但在這裡的超平面則是希望能夠使所有資料對此超平面的加總距離越近越好。因此，我們能夠透過該超平面進行預測。

#### ● 建模流程

SVR 建模流程與 XGBoost 與 MLP 模型不同，在此因為 SVR 模型是具有迴歸概念的機器學習模型，並不像 SVM 模型是進行分類的機器學習模型。所以在這邊我們不是預測上漲與下跌，而是改為預測報酬率。而在此不選擇使用 SVM 進行漲跌預測的原因是因為本研究希望能夠使用集成學習之概念建置投資組合，所以應該盡量避免都使用預測漲跌的機器學習模型，希望透過增加一個預測報酬率的模型來學習到被前兩個模型忽略掉的特徵。

本研究 SVR 建模是透過訓練模型進行預測報酬率，再從預測報酬率較高之股票中挑選出數檔進行回測。而 SVR 模型是透過資料建置最適超平面，並不會有模型訓練結果不穩定的問題，故在這裡 SVR 模型只會訓練一次。圖 7 為 SVR 模型的建模流程：

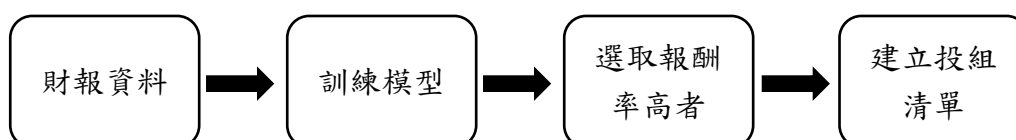


圖 7、SVR 模型訓練流程

● 模型預測力

不同於前兩個模型，SVR 模型預測的是股票在再平衡日至下一次財報發布的報酬率，並透過選取較高報酬率的股票作為投資組合清單。在圖 8 的累積報酬率中可以看到重複率最高的前 10 檔、20 檔、30 檔所組合而成的投資組合清單累積報酬率表現，表 8 則是在回測期間每個投資組合內部各檔股票上漲的比率。

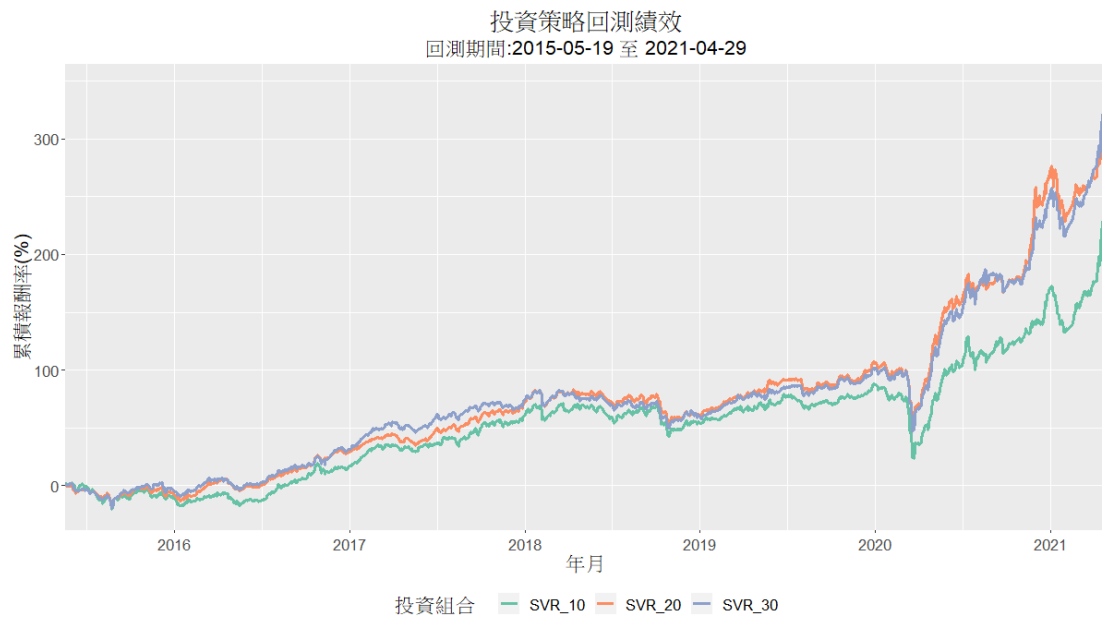


圖 8、SVR 回測績效

表 8、SVR 上漲比率

投資組合名稱	上漲比率
SVR_10	60.70%
SVR_20	59.47%
SVR_30	57.62%

## 第四節 集成學習選股

本節將討論如何使用集成學習概念建立投資組合清單，集成學習的概念為透過結合多個不同的機器學習模型，解決單個模型常遇到的問題如過度擬合、訓練結果不穩定等。第一個部分會先討論單一種模型的評分，也就是前一節 XGBoost 模型與 MLP 模型訓練十次後建置投資組合的詳細過程。而在第二部分則是承接第一部分的內容討論多個模型的綜合評分。將前一節三個模型預測的結果進行整合，透過三個模型的預測結果為上漲潛力較高的股票進行綜合性的評分，希望能夠建立比使用單一模型還要好的投資組合。

### 一、單一種模型評分

在前一節的 XGBoost 與 MLP 模型建模流程中本研究採取重複訓練模型十次來增加模型預測之穩定度。而在訓練完十次後詳細的選股方式如圖 9 所示，XGBoost 與 MLP 模型在進行預測時會產出測試集落在上漲、不漲不跌、下跌三群中的機率。接著從每個模型中選出落在上漲群機率最高的前 30 檔股票，並給予這三十檔股票 1 分。將十個模型都重複上述步驟即可獲得上漲潛力高的股票的評分。



圖 9、單一種模型評分流程示意圖

## 二、多種模型綜合評分

在上述單一種模型評分可以看到評分方式是透過選取每單一個模型上漲機率前三十名高的股票，藉此建立出一個高上漲機率的股票清單。接著再依照各檔股票出現在清單的個數給予評分。而在多種模型評分也採取類似的概念為這些股票進行綜合評分。首先，本段落會先介紹 XGBoost 模型與 MLP 模型的綜合評分，再來介紹加入 SVR 模型的綜合評分。

在 XGBoost 模型與 MLP 模型的綜合評分中，如圖 10 所示，採用的是將單一種模型的評分表直接合併的方式來給予綜合評分。也就是將兩個模型所對應的股票評分表進行相對應的分數加總。

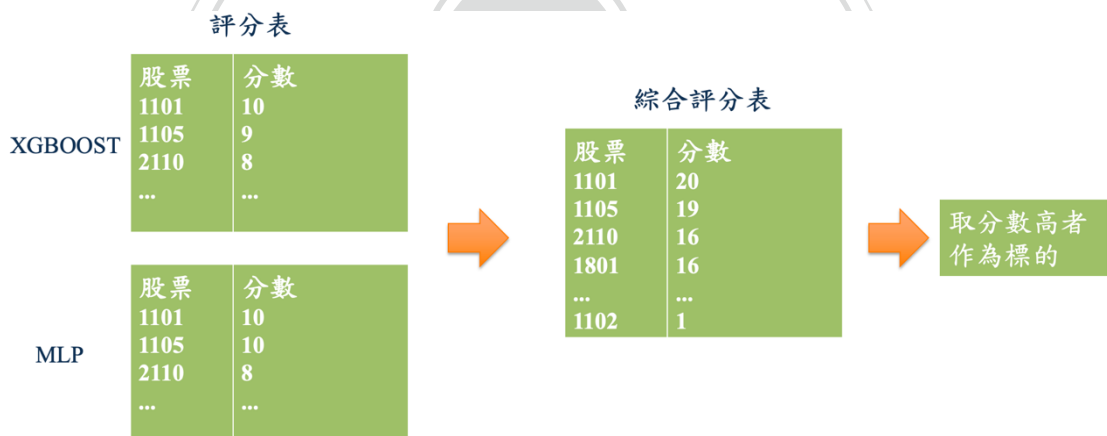


圖 10、XGB 與 MLP 綜合評分示意圖

上述為 XGBoost 與 MLP 模型綜合評分之方式，而在加入 SVR 之方式則與上述不同。因為在前一節 SVR 模型建模流程中，不像前兩種模型有訓練十次模型取而代之的是訓練一次 SVR 模型，並且該模型是用來預測股價的報酬率。因此在加入 SVR 模型的綜合評分中，如圖 11 所示，是透過將預測的報酬率進行由高到低的排名，並選取預測報酬率前五十名給予相對應的分數。為了讓 SVR 模型在評分中有一定的影響力，因此給予預測報酬率前十名之股票 5 分十一至二十名的股票給予 4 分，依此類推倒數十名則給予 1 分。最後，再將 SVR 的評分加

到 XGBoost 模型與 MLP 模型的綜合評分表上，組成最後三個模型的綜合評分表。在這邊的集成學習除了考量到補足不同模型的缺點外，還有考量到模型在預測報酬率與漲跌上可能會重視不同特徵的問題，也就是若只有使用模型預測漲跌可能會導致模型選出上漲率高但不一定報酬率高的股票。

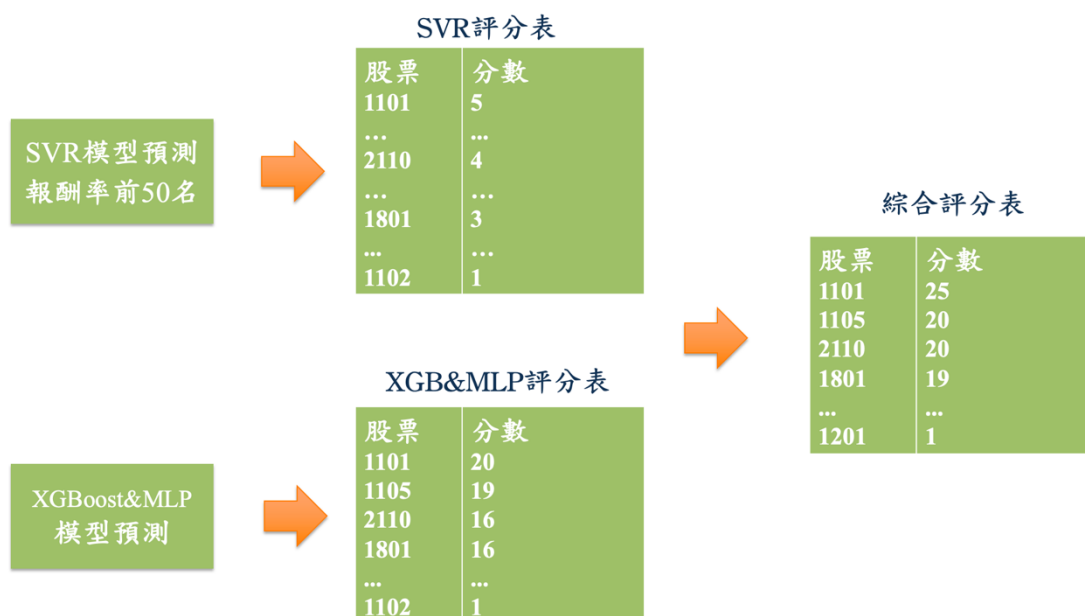


圖 11、綜合評分流程示意圖

## 第五節 資產配置策略

在第三節分別使用了三種模型預測股價至下一季財報發布日的漲跌以及報酬率，在第四節則是透過第三節的模型進行集成學習概念的綜合評分。接著本節希望從前一節選出高綜合評分之股票進行權重配置。本研究採取兩種權重配置方法，分別為平均-變異模型 (Mean Variance) 以及最小風險模型 (Global Minimum Risk)，希望透過兩種不同的配置權重方式能夠提供不同風險等級之投資組合。

美國經濟學家 Markowitz (1952) 提出投資組合理論，在假設市場投資人皆為風險趨避者，當不同投資組合有相同預期報酬率時，投資人將傾向選擇風險最小之投資組合。也就是說若有相同風險之投資組合，投資人將傾向選擇預期報酬率最高之投資組合。若投資人並非投資在上述兩種投資組合上，則代表該筆投資為不效率的投資。藉此則可以延伸出 Markowitz 提出的效率前緣 (Efficient Frontier) 概念，透過將上述所有效率的投資組合呈現在預期報酬率與風險的座標軸上則可以得到效率前緣，本研究之股票權重配置則會採用效率前緣上的投資組合，分別為平均-變異最優投資組合與最小風險投資組合。

平均-變異最優投資組合又稱為相切組合 (Tangency Portfolio)，為效率前緣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，因此也稱之為最大夏普比率投資組合 (Maximum Sharpe Portfolio)。最小風險投資組合為效率前緣中風險最小的投資組合，代表波動度最低的投資組合，也是在效率前緣中預期報酬率最低的投資組合。圖 12 為平均-變異最優投資組合與最小風險投資組合與效率前緣之關係，在效率前緣最左邊的點代表最小風險投資組合，而效率前緣與市場相切的點為平均-變異最優投資組合。



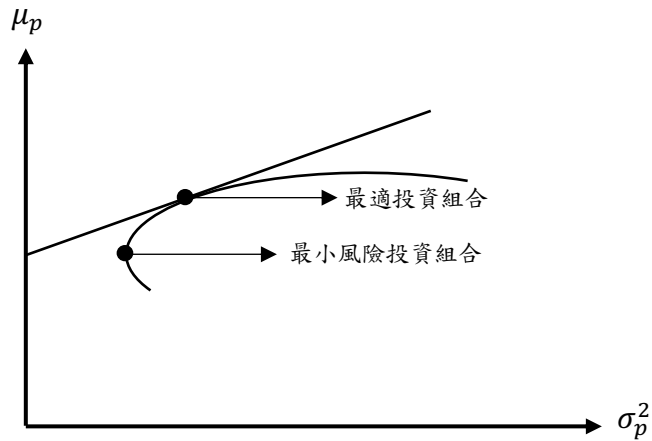


圖 12、最適投資組合與最小風險投資組合

本研究希望透過前一節股票綜合評分中選出前 10、20、30 名高分之股票，並使用本節所介紹的平均-變異最適投資組合以及最小風險組合進行投資組合權重配置。其中判斷投資清單中股票的波動度以及股票間之相關性是使用過去 500 個歷史交易日資料進行計算。另外，為防止某些股票權重配置過低，配置時限制每檔股票必須要配置 2 倍投資檔數倒數之權重。例如投資 10 檔股票則每檔股票最小投資權重必須要有 5%。



## 第六節 績效指標說明

### 一、 累積報酬率

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

其中  $n$  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

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

其中  $n$  為回測期間總交易日數。

### 三、 年化標準差

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

其中  $n$  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。

### 四、 年化夏普比率

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

其中  $E(R_p)$  為年化報酬率， $r_f$  為無風險利率， $\sigma_p$  為年化標準差。

### 五、 最大回撤率

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

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

## 第四章 實證結果

本章將呈現使用第三章研究方法之集成選股方式之實證結果，本章總共分為兩節。第一節先將流動性較差之股票進行篩選後，以篩選後之股票進行模型訓練建立投資組合並進行回測。第二節則是以台灣公司市值前 100 名作為篩選標準，透過兩種不同篩選方式得到兩種不同的投資策略，第一節為追蹤台股整體之投資策略，而第二節為追蹤大盤之投資策略。

### 第一節 流動性篩選選股

因在實務運作上投資人只能在市場上買到流動性較佳之股票，因此在建立投資組合時也應該考量到股票流動性的問題。本節分為兩種流動性篩選方式第一種篩選方式使用交易量與交易張數為篩選標準，第二種方式則為單純使用交易量為篩選標準。

第一種篩選方式使用單日交易張數需大於三百張或交易量為前 20% 之股票。雖然以台股歷史交易資料來看，單日交易張數大於三百張之標準已能夠避免選取到流動性差之股票，但卻會使一些流動性好但市值較高之股票被篩選掉，如大立光電股價在上千元其因為市值非常高導致交易張數較低，但這並不代表大立光電股票流動性差，所以在篩選交易張數後應將交易量在前 20% 的股票加回來。圖 13 為使用流動性篩選後的建模與資產配置流程。

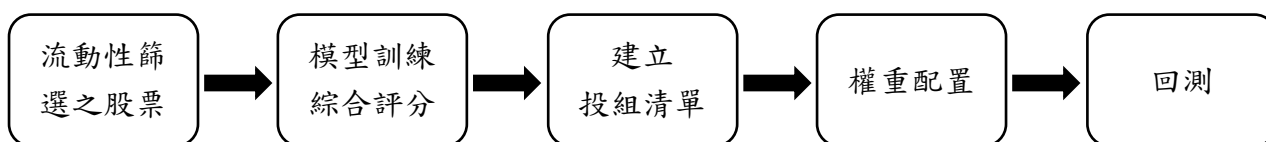


圖 13、流動性篩選建模與資產配置流程

圖 14 為使用第一種篩選方式之最適投資組合之回測績效，圖 15 則為最小風險投資組合回測績效，表 9 則為兩投資組合之績效。分別有選出 10 檔、20 檔與 30 檔之投資組合。

投資策略回測績效  
回測期間:2015-05-18 至 2021-04-29

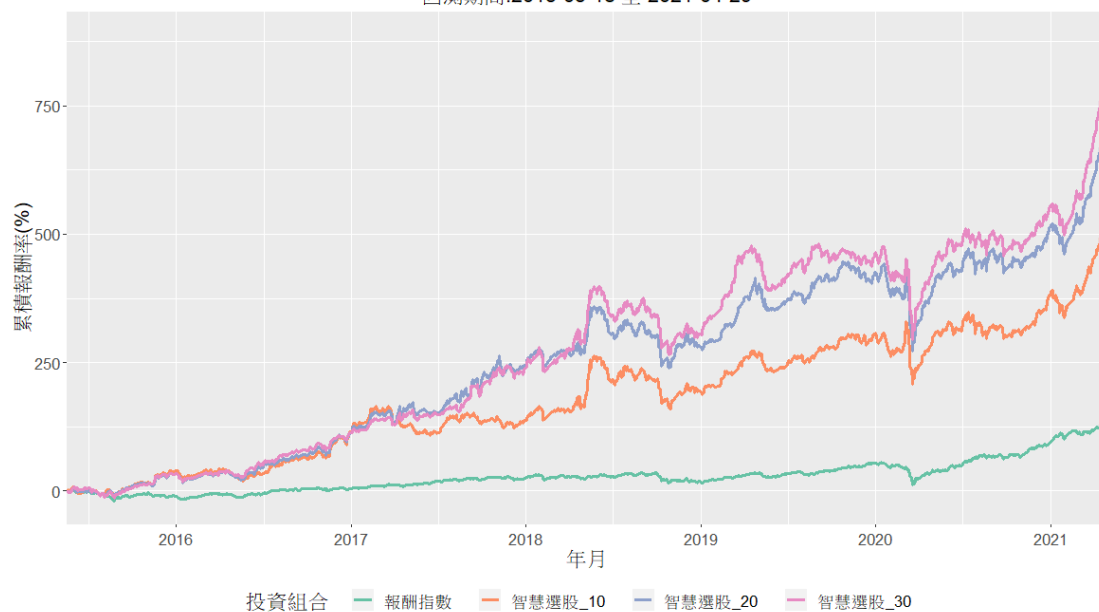


圖 14、流動性篩選最適投資組合績效-1

投資策略回測績效  
回測期間:2015-05-18 至 2021-04-29

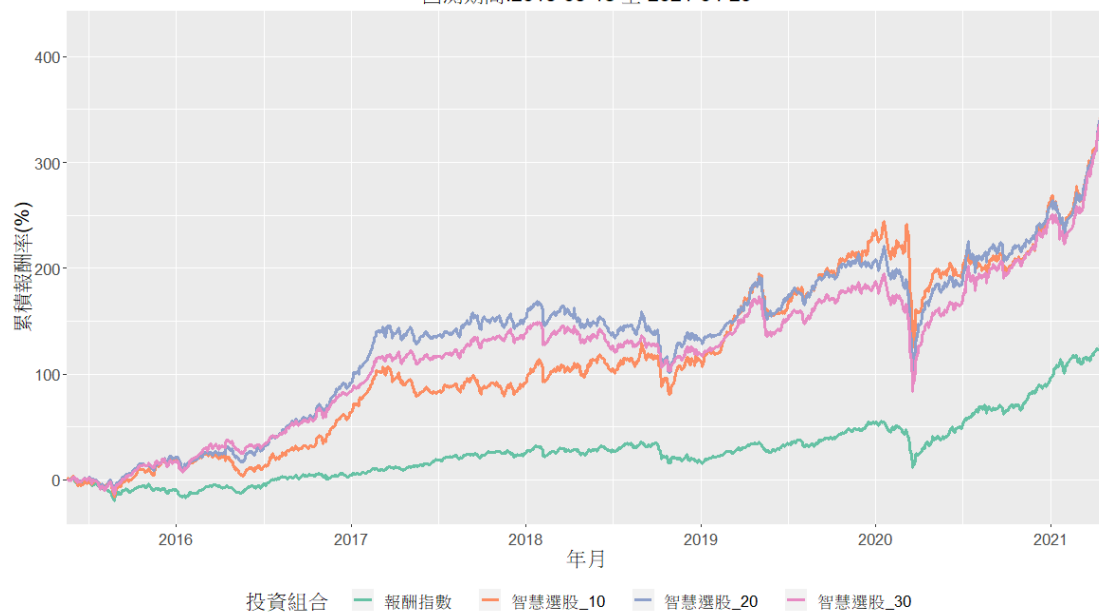


圖 15、流動性篩選最小風險投資組合績效-1

表 9、流動性篩選投資組合績效-1

投組名稱	累積報酬率	年化報酬率	年化標準差	夏普比率	最大回撤率
報酬指數	132.92%	15.75%	0.1508	1.0457	-28.55%
智慧選股 10_tangency	587.88%	39.62%	0.2529	1.4472	-28.57%
智慧選股 20_tangency	769.47%	45.40%	0.2449	1.6525	-31.77%
智慧選股 30_tangency	854.11%	47.76%	0.2350	1.7803	-32.06%
智慧選股 10_GMR	420.25%	33.03%	0.2187	1.4160	-34.60%
智慧選股 20_GMR	402.84%	32.25%	0.1972	1.5174	-37.72%
智慧選股 30_GMR	394.55%	31.87%	0.1838	1.5989	-38.00%

上圖與上表中的報酬指數為台灣股市大盤的績效，智慧選股中的 tangency 策略為最適投資組合也稱為相切投資組合，而智慧選股中的 GMR 策略為最小風險投資組合。從上圖與上表中都可以發現不論是最適投資組合還是最小風險投資組合其年化報酬率與夏普比率皆大於台灣股市大盤績效，另外，從上圖也可以看到兩種投資組合的累積報酬率走勢也是優於在前一章圖 4、圖 6、圖 8 單一模型下使用最小風險模型配置之績效走勢，代表使用集成學習方法能夠有效的提升投資組合之績效。

第二種方法為單純使用日交易量前 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。而此篩選法的好處在於能夠更加確保投資者在股票市場購買到相對應張數的股票。若未來希望建立理財機器人，將本研究之策略應用在實務上，則使用此篩選方法能夠更加確保實務上再平衡之可行性。圖 16 為使用第二種流動性篩選方式之最適投資組合績效，圖 17 為使用第二種流動性篩選方式之最小風險投資組合績效，表 10 為兩種投資組合之績效。分別選出 10 檔、20 檔與 30 檔之投資組合。

投資策略回測績效  
回測期間:2015-05-18 至 2021-04-29

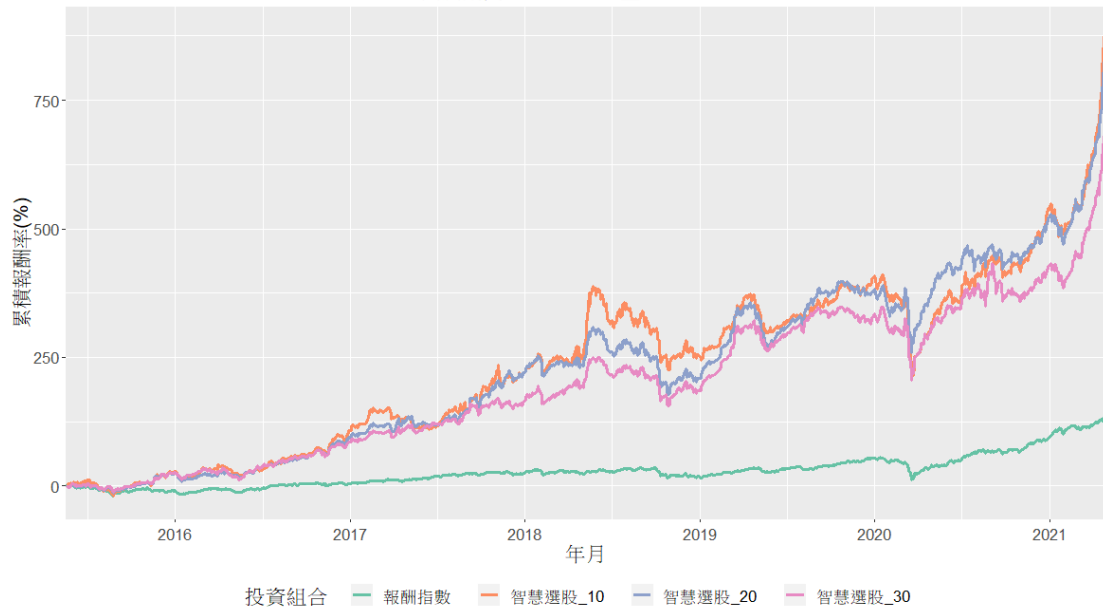


圖 16、流動性篩選最適投資組合績效-2

投資策略回測績效  
回測期間:2015-05-18 至 2021-04-29

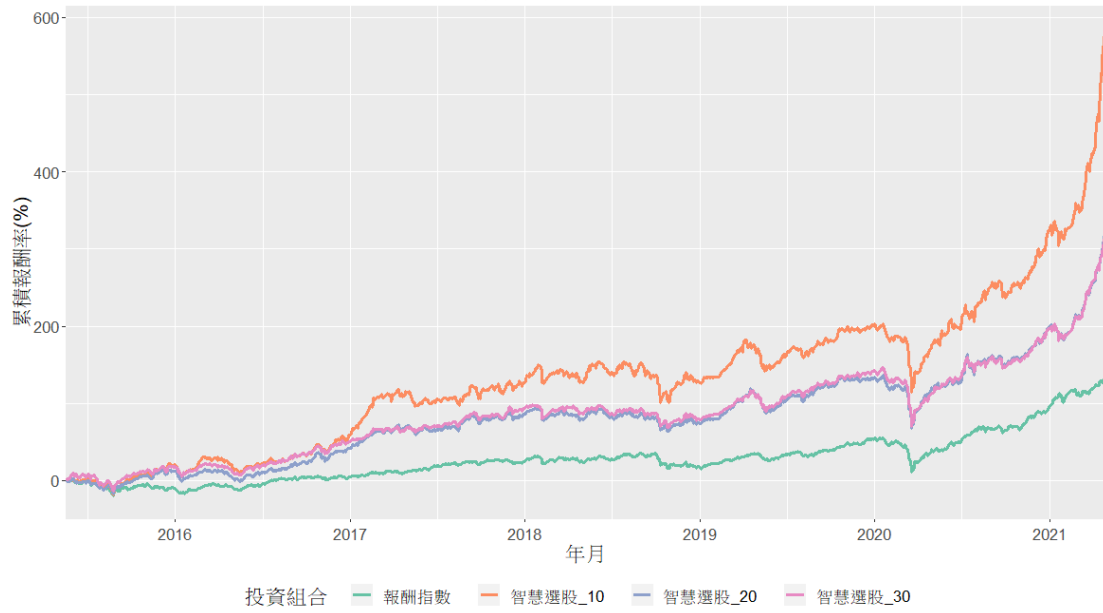


圖 17、流動性篩選最小風險投資組合績效-2

表 10、流動性篩選投資組合績效-2

投組名稱	累積報酬率	年化報酬率	年化標準差	夏普比率	最大回撤率
報酬指數	132.92%	15.75%	0.1508	1.0457	-28.55%
智慧選股 10_tangency	849.49%	47.63%	0.2732	1.5641	-39.76%
智慧選股 20_tangency	823.45%	46.92%	0.2400	1.7252	-32.13%
智慧選股 30_tangency	670.74%	42.40%	0.2335	1.6323	-31.94%
智慧選股 10_GMR	570.53%	39.01%	0.20756	1.69209	-29.21%
智慧選股 20_GMR	331.75%	28.81%	0.18006	1.49723	-29.34%
智慧選股 30_GMR	323.49%	28.38%	0.17838	1.49099	-30.92%

上圖與上表呈現的為使用日交易量前 50%的績效，與前一種流動性篩選之績效相比，此篩選方式不論在最適投資組合或是在最小風險投資組合中，投資 10 檔股票的投資組合之績效都有顯著的提升，年化報酬率與夏普比率皆是如此。可以看出若排除一些流動性稍差的股票，模型的預測效果更好。另外，在此方法的最小風險投資組合中年化標準差也較第一種流動性篩選方式來的低。這也可以滿足風險屬性較低的民眾的需求。綜上所述，若要以本研究方法建置一個理財機器人，使用第二種流動性篩選除了在實務上做再平衡更加容易外，還能夠滿足更多風險屬性的民眾。

## 第二節 市值篩選選股

本節篩選股票方式為使用公司市值前 100 名的公司股票進行選股。此篩選方式主要是希望能夠以追蹤大盤的方式建立投資組合，其概念就如同是在追蹤台灣 50 一樣。另外，因市值前 100 名公司之特性可能與整個台灣股市有所不同，因此本研究使用同樣的研究方法針對市值前 100 名的公司進行建模與建立投資組合，圖 18 為本節建模與資產配置流程：

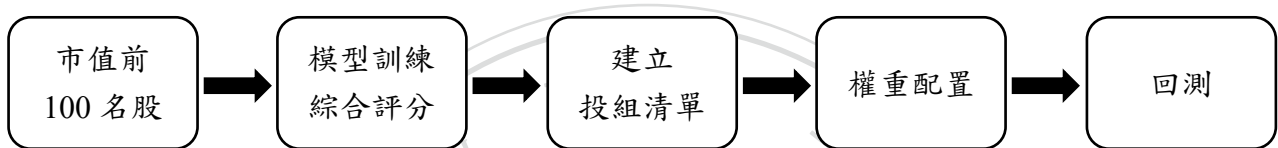


圖 18、市值前 100 之建模與資產配置流程

以上述方式進行建模與資產配置後，圖 19 為市值前 100 名最適投資組合績效，圖 20 為市值前 100 名最小風險投資組合績效，表 11 為兩種投資組合之績效，分別有投資 5 檔、10 檔、20 檔之績效。

若將上述建模流程中三個模型分開來看，就如同在前一章的機器學習模型中檢測模型預測力一樣。在表 12 中可以看到 XGBoost 模型與 MLP 模型與 SVR 模型在最小風險投資組合選取 5 檔、10 檔、20 檔之績效。可以發現在 SVR 模型的表現較差。

若將 SVR 去除，只是用 XGBoost 模型與 MLP 模型進行建模與資產配置後，圖 21 為 XGB 與 MLP 模型最適投資組合績效，圖 22 為 XGB 與 MLP 模型最小風險投資組合績效，表 13 為兩種投資組合之績效。可以看到相較於用三種模型進行建模，兩種模型的績效是比較好的。因此可以推論在本研究中使用集成學習概念下建模，若模型之預測力不佳時將會導致模型整體預測力下降。所以若使用本研究集成學習概念進行綜合評分時，須先確保各種模型有一定的預測力才能夠使模型預測力提升。



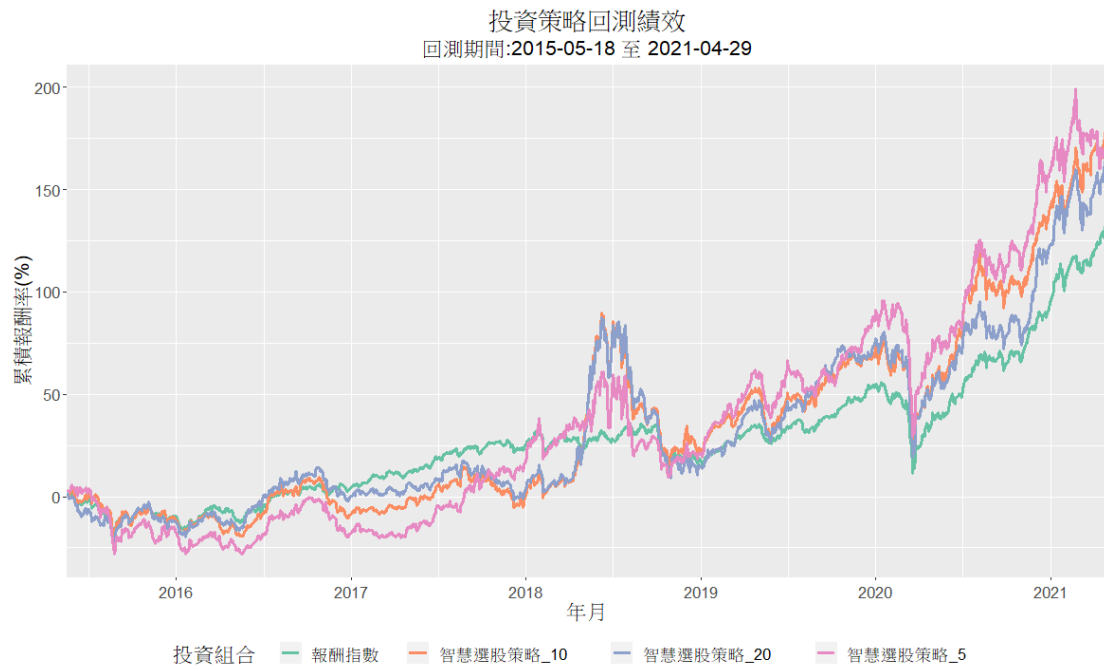


圖 19、市值前 100 最適投資組合績效

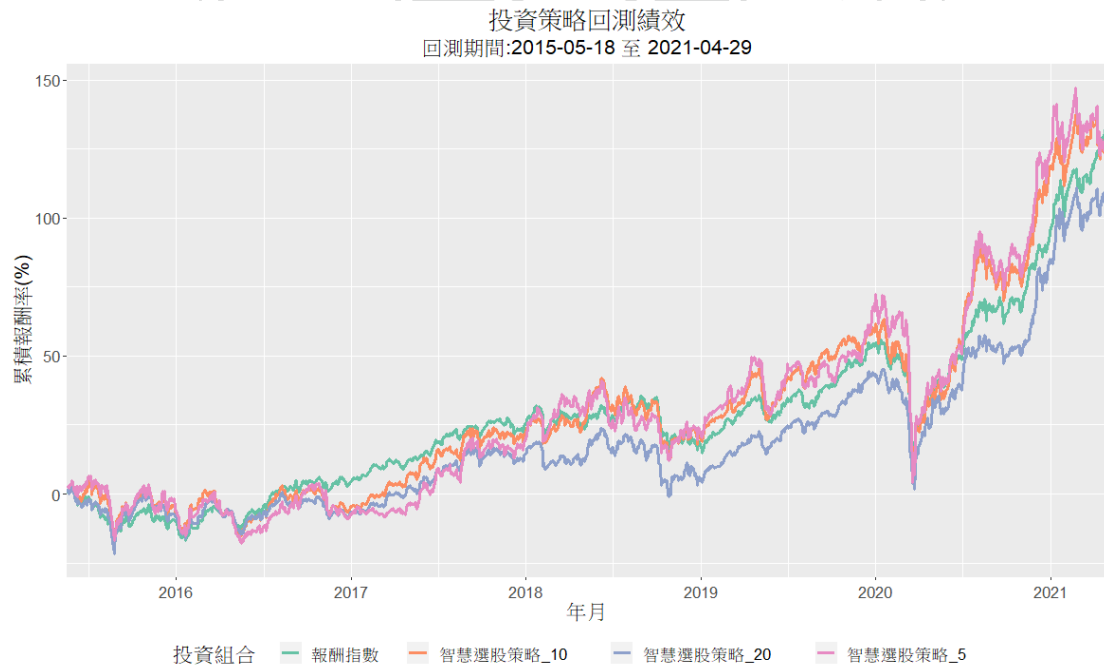


圖 20、市值前 100 最小風險投資組合績效



表 11、市值前 100 投資組合績效

投組名稱	累積報酬率	年化報酬率	年化標準差	夏普比率	最大回撤率
報酬指數	132.92%	15.75%	0.1508	1.04572	-28.55%
智慧選股 5_tangency	173.39%	19.01%	0.26723	0.78547	-33.95%
智慧選股 10_tangency	186.31%	19.97%	0.25951	0.83186	-39.70%
智慧選股 20_tangency	170.63%	18.81%	0.25111	0.81212	-41.97%
智慧選股 5_GMR	128.49%	15.37%	0.2282	0.7413	-39.67%
智慧選股 10_GMR	129.56%	15.47%	0.1979	0.8264	-33.51%
智慧選股 20_GMR	110.52%	13.75%	0.1694	0.8456	-29.94%

表 12、市值前 100 各模型投資組合績效

投組名稱	累積報酬率	年化報酬率	年化標準差	夏普比率	最大回撤率
XGB_5	55.91%	7.99%	0.2065	0.4759	-36.03%
XGB_10	124.99%	15.07%	0.1784	0.8762	-31.16%
XGB_20	86.07%	11.35%	0.1666	0.7289	-31.98%
MLP_5	100.99%	12.84%	0.2076	0.6863	-33.84%
MLP_10	125.36%	15.10%	0.1851	0.8530	-31.35%
MLP_20	74.56%	10.12%	0.1549	0.7003	-29.59%
SVR_5	49.41%	7.20%	0.2233	0.4233	-43.43%
SVR_10	39.90%	5.98%	0.1815	0.4113	-34.42%
SVR_20	38.36%	5.78%	0.1398	0.4722	-26.35%

投資策略回測績效  
回測期間:2015-05-18 至 2021-04-29

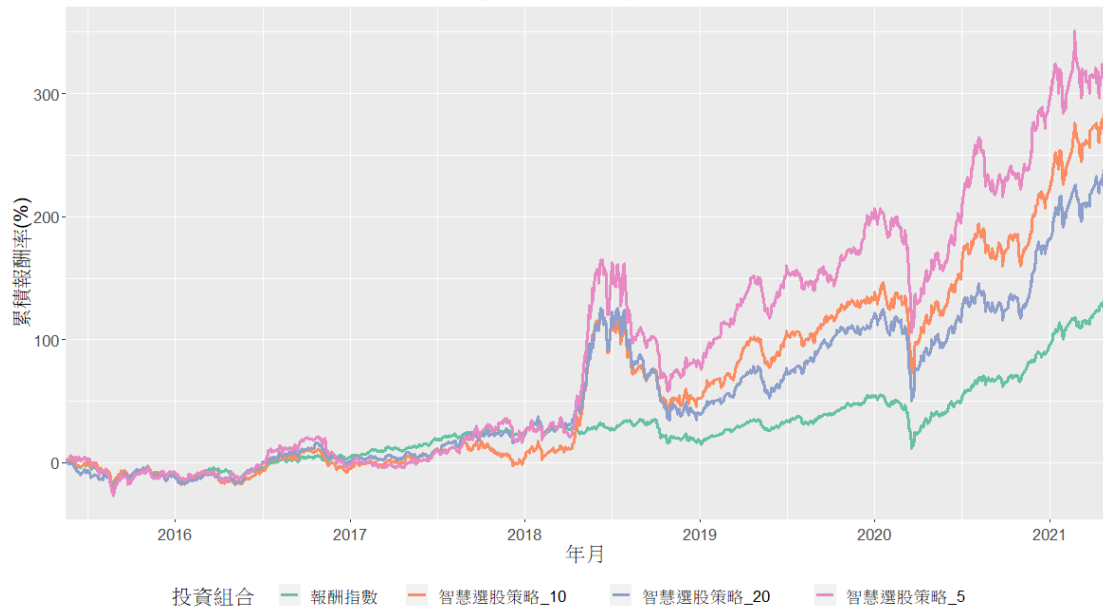


圖 21、XGB 與 MLP 綜合評分最適投資組合績效

投資策略回測績效  
回測期間:2015-05-18 至 2021-04-29



圖 22、XGB 與 MLP 綜合評分最小風險投資組合績效

表 13、市值前 100 名 XGB 與 MLP 綜合評分投資績效

投組名稱	累積報酬率	年化報酬率	年化標準差	夏普比率	最大回撤率
報酬指數	132.92%	15.75%	0.1508	1.0457	-28.55%
智慧選股 5_tangency	325.96%	28.51%	0.2874	1.0166	-40.42%
智慧選股 10_tangency	294.69%	26.82%	0.2605	1.043	-36.01%
智慧選股 20_tangency	247.13%	24.04%	0.2486	0.9912	-40.51%
智慧選股 5_GMR	182.12%	19.66%	0.2271	0.9046	-33.92%
智慧選股 10_GMR	129.03%	15.42%	0.1955	0.8320	-33.05%
智慧選股 20_GMR	135.73%	16.00%	0.1735	0.9428	-30.93%

## 第五章 結論與未來方向建議

本研究使用台灣上市公司的財報以集成學習概念對股票進行評分，使用 XGBoost 模型與 MLP 模型以及 SVR 模型對股票進行綜合評分。XGBoost 模型與 MLP 模型進行股票漲跌預測，SVR 模型則進行股票報酬率預測。相較於傳統使用單一機器學習對股票預測漲跌或是報酬率，本研究使用多種模型同時預測股票漲跌以及股票報酬率。除了減少單一模型訓練上常見的問題，如過度擬合與梯度下降法掉入局部極小值等問題，還考慮了預測股票漲跌與報酬率時可能重視不同特徵導致模型預測失準的問題。實證結果也證實使用多種模型預測股票漲跌與報酬率能夠有效的提升投資組合的報酬率。

在回測績效方面使用 Markowitz 提出的效率前緣概念進行資產權重配置，可以發現最適投資組合之績效比最小風險投資組合來得好，但最小風險投資組合的年化標準差也較小。

最後在市值前 100 名公司股票的回測中發現使用 XGBoost 模型與 MLP 模型綜合評分之績效比使用三種模型綜合評分績效更好，其原因為 SVR 模型在市值前 100 名公司股票的預測力較差。相較於 XGBoost 模型與 MLP 模型在市值前 100 名的回測績效，可以發現 SVR 模型的績效明顯較前兩者差。因此若要使用本研究方法對股票進行綜合評分時，應考慮各模型是否有一定的預測能力。若某個模型回測績效明顯低於其他模型回測績效，則可能代表其預測能力較差。在進行綜合評分時可能會使整體績效下降。

本研究以財報參數作為特徵並使用多個機器學型模型預測股票報酬率與漲跌，並未使用到股票的籌碼面與技術面的指標，在未來可以嘗試以不同指標並使用本研究集成學習概念進行股票預測。使投資標的轉換期間不再受限於財報發布時間，建立更彈性的再平衡頻率。而在模型訓練上除了本研究所使用的三個模型外也可以增加更多不同的模型來增加預測的準確度。

## 參考文獻

1. Alberg, J., & Lipton, Z. C. (2017). Improving factor-based quantitative investing by forecasting company fundamentals. *arXiv preprint arXiv:1711.04837*.
2. Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47, 552-567.
3. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
4. Clarke, R. G., De Silva, H., & Thorley, S. (2006). Minimum-variance portfolios in the US equity market. *The journal of portfolio management*, 33(1), 10-24.
5. Dietterich, T. G. (2002). Ensemble learning. *The handbook of brain theory and neural networks*, 2(1), 110-125.
6. Emerson, S., Kennedy, R., O'Shea, L., & O'Brien, J. (2019, May). Trends and applications of machine learning in quantitative finance. In *8th international conference on economics and finance research (ICEFR 2019)*.
7. Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14-15), 2627-2636.
8. Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003*.
9. Markowitz, H. (1952). The utility of wealth. *Journal of political Economy*, 60(2), 151-158.
10. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3), 199-222.

11. Tokat, Y., & Wicas, N. W. (2007). Portfolio rebalancing in theory and practice. *The Journal of Investing*, 16(2), 52-59.

