

# 使用參數化邊界支持向量分類機建立股票預測模型

郝沛毅 江旻緯 簡紫婷 曾麒璋 曾國軒 劉哲宇

國立高雄應用科技大學資訊管理學系

{ haupy、1098137220、1101137211、1101137240、1101137111、1101137122 }@cc.kuas.edu.tw

**Abstract**—股票在投資的市場中，有著迷人的高報酬率但相對的也伴隨著高度的風險，如何準確地預測未來股價的變動一直是投資者所關注的目標。本研究使用改良式參數化邊界支持向量分類機(par-v-SVC)對技術面分析中的乖離率指標(BIAS)、相對強弱指標(RSI)、威廉指標(WMS%R)…等，進行預測模型的建立，並將實驗分為兩部分：(1)將技術指標反應時間分為一天、三天、五天，並將股票漲跌幅度設定為 1%與 3%(股票股價漲跌幅度超過 1%或 3%才會判定為漲或跌，其餘皆歸類為持平)進行分析與預測。(2)與其他不同分類演算法(類神經網路、簡單貝氏分類、決策樹)進行正確率之比較。經實驗結果發現，改良式 SVM(par-v-SVC)對於股票股價漲跌預測之正確率，相較於其他分類演算法來得較佳。並從實驗中觀察發現，漲跌幅度設定為 3%之資料集，整體預測正確率比 1%之資料集高，而反應天數在漲跌幅度為 1%的資料集中三天與五天為佳；相反的，在漲跌幅度為 3%的資料集中，反應時間為一天之正確率為最高。

**Keywords:** 股價預測、支持向量機、資料探勘

## I. 前言

股票為一種有價證券，是公司為募集資金，而發給投資者作為公司資金部分所有權的憑證。近年來股票成為投資者投資與投機的重要理財工具，股票有著高額的報酬率但相對伴隨著高度的風險，對多數投資人來說無不希望從股票市場獲得超額的利潤，投資人為獲得超額利潤必須從多個角度考慮與預測股票的漲跌，如資金面、政治面、技術面、心理面等種種因素等，多數投資人在選擇股票投資時主要是採用基本面分析和技術面分析兩種，基本面分析是分析企業的本質，屬於中長期的判斷，然而基本面分析無法提供明確的買賣點。而技術面分析可清楚地指出買賣的時間點，不同的技術指標甚至可以進行不同期間的預測。在技術分析方面實務上採用的指標相當多，然而在選取時必須要考慮個股特性，以選擇合適的指標，才能獲得較高的報酬。

拜科技進步所賜，使得人們解決問題不再只是單憑過去的經驗，因為股票市場的特性，股票預測學派提出股價是有跡可循的，歷史會一再的重演並遵循著「漲高回檔，跌深反彈」的特性，過去與未來之間或許存在著潛在的關聯，單憑著過去的數據與投資人的經驗，不但不容易掌握其中的關聯性，更容易添加自我主觀的意識而影響投資人的評斷，所以多數的企業與專家都會藉由

電腦做為輔助投資決策的工具，因此各種不同的資料探勘(Data Mining)工具，如：時間序列分析、ARIMA、類神經網路、模糊理論…等，被用在股票預測的決策工具之中。這些系統可以幫助投資者做出更客觀準確的判斷，透過電腦也可以更方便的蒐集資料，找出過去與未來軌跡，減少人為的錯誤與疏失。

在眾多資料探勘工具中，支持向量機(SVM)被大量使用在解決分類與迴歸問題，支持向量機是一種監督式學習的方法，主要用在分類(Classification)和迴歸(Regression)上。SVM有兩個特性：(一)能有效的解決線性與非線性問題。(二)SVM學習可以利用已知的有效演算法發現目標函數的最小值。相較於其他演算法(如：類神經網路)不會出現極值的問題。本研究使用改良式支持向量機(par-v-SVC)進行股價漲跌預測，par-v-SVC不同於傳統SVM，使用管狀式的邊界分割類別，而par-v-SVC使用參數化任意形狀邊界分割類別，par-v-SVC於錯誤是異質的情況下特別有用，換句話說，就是錯誤發生的變異質隨著資料的位置不同而改變有著良好的幫助。

## II. 文獻探討

### A. 技術指標

證券市場的股票價格因為要考慮太多因素，所以股價是複雜多變的，投資者在投資時需要有一套方法來制定投資策略，技術分析是投資市場中非常普及應用的分析方法。所謂的技术分析是相對於基本分析而言的，基本分析著重於一般經濟情況以及各個公司的經營狀況，行業動態等因素進行分析，而技術指標是透過圖表或技術指標的紀錄，研究市場過去與現在的行為反應，以推測未來股價的趨勢與走向，其依據的內容主要是由股價、成交量、或漲跌指數計算而來，技術分析只會關心證券市場本身的變化，而不考慮經濟與政治等外部因素的影響。基本分析與技術分析都認為股價是由供需關係所決定。基本分析是根據影響供需關係的因素分析進行預測，而技術分析是根據股價本身的變化來預測走勢。技術分析的基本觀點是：所有股價的實際供給需求及與其引導的因素如：股票市場上投資者對未來的希望、擔心、恐懼等等都反映在價格和交易量上。

著名的經濟學家凱恩斯於 1936 年提出空中樓閣理論，是技術理論的基礎，該理論完全拋開股價的內在價值，強調心理建構的出來的空中樓閣。投資者會以一定

的價格購買股票，是因為他們相信有其他人會以更高的價格來購買他所持有的股票，至於價格的高低並不重要，重要的是存在著更大的投資者以更高的價格向你購買股票，所以一般投資者必須搶在最大的投資者購買之前買進股票，並在股票的最高點賣給願意出最高價的投資者。在股票市場中不同的學者對於股票有不同的觀點，Cootner 於 1964 在效率市場中(Efficient Market) 股價在任何時間點均已反應所有訊息[1]，所以股價是不可預測的，Brealey & Mayer 學者認為最佳的獲利策略是長期買進並持有股票[2]，以股票股利賺取超額利潤，亦有學者 Edwards & Magee 認為股票股價是有軌跡可循[3]，股票歷史會一再重演，並試圖藉由大量的統計資料來預測行情走勢，此為技術分析學派學者的觀點，此觀點為本研究動機之一。

本論文使用下列常用以指標判斷市場強弱、買賣訊號…等以協助投資者買賣決策。本研究根據過去文獻[4-7]整理出常用以判斷股票股價漲跌技術指標如下：

### A.1 乖離率 BIAS

乖離率(BIAS)是由 Granville 所提出的葛蘭碧八大法則(J. Granville Rule)[8]中的兩大法則推演而來，乖離率的基本理論為當股價或指標偏離平均線太遠時，便會向平均線移動，為了測量此偏離的程度，於是發展出乖離率指標。

股價收盤價離開移動平均線之值稱為乖離，而乖離再除以移動平均值稱為乖離率。移動平均值為一段時間收盤價的平均值，可代表投資者在這段期間內以收盤價購買股票的平均成本。而乖離率可代表投資者的平均報酬，股價在移動平均線之上代表乖離率為正，表示目前的股價高於購買股票時的價格。而正值越大投資者的獲利也越高，但因為投資者想要了結獲利的心態越重，投資者賣出壓力也逐漸提高，股票回檔的機率也相對提高；相反的股票在移動平均線之下代表乖離率為負，表示目前的股票低於購買股票時的價格。而負值越大表示虧損嚴重，投資者想要出脫的心態也逐漸降低，賣壓減輕，買方則認為股價跌幅大，買入成本極低，買入意願加強，股票行情回升的可能性越高，股乖離率負值越高股價反彈的機率提高。乖離率計算公式如下

$$N\text{日移動平均線指數} = \left( \frac{N\text{日收盤價加總}}{N} \right)$$

$$N\text{日乖離率} = \left( \frac{\text{當日收盤價} - N\text{日移動平均線指數}}{N\text{日移動平均線指數}} \right) \times 100\%$$

本論文則採用 6 日與 12 日作為乖離率之參數，並以 6 日與 12 日之乖離率做為輸入特徵[10]。

### A.2 相對強弱指標(RSI)

相對強弱指標(RSI)公式是由 J. Welles Wilder, Jr.(1978)所提出，但普遍投資者採用的是 Andrew Cardwell 的價格預測方法 [8]。RSI 指標是利用股票上整或下跌趨勢來判斷超買或超賣，用以判斷股票是漲或是跌。RSI 計算公式如下

$$N\text{日RSI} = \frac{N\text{日收盤價上漲平均數}}{|N\text{日收盤價變動的平均數}|} \times 100$$

若 RSI 很小(例如 RSI<20)：一般認為股價下跌已深，市場超賣建議投資者買進。若 RSI 很大(例如 RSI>80)：一般認為股價上漲幅度高，市場超買建議投資者賣出。RSI 值投資者可自行設定買點與賣點，如 RSI<10 買進，RSI>90 賣出。本研究使用 6 日之 RSI 做為輸入特徵 [8]。

### A.3 威廉指標(Williams' Oscillator)

威廉指標是由美國投資專家 Larry Williams 提出 [9]，威廉指標利用擺動原理來判斷超買或是超賣現象，做為決策參考，也可用來預測股市循環週期內的高點與低點，其值介於 0 至 100 之間。威廉指標通常為股市循環週期的一半，以台灣股市為例，通常採用 12 日威廉指標(一個月為 4 周共有 24 個營業日，取一半為 12 日)。計算公式如下：

$$12\text{日威廉指標} = \frac{12\text{日內最高價} - 12\text{日收盤價}}{12\text{日內最高價} - 12\text{日內最低價}} \times 100\%$$

威廉指標異常值：當 12 日%R 值大於 80 時，代表股市呈現超賣現象，為買進時機。反之當 12 日%R 值小於 20 時，代表股市出現超買現象。為賣出時機。當 12 日威廉指標從超賣區向上升高，代表行情將止跌回穩。如突破 50(中央線)代表股市轉強可追加買進。反之若 12 日威廉指標從超買區向下滑落，且低於中央線代表股市跌勢轉強，可追跌賣出。本研究使用 6 日與 12 日之威廉指標做為輸入特徵[8]。

### A.4 KD 隨機指標

KD 隨機指標(Stochastic Oscillator)原名 %K&%D，是由 George C. Lane 於 1950 年推廣使用[8]，是技術分析中一種動量分析方法，隨機一詞是指標在一段時間內相對於其波動範圍的位置。公式計算方法是先計算出未成熟隨機指標(Raw Stochastic Value, RSV)，在求出當日的 K 值與 D 值，K 值為快速平均值，反應較為靈敏；D 值為慢速平均值，反應較為不靈敏。若 K 值大於 D 值表示股票處於漲勢；反之則為跌勢。計算公式如下

$$RSV = \frac{\text{第}N\text{日收盤價} - \text{最近}N\text{日內最低價}}{\text{最近}N\text{日內最高價} - \text{最近}N\text{日內最低價}} \times 100$$

$$K值(\%K) = \frac{2}{3} \times \text{前一日之}K值 + \frac{1}{3} \times RSV$$

$$D值(\%D) = \frac{2}{3} \times \text{前一日之}D值 + \frac{1}{3} \times \text{當日}K值$$

若無前一日之 K 或 D 值分別為 50 代入。當 K 值大於 D 值為漲勢，因此當 K 線(短期)由下向上突破 D 值(長期)時為黃金交叉點，股市看漲；相反的當 D 值大於 K 值為跌勢，因此當 K 線由上向下跌破 D 線時為死亡交叉，股市看跌。若  $KDV_n$  為正  $KDV_{n+1}$  為負代表 K 線由上方向下跌落 D 線，為死亡交叉，KD 值標記為 -1，若  $KDV_n$  為負  $KDV_{n+1}$  為正，代表 K 線由下方超越 D 線，此時為黃金交叉，KD 值標記為 1，若  $KDV_n$  與  $KDV_{n+1}$  同時為正或負，則代表沒死亡交叉或黃金交叉發生，KD 值標記為 0。本研究使用 9 日之 KD 值做為輸入特徵[8]。

### A. 5 心理線(PSY)

心理線(Psychology line)是一種建立在研究投資者心理趨向基礎上，將某段時間內投資者傾向買方還是賣方的心理與事實轉化為數值，形成人氣指標，作為買賣股票的參數。計算公式如下

$$N日PSY = \frac{N日內上漲的天數}{N} \times 100\%$$

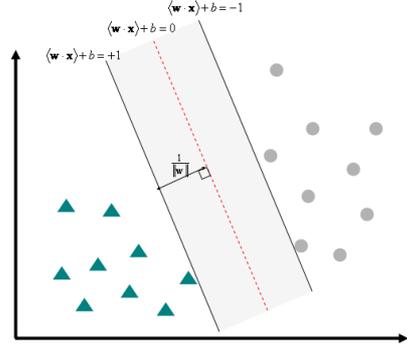
心理線一般常態分佈在 25%-75%之間，但是心理線的超買或超賣應看多頭市場或是空頭市場。當心理線低於 10%或高於 90%時，非常可能是超買或超賣現象，行情反轉的可能性高，故此時買賣最適宜。本研究使用 6 日之心理線做為輸入特徵。[8]

### B. 支持向量機

支持向量機(Support Vector Machines, SVM)是一種分類(Classification)演算法，可處理分類兩位的問題，是由Vapnik等在1995年根據統計學習理論中結構化風險最小誤差法(Structural Risk Minimization, SRM)所提出的一種新的機器學習方法。[15]

我們想要將圓與三角形做分類，作法是在兩群中間找出一個最佳的超平面(Hyper Plane)來分割這兩類的資料，圖一顯示最佳的分割超平面，即是邊界(margin)最大的超平面，最佳超平面可透過下述最佳化問題求解出：

$$\begin{aligned} & \underset{w, b, \xi_i}{\text{minimize}} && \frac{\|w\|^2}{2} + C \sum_{i=1}^N \xi_i \\ & \text{subject to} && y_i (\langle w \cdot \Phi(x_i) \rangle + b) \geq 1 - \xi_i \\ & && \xi_i \geq 0 \quad \forall i \end{aligned} \quad (1)$$



圖一、SVM中的最佳分割超平面

使用Lagrange理論，我們可以得到下列對偶問題。

$$\begin{aligned} & \underset{\lambda_i}{\text{maximize}} && -\frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j \langle \Phi(x_i) \cdot \Phi(x_j) \rangle + \sum_{i=1}^N \alpha_i \\ & \text{subject to} && \sum_{i=1}^N y_i \alpha_i = 0 \\ & && \alpha_i \in [0, C] \quad \forall i \end{aligned} \quad (2)$$

$\alpha_i, i=1, 2, \dots, m$  為Lagrange multipliers，如果有一個  $x_i$  的  $\alpha_i > 0$ ，則該  $x_i$  即可稱為支援向量。我們找出支援向量後，就可以找到最大邊界的分割超平面。最後我們可以得到一個分類處理的函數：

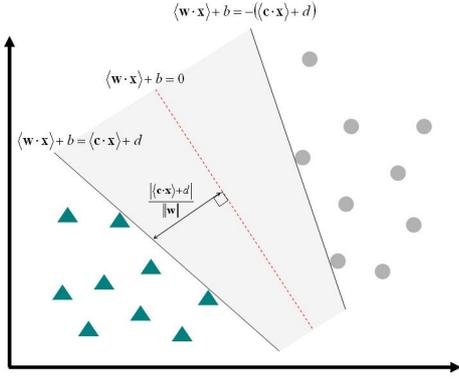
$$f^*(x) = \text{sgn} \left( \sum_i \alpha_i y_i k(x_i, x) + b \right)$$

其中  $k(x, y) \equiv \langle \Phi(x) \cdot \Phi(y) \rangle$  為核心函數。

### III. 研究方法

本研究使用一個新的SVM演算法(par-v-SVC)來建立股價預測模型，此演算法與其他SVM演算法的差異在於，par-v-SVC擁有可調適錯誤的功能，換句話說，使用者可自行定義一個  $v$  參數來控制錯誤向量的上界以及支援向量數目的下界，此外par-v-SVC演算法可使用一個參數化的任意形狀邊界來取代傳統SVM的管狀邊界，用以分割兩個類別，這對於錯誤是異質的情況下特別有用。Par-v-SVC的最佳超平面可透過下述最佳化問題求解出：

$$\begin{aligned} & \underset{w, c, b, d, \xi_i, \xi_i^*}{\text{minimize}} && \frac{1}{2} \|w\|^2 + C \left( -v \cdot \left( \frac{1}{2} \|c\|^2 + d \right) + \frac{1}{N} \sum_{i=1}^N \xi_i \right) \\ & \text{subject to} && y_i (\langle w \cdot \Phi(x_i) \rangle + b) \geq (\langle c \cdot \Phi(x_i) \rangle + d) - \xi_i \\ & && d \geq 0, \xi_i, \xi_i^* \geq 0 \quad \text{for } i=1, \dots, N. \end{aligned} \quad (3)$$



圖二、使用參數化邊界來分割兩個類別。

使用Lagrange理論，我們可以得到下列對偶問題。

$$\text{maximize}_{\alpha_i} \begin{cases} -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \\ + \frac{1}{2Cv} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \end{cases} \quad (4)$$

subject to

$$\sum_{i=1}^N y_i \alpha_i = 0, \quad \sum_{i=1}^N \alpha_i \geq C \cdot v, \quad \alpha_i \in \left[0, \frac{C}{N}\right].$$

$\alpha_i, i=1, 2, \dots, m$  為Lagrange multipliers，如果有一個  $x_i$  的  $\alpha_i > 0$ ，則該  $x_i$  即可稱為支援向量。求解出  $\alpha_i$  後，我們可以得到權重向量為：

$$\mathbf{w} = \sum_{i=1}^N y_i \alpha_i \Phi(\mathbf{x}_i) \quad \text{與} \quad \mathbf{c} = \frac{1}{C \cdot v} \sum_{i=1}^N \alpha_i \Phi(\mathbf{x}_i).$$

得到權重向量  $\mathbf{w}$  與  $\mathbf{c}$  後，我們可以使用下述KKT(Karush-Kuhn-Tucker)最佳化條件求出偏移量  $b$  與  $d$

$$\alpha_i [y_i (\langle \mathbf{w} \cdot \Phi(\mathbf{x}_i) \rangle + b) - (\langle \mathbf{c} \cdot \Phi(\mathbf{x}_i) \rangle + d) + \xi_i] = 0 \quad (5)$$

$$\left(\frac{C}{N} - \alpha_i\right) \xi_i = 0 \quad (6)$$

對於某些滿足  $\alpha_i \in \left(0, \frac{C}{N}\right), y_i = +1$  與  $\alpha_j \in \left(0, \frac{C}{N}\right), y_j = -1$ ，我們有  $\xi_i = \xi_j = 0$  而且公式(5)中第 2 項值為 0，所以偏移量  $b$  與  $d$  可以使用下式計算出來。

$$b = \frac{-1}{2} \left[ \langle \mathbf{w} \cdot \Phi(\mathbf{x}_i) \rangle + \langle \mathbf{w} \cdot \Phi(\mathbf{x}_j) \rangle \right], \quad (7)$$

$$d = \frac{1}{2} \left[ \langle \mathbf{w} \cdot \Phi(\mathbf{x}_i) \rangle - \langle \mathbf{w} \cdot \Phi(\mathbf{x}_j) \rangle \right], \quad (8)$$

對於某些  $i, j$  使得  $\alpha_i \in \left(0, \frac{C}{N}\right), y_i = +1$  與  $\alpha_j \in \left(0, \frac{C}{N}\right), y_j = -1$ ，

因此，最終的決策函數  $f^*$  與參數化邊界函數可以使用下列公式得到：

$$f(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i k(\mathbf{x}, \mathbf{x}_i) + b, \quad (9)$$

$$f^*(\mathbf{x}) = \text{sgn}(f(\mathbf{x})), \quad (10)$$

$$\text{與} \quad g(\mathbf{x}) = \left( \frac{1}{C \cdot v} \sum_{i=1}^N \alpha_i k(\mathbf{x}_i, \mathbf{x}) + d \right). \quad (11)$$

我們找出支援向量後，就可以找到最大邊界的分割超平面。最後我們可以得到一個分類處理的函數：

$$f^*(\mathbf{x}) = \text{sgn} \left( \sum_i \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \right)$$

其中  $k(\mathbf{x}, \mathbf{y}) \equiv \langle \Phi(\mathbf{x}) \cdot \Phi(\mathbf{y}) \rangle$  為核心函數。當  $f(\mathbf{x}) > 0$  時，表示該筆資料與標註為“+1”的資料同屬一類；反之則是屬於另外一類。

#### IV. 實驗結果

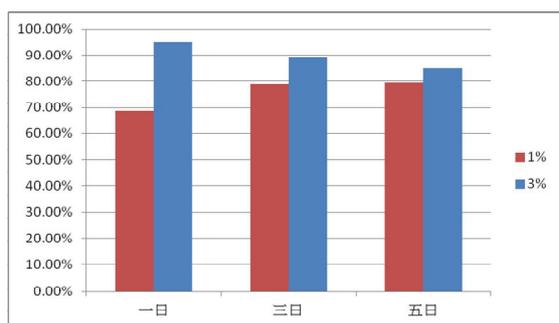
本研究資料透過台灣經濟新報資料庫(TEJ+)：「TEJ Equity」的「調整股價(日)-除權息調整」資料庫。取得公司過去每日的歷史股價資料(開盤價、當日最高價、當日最低價、收盤價)，並且透過 NetBean 7.4 版使用 Java 語言對過去歷史股價資料進行進一步的資料整理與技術指標的處理。缺漏值的部分，若該個股當天沒有開盤沒有取得開盤價、當日最高最低價與收盤價時，TEJ+會自動將缺漏之部分補齊為前一天之資料。本資料期間為台股數間公司十年間股價資料(2003年1月1日至2012年12月31日)。首先對蒐集資料進行資料前處理，以個股公司股價資料進行技術指標運算，整理研究所需之交易資料，接著透過股價漲跌範圍(單純上漲下跌、漲跌幅度 1%、漲跌幅度 3%)標記資料類別(漲、跌、持平)，並轉換為支援向量機演算法可以分析之格式。再以相同資料利用 Weka 程式中的其他分類器：類神經網路(Neural Network)、貝氏分類器(Bayesian Classifier)、決策樹(Decision Tree)進行分類正確率比較。本研究使用技術指標(表一，T1-T7)做為判斷一、三、五日內「上漲」、「持平」與「下跌」三個類別(表 1，O1-O3)的輸入特徵，本研究所使用的特徵變數表如下表一所示。

表一、輸入與欄位資料

變數(T)	代表變量
T1	六日乖離率指標值
T2	十二日乖離率指標值
T3	十二日威廉指標值
T4	六日相對強弱指標值
T5	十二日相對強弱指標值
T6	心理線值
T7	KD 線值
O1	隔天股價漲跌
O2	第三天股價漲跌
O3	第五天股價漲跌

表一中，欄位 O 為股價漲跌標記，分別為隔日、三日與五日，以股票漲跌幅不同判定股價漲跌或持平，漲為 1、持平為 0、跌為-1，若漲跌幅度大於預設漲幅(1%、3%)則標記為漲；若小於預設漲跌幅度(-1%、-3%)則標記為跌。

實驗對象 A：台積電，資本額：2,591.5 億元，樣本數為：987 筆



圖三、漲跌幅度為 1%與 3%之正確率比較(台積電)

從圖三能看出，實驗漲跌幅度設定為 3%之平均正確率較佳，雖然漲跌幅度為 3%反應時間為一日之正確率為最佳，但這是因為「持平」類別過多的影響，然而在漲跌幅度設定為 1%，反應時間相同為一日的實驗，正確率卻是最低。觀察後發現，漲跌幅度設定為 1%與 3%反應時間為一日之實驗，實驗結果相同皆因「持平」類別資料集過多，導致上漲與下跌類別資料集過少，但是反應時間為一日漲跌幅度為 1%之資料集是因為，在原始資料中隔日的漲跌次數較少，導致「上漲」與「下跌」類別的筆數過少，但是漲跌幅度為 3%之實驗，因為將漲跌幅度設定為 3%因此「上漲」類別需要漲到一定程度才會被判定為漲，所以雖然訓練不足，但是特徵較為明顯，相較於漲跌幅度設定為 1%之實驗結果效果較佳。

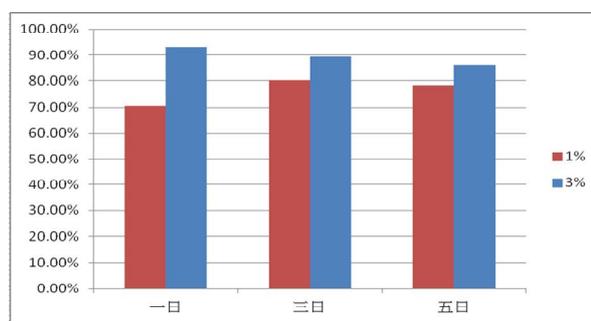
接著，利用相同的資料集與相同的設定(漲跌幅度為 1%、3%反應時間為一、三、五天)，與其他常見之分類器(類神經網路、簡單貝氏、決策樹)進行比較，由表

二中可以發現，本研究改良式 SVM 整體正確率皆優於其他三種分類器，從表二中也可看出，除了漲跌幅度為 1%反應時間為三日之正確率，略為低於類神經網路外，其餘皆優於其他分類演算法。

表二、不同分類演算法之比較(台積電)

分類器	漲跌幅度為 1%			漲跌幅度為 3%		
	一日	三日	五日	一日	三日	五日
par-v-SVC	68.59%	78.91%	79.43%	94.83%	89.16%	85.11%
NN	65.65%	79.84%	78.82%	94.83%	85.82%	83.28%
DT	37.99%	62.72%	67.48%	94.22%	82.78%	57.35%
bayesian	47.32%	65.75%	75.89%	82.07%	69.40%	74.06%

實驗對象 B：正新，資本額：83 億，樣本數為：987 筆



圖四、漲跌幅度為 1%與 3%之正確率比較(正新)

從圖四中可以明顯看出，漲跌幅度設定為 3%之正確率明顯優於 1%之正確率，雖然實驗結果反應時間為一日漲跌幅度為 1%與 3%之實驗結果剛好相反，一個為最差，一個為最佳，但結果與實驗對象 A 相似，因為「持平」類別資料集過多，導致其餘兩個類別訓練不足，但因為漲跌幅度設定為 3%之實驗，雖然因為訓練不足，但是因為漲跌判定較為明顯，以至於正確率與 1%相比較佳。從三中可以看出，實驗對象 B 與實驗對象 A 一樣，改良式 SVM 預測股票漲跌之正確率皆優於其餘分類器。而表中除了漲跌幅度為 1%反應時間為五日之模型正確率不高，低於類神經網路外，其餘模型皆高於其他分類演算法。

表三、不同分類演算法之比較(正新)

分類器	漲跌幅度為 1%			漲跌幅度為 3%		
	一日	三日	五日	一日	三日	五日
par-v-SVC	70.62%	80.14%	78.32%	93.01%	89.36%	86.12%
NN	65.35%	77.10%	80.24%	91.89%	86.52%	84.80%
DT	50.46%	64.64%	70.82%	88.86%	75.08%	57.55%
bayesian	67.58%	56.94%	74.57%	75.79%	74.97%	78.22%

## V. 結論與建議

高報酬的股票投資，一直都是投資人想要獲得高額利潤的投資工具，但也因為伴隨著高度的風險，使得在股票投資上不但沒有獲得高額利潤還慘賠的投資人也

不在少數，所以如何提高股票股價的預測能力與正確率，一直都是投資人迫切關注的一環，然而技術指標雖然種類眾多，但如何應用與建立一套強大的預測模型，是近年來研究的熱門議題，而人工智慧的興起，與分類器演算法工具的日益成熟，如何將股票漲跌有效的分類並提高分類的正確率，是資料探勘領域中一直在嘗試的方向。

前面提到分類器的興盛，對於許多以往難以獲得良好解答的複雜問題如：股票預測、人臉辨識等，都有了良好的解答方式，其中又以 SVM 對於分類與迴歸問題有著傑出的解決能力，本研究使用改良式的 SVM(par- -SVC)，與傳統管狀的分類邊界不同，par- -SVC 使用任意形狀之分類參數，對於具有異質性雜訊的問題有著傑出的改善。本研究針對數家上市櫃公司進行實驗，透過觀察技術指標的反應時間，與不同漲跌幅度的設定發現了，在實驗一中，針對不同的反應時間(一、三、五日)與漲跌幅度設定(1%與 3%)中，漲跌幅度 1%與 3%比較結果，漲跌幅度 3%的整體正確率，明顯優於漲跌幅度為 1%之資料集，原因在於將漲跌幅度設為 3%之資料，若無明顯漲跌(漲跌幅度低於 3%)，皆被判斷為持平類別，唯有明顯漲跌幅之資料才會被判定為上漲或下跌，雖然持平類別資料過多，可能會造成另外兩個類別訓練不足，但因為漲跌幅度設為 3%特徵較為明顯，所以相較於漲跌幅度為 1%之資料，正確率較佳。而反應時間在一、三、五日的設定下，因為隔日漲跌幅度相較於三日五日來說，變動相對沒有那麼大，所以導致上漲與下跌類別資料不足，以至於訓練不足正確率不佳，反應時間三、五日中，因為資料分配較為平均，模型訓練較足，所以正確率比反應時間為一日之資料來的優秀。實驗二中，本研究改良式 SVM 在帳跌幅度設定為 1%與 3%，反應時間為一三五日中，平均正確率皆高於其他分類演算法(類神經網路、簡單貝氏、決策樹)。其中又以與貝氏分類器和決策樹比較之正確率相對來的較佳，與類神經網路之正確率較為接近也相對較高。

## VI. 參考文獻

1. Coonter, Paul, The Random Character of Stock Marker Prices. London: Risk, 1964
2. Brealey, Richard A, and Stewart C. Mayers, Principles of Corporate Finance. NY: McGraw-Hill, 1991.
3. Edward, Robert D., John Magee, and W. H. C. Bassetti, Technical Analysis of Stock Trends. Boca Raton, NY: AMACOM, 1984.
4. 洪志豪，技術指標 KD、MACD、RSI 與 WMS%R 之操作績效實證，國立台灣大學企業學系碩士論文，1999。
5. 簡辰丞，結合 MACD 與類神經模糊技術之股票預測模型-以台灣金融股為例，靜宜大學企業管理學系碩士論文，2001。
6. 喻欣凱，運用支援向量機與文字探勘於股價漲跌趨勢之預測，輔仁大學資訊管理學系碩士論文，2008。
7. 陳鄴貞，以財務指標及技術指標建構股價預測模型-類神經網路模型之應用，國立台北大學國際財務金融學系碩士論文，2011。
8. Brown, Constance, Technical analysis for the trading professional, NY: McGraw-Hill, 1999.
9. Willians, Larry, How I made one million dollars in the commodity market last year, Carmel Valley, CA: Conceptual Management, 1973.
10. 林宗永，證券投資技術分析指標獲利性之研究，政治大學企業管理系碩士論文，1989。
11. 蔡宜龍，台灣股票市場技術分析指標有效之衡量，成功大學工業管理系碩士論文，1989。
12. 方國榮，證券投資最適指標之研究：技術面分析，台灣大學商學系碩士論文，1991。
13. 蘇子龍，證券市場技術分析指標有效性之探討：以乖離率、威廉指標為例，中山大學企業管理系碩士論文，1992。
14. 高梓森，台灣股市技術分析之實證研究，台灣大學財務金融學系碩士論文，1994。
15. Burges C., A tutorial on support vector machines for pattern gnition, Data mining and Knowledge discover, 1998.