

南華大學科技學院資訊管理學系

碩士論文

Department of Information Management

College of Science and Technology

Nanhua University

Master Thesis

應用模糊分類技術與技術指標於股價漲跌預測之研究

A Study on Application of Fuzzy Classification

Techniques and Technical Indicators to Stock Price

Prediction

郭俊億

Jyun-Yi Guo

指導教授：王佳文 博士

Advisor: Jia-Wen Wang, Ph.D.

中華民國 109 年 1 月

January 2020

南華大學
科技學院資訊管理學系
碩士學位論文

應用模糊分類技術與技術指標於股價漲跌預測之研究
A study on application of fuzzy classification techniques
and technical indicators to stock price prediction

研究生：鄧俊億

經考試合格特此證明

口試委員：陳智賢

洪銘建
孫文文

指導教授：孫文文

系主任(所長)：陳智賢

口試日期：中華民國 109 年 01 月 03 日

南華大學資訊管理學系碩士論文著作財產權同意書

立書人：_____郭俊億_____之碩士畢業論文

中文題目：

應用模糊分類技術與技術指標於股價漲跌預測之研究

英文題目：

A study on application of fuzzy classification techniques
and technical indicators to stock price prediction

指導教授： 王佳文 博士

學生與指導老師就本篇論文內容及資料其著作財產權歸屬如下：

- 共同享有著作權
- 共同享有著作權，學生願「拋棄」著作財產權
- 學生獨自享有著作財產權

學 生： 郭俊億 (請親自簽名)

指導老師： 王佳文 (請親自簽名)

中華民國 109年 1月 14月

南華大學碩士班研究生
論文指導教授推薦函

資訊管理系碩士班郭俊億君所提之論文
應用模糊分類技術與技術指標於股價漲跌預測之研究
A study on application of fuzzy classification techniques and
technical indicators to stock price prediction
係由本人指導撰述，同意提付審查。

指導教授 班級

109年1月14日

誌 謝

首先要感謝我的指導老師王佳文教授，在論文撰寫中，老師不辭辛勞的指導與討論，才能使得我完成本篇論文。在這兩年中發生了許多事情，讀書、休學、復學，也換了指導老師，並從 AI 機器人程式中轉到大數據研究，跟著王佳文老師一直學習著相關知識，從不太懂到現在能自己跑完分析寫完一篇論文，除了學習到相關知識，更能擁有獨立思考問題並解決問題的能力，而在平時相處討論中，也學習到待人處事的態度。

再來則要感謝陳萌智老師，雖然沒有跟您到最後，但也是因為您我才會讀研究所，而讓我學習到跟現在趨勢相關的 AI 機器人相關領域，讓我更有機會去接觸自己平時接觸不到的事物。

另外還要感謝資訊管理系的系主任、系助理以及各位小義工們，在最後一學期中，互相幫忙以及解決系上各種事物也擁有很寶貴的經驗，以及黃信凱同學、蕭憶臻在口試當天臨時的幫忙，也要謝謝陳又嘉同學，常常麻煩妳幫忙一些雜事。

最後要感謝同學張育嘉、廖偉哲、董紀延、王睿宇、魏宏恩以及溫志仁，在低潮時能互相鼓勵以及給予意見，並且隨時都能互相幫助，擁有一群好朋友的陪伴，在任何事情裡都不無趣了。

此外要感謝我的家人，擁有你們的支持、鼓勵與陪伴我才能安安心心地度過生命的每一個階段，也因為你們才能在我求學的階段沒有後顧之憂。最後，謹以本篇論文獻給所有曾經幫助過我的人，在此獻上最誠摯的謝意。

應用模糊分類技術與技術指標於股價漲跌預測之研究

學生：郭俊億

指導教授：王佳文

南 華 大 學 資 訊 管 理 學 系 碩 士 班

摘 要

技術分析學派投資人認為股價會一再複製過去的歷史，因此可以利用此特性找出一定的規律。本研究篩選出常見技術指標做為分析變數，並採用均等模糊化以及 CPDA 模糊化方法進行模糊化，且將其漲跌幅度設定為分類目標進行預測。在方析方法上，本研究採用隨機森林、支持向量機做為分析工具。在實證方面採用2014年到2018年實際股市資料來進行驗證，並比較分析其結果。本研究貢獻有：

1. 應用均等模糊化與 CPDA 模糊化方法建立分類模型。
2. 採用台灣股市2014~2018年實際股市資料進行實例分析。
3. 比較不同分類技術於模糊化與未模糊化資料集之效益，並採用10-folds 進行驗證比較。

關鍵字：股市技術指標、模糊理論、模糊分類技術、股價預測

A study on application of fuzzy classification techniques and technical indicators to stock price prediction

Student : JYUN-YI GUO

Advisor: JIA-WEN WANG, Ph. D.

Department of Information Management
Nanhua University
Master Thesis

ABSTRACT

The study screens out the technical indicators as analysis parameters. Using equal fuzzification and fuzzification of cumulative probability distribution approaches (CPDA) to forecast the stock fluctuations. In the method of analysis, this study uses random decision forests and support vector machine (SVM) as the analysis tools. In the empirical aspect, compared and analyzed the result from the stock data of 2014-2018 in Taiwan. It has three advantages,

1. Building the fuzzy classification model base on equal fuzzification and CPDA methods.
2. Using the practical data to compare the results from the stock data of 2014-2018 in Taiwan.
3. Comparing the different classification techniques for fuzzified and non-fuzzified data sets by 10-folds cross validation.

Keywords: Stock indicators, Fuzzy Theory, Fuzzy Classification

Techniques, Stock price forecasting

目錄

論文著作財產權同意書	I
論文指導教授推薦函	II
誌謝	III
中文摘要	IV
英文摘要	V
目錄	VI
表目錄	VIII
圖目錄	XI
第一章緒論	1
1.1. 研究背景與動機	1
1.2. 研究目的	3
1.3. 研究結構	3
第二章 文獻探討	4
2.1 常見技術指標	4
2.2 技術指標相關文獻回顧	12
2.3 分類技術應用之相關文獻	15
2.4 驗證方法	19
2.5 模糊理論介紹以及相關文獻	21
第三章 研究方法	25
3.1 研究架構	25
3.2 累積機率分配方法	29
3.3 例子說明	32
3.4 小結	35
第四章實證比較	37
4.1. 數據分析	37
4.2. 研究結果分析	51
第五章結論	58

參考文獻	-----	60
附 錄 一	-----	65
附 錄 二	-----	71



表 目 錄

表 2.1 技術指標文獻回顧整理	14
表 2.2 Precision and Recall	21
表 3.1 各欄位資料	27
表 3.2 明確值資料表	28
表 3.3 模糊化資料	28
表 3.4 五個語言值的語言區間值	30
表 3.5 K(9)模糊集	31
表 3.6 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	33
表 3.7 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)	33
表 3.8 SVM 之明確值混淆矩陣(以 10-Folds 為例)	33
表 3.9 隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	34
表 3.10 隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)	34
表 3.11 隨機森林之明確值混淆矩陣(以 10-Folds 為例)	35
表 3.12 在 1%漲跌幅中鴻海 SVM 以及隨機森林正確率比較	35
表 4.1 鴻海之 SVM 之明確值混淆矩陣(以 10-Folds 為例)	38
表 4.2 鴻海之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)	38
表 4.3 鴻海之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	39
表 4.4 鴻海之隨機森林之明確值混淆矩陣(以 10-Folds 為例)	39
表 4.5 鴻海之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)	40
表 4.6 鴻海之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	40
表 4.7 鴻海 SVM 以及隨機森林正確率比較	40
表 4.8 台塑之 SVM 之明確值混淆矩陣(以 10-Folds 為例)	41
表 4.9 台塑之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)	42
表 4.10 台塑之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	42
表 4.11 台塑之隨機森林之明確值混淆矩陣(以 10-Folds 為例)	43
表 4.12 台塑之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)	43

表 4.13	台塑之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	43
表 4.14	台塑 SVM 以及隨機森林正確率比較	44
表 4.15	台積電之 SVM 之明確值混淆矩陣(以 10-Folds 為例)	45
表 4.16	台積電之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)	45
表 4.17	台積電之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	45
表 4.18	台積電之隨機森林之明確值混淆矩陣(以 10-Folds 為例)	46
表 4.19	台積電之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)	46
表 4.20	台積電之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	47
表 4.21	台積電 SVM 以及隨機森林正確率	47
表 4.22	南亞之 SVM 之明確值混淆矩陣(以 10-Folds 為例)	48
表 4.23	南亞之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)	48
表 4.24	南亞之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	49
表 4.25	南亞之隨機森林之明確值混淆矩陣(以 10-Folds 為例)	49
表 4.26	南亞之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)	50
表 4.27	南亞之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	50
表 4.28	南亞 SVM 以及隨機森林正確率比	50
表 4.29	鴻海在持平之中的明確值與模糊值之 Recall 比較	53
表 4.30	鴻海在持平之中的明確值與模糊值之 Precision 比較	53
表 4.31	鴻海在下跌之中的明確值與模糊值之 Recall 比較	54
表 4.32	鴻海在下跌之中的明確值與模糊值之 Precision 比較	54
表 4.33	鴻海在上漲之中的明確值與模糊值之 Recall 比較	55
表 4.34	鴻海在上漲之中的明確值與模糊值之 Precision 比較	55
表 4.35	鴻海之 SVM 之明確值混淆矩陣(以 10-Folds 為例)	57
表 4.36	鴻海之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	57
表 6.1	台塑在持平之中的明確值與模糊值之 Recall 比較	65
表 6.2	台塑在持平之中的明確值與模糊值之 Precision 比較	65
表 6.3	台塑在上漲之中的明確值與模糊值之 Recall 比較	66
表 6.4	台塑在上漲之中的明確值與模糊值之 Precision 比較	66
表 6.5	台塑在下跌之中的明確值與模糊值之 Recall 比較	67

表 6.6 台塑在下跌之中的明確值與模糊值之 Precision 比較	67
表 6.7 南亞在下跌之中的明確值與模糊值之 Recall 比較	68
表 6.8 南亞在上漲之中的明確值與模糊值之 Precision 比較	68
表 6.9 南亞在上漲之中的明確值與模糊值之 Recall 比較	69
表 6.10 南亞在上漲之中的明確值與模糊值之 Precision 比較	69
表 6.11 南亞在下跌之中的明確值與模糊值之 Recall 比較	70
表 6.12 南亞在上漲之中的明確值與模糊值之 Precision 比較	70
表 7.1 南亞之 SVM 之明確值混淆矩陣(以 10-Folds 為例)	71
表 7.2 南亞之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)	71
表 7.3 南亞之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	71
表 7.4 南亞之隨機森林之明確值混淆矩陣(以 10-Folds 為例)	72
表 7.5 南亞之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)	72
表 7.6 南亞之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	72
表 7.7 台積電之 SVM 之明確值混淆矩陣(以 10-Folds 為例)	73
表 7.8 台積電之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)	73
表 7.9 台積電之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	73
表 7.10 台積電之隨機森林之明確值混淆矩陣(以 10-Folds 為例)	74
表 7.11 台積電之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)	74
表 7.12 台積電之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	74
表 7.13 台塑之 SVM 之明確值混淆矩陣(以 10-Folds 為例)	75
表 7.14 台塑之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)	75
表 7.15 台塑之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	75
表 7.16 台塑之隨機森林之明確值混淆矩陣(以 10-Folds 為例)	76
表 7.17 台塑之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)	76
表 7.18 台塑之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)	76

圖目錄

圖 2.1 SVM 概念	17
圖 2.2 均等模糊函數(以 5 個語意為例)	24
圖 3.1 研究架構	26
圖 4.1 四家公司收盤價波動圖	52



第一章緒論

1.1.研究背景與動機

在現今的社會，投資理財已經是現代人日常之一，而投資理財領域中，包含著各種理財方式，例如：外匯、股票、基金等等，而一直以來股票則是為大部分頭資者的理財工具之一。股票，是一種有價證券，是為一股票公司需要籌措資金將所有權藉由這種有價證券方式進行分配，然後將股票發給投資者作為資本憑證，投資者並且可以藉由公司獲利而分享股利，並分享公司成長或交易市場波動帶來的利益，但同時也須為了公司而承擔失敗的風險。

而在股票投資分面存在著高報酬以及高風險，所以為了讓投資者能有依據以及降低風險的目的下，過去文獻則會利用股市的歷史資料來分析預測漲跌(吳嘉川，2008；江旻緯，2014)，並利用開盤價換算各種新指標，我們把這些指標稱技術分析指標，例如：KD、RSI、MACD 等等，以利於各種股市環境之預測。

機器學習為近年興起的領域，涉及機率學、統計學、計算複雜理論等等多門學科，機器學習主要概念是在數據中分析找出規律，並透過規律來對未來進行預測。而機器學習演算法能在大數據中分析出有效並且重要的結果，有助於預測金融相關產業之交易(Gerlein et.al, 2016)。吳嘉川在 2008 年利用技術指標以及機器學習來預測數家上市

公司漲跌幅來預測漲跌幅，並且利用 5-Folds 進行驗證，由結果得知，SVM 分類結果有不錯的表現。江旻偉在 2014 年利用技術指標、改良式 SVM 以及其他機器學習方法來預測數家上市公司以漲跌幅來預測漲跌幅，結果得知，改良 SVM 均優於其他方法。謝政翰和陳振東在 2019 年利用技術指標以及機器學習來預測數家上市公司，並以當日技術指標(開盤價差、最高價差、最低價差.....)來預測當日漲跌幅，並且以模糊化方法再進行比較，其結果發現利用技術指標來預測當日漲跌幅擁有較高之準確率並且結合模糊理論能更有效地進行預測。

模糊理論由有許多前人研究證實，不論是應用在預測歷年大學入學人數(Chen, 1996)、氣溫(林宇恆, 2015)、以及台灣加權股價指數(TAIEX)等，都有良好的表現。至今已有許多研究者進行模糊化改良(Chen, 1996)，皆使其能得到更好的結果，本研究利用 Chung-Ho Su 等人於 2010 年提出之累積機率分配 (Cumulative Probability Distribution Approaches, CPDA) 來進行模糊化，從而得到了更好的預測表現，而數值能運用 CPDA 方法來產生歸屬函數並得到更主觀的歸屬值(謝睿帆, 2008)。

綜合上述，本研究利用上述所提到之機器學習分類技術以及模糊理論以股市技術指標來進行股票漲跌幅預測，探討其中之相關影響性。

1.2.研究目的

本研究先將全曜財經資訊股份有限公司(簡稱 CMoney)所開發的 CMoney 投資決策支援系統中擷取股市技術指標，並利用文獻探討法針對常見之技術指標進行挑選，最後從 97 個技術指標篩選出 28 個常見技術指標，再利用模糊化與分類技術方法(SVM、隨機森林)來做分析工具，最後利用交叉驗證(10-Folds)、Recall 以及 Precision 來結果評估，本研究目的如下：

1. 應用均等模糊化以及 CPDA 模糊化方法建構分類模型，進行實例股市分析比較。
2. 採用 2014~2018 年實際四家股市資料進行實例分析。
3. 利用模糊分類技術與股市技術指標結合 10-Folds 交叉驗證，來探討模糊值與明確值之關聯性。

1.3 研究結構

本研究共分為五個章節，第一章為緒論：主要闡述研究背景動機與目的，第二章為文獻探討：主要介紹常用技術指標、分類技術、模糊技術以及與本研究相關之各種文獻，第三章為研究方法：介紹本研究之架構與流程，並帶入一小例子進行實例說明，第四章為實證分析：採用四家公司 2014~2018 年進行分析比較，第五章為結論：說明本研究之發現並提出未來展望與限制。

第二章 文獻探討

本章第一節為常見技術指標、第二節技術指標相關文獻回顧、第三節為分類技術應用之相關文獻、第四節為驗證方法與第五節為模糊理論介紹以及相關文獻。

2.1 常見技術指標

技術分析學派投資人認為股價會一再複製過去的歷史(周聖鈞，2012)，因此可以利用此特性找出一定的規律而從中獲得利潤。股市市場上較著名之技術指標有 KD 隨機指標、RSI(Relative Strength Index)、MACD(Moving Convergence-Divergence)、乖離率 (BIAS)等等，但各個技術指標都有其優缺點。因此本研究將以股市技術指標進行詳細說明。

➤ KD 隨機指標

KD 線由美國作者(George Lane)在 1957 年所提出，而目前正在使用的 KD 指標則是 1986 年所改良，其融合了動量觀念、移動平均線及相對強弱指標原理之優點發展而成，屬於一種短期而敏感的指標，適合應用於短期波動幅度迅速的商品。其理論基礎則是股價上漲時，當日收盤價總是朝著向當日價格波動的最高價接近。相反的，當股價下跌時，當日收盤價總是朝著向當日價格波動的最低價接近。

此指標需先計算未成熟隨機值(Raw Stochastic Value, RSV)，再以

RSV 的三日指數平滑移動平均線(EMA)，分別可算出快速隨機指標(K線)及慢速隨機指標(D線)，最後則運用 KD 線的交叉點來研判股市之間的買賣時機。KD 隨機指標融合了相對強弱指標買超及賣超現象及移動平均線的觀念。而 KD 線的隨機指標的計算分是下方所示：

步驟 1. 先求出 T 日之未成熟隨機值(Raw Stochastic Value)

$$RSV_t = \frac{C - TL}{TH - TL} \times 100 \quad \text{其,} \quad (1)$$

C=當日收盤價

TH=最近 T 天的最高價

TL=最近 T 天的最低價

步驟 2. 再求出平滑隨機值，K(其中平滑係數通常設為 1/3，第一個 K 值通常設為 50。)

$$K_t = K_{t-1} \times \frac{1}{3} + RSV_t \times \frac{1}{3} \quad (2)$$

步驟 3. 再求出平滑隨機值，D(其中平滑係數通常亦設為 1/3，第一個 D 值通常亦設為 50。)

$$D_t = K_{t-1} \times \frac{1}{3} + D_{t-1} \times \frac{2}{3} \quad (3)$$

步驟 4. 由 K 與 D 的計算公式得知，K 值的變動比 D 值的變動大，通常當 K 值由上跌破 D 值往下時為賣出的訊號；而當 K 值由下突破 D 直往上時為賣出的訊號。

➤ 移動平均線(Moving Average, MA)

移動平均線(MA)所代表的正是某一期間投資人的平均成本，而研究常把移動平均線的參數（天數）設為 10 天、20 天、60 天，可以使其更快速的看出短、中、長期的價格趨勢。

王培宇(2012)將提到葛藍碧八大法則利用價格與其移動平均線的關係作為買進與賣出訊號的依據。認為價格的波動具有某種規律，但移動平均則代表著趨勢與平均的買進成本。

➤ 指數平滑異同移動平均線 (Moving Average Convergence-Divergence, MACD)

MACD 為 Appel & Hitschler (1979)所提出的交易法，其原理是利用快速及慢速兩條指數平滑移動平均線，計算兩者的差離值 (DIF)，再利用差離值即差離值平均值 (DEM) 的收斂及發散徵兆，因而找出股市或個股的買進或賣出時機。其公式如下：

步驟 1. 計算需求指數(Demand Index, DI)：

$$DI_t = \frac{H_t + L_t + 2 \times C_t}{4} \quad \text{其,} \quad (4)$$

H_t ：第 t 期最高價

L_t ：第 t 期最低價

C_t ：第 t 期收盤價

步驟 2. 計算 T1、T2 日 EMA(T1 短天期、T2 長天期)

$$EMA_t = EMA_{t-1} + \alpha(DIF_t - EMA_{t-1}) \quad \text{其,} \quad (5)$$

$\alpha = 2 / (1 + \text{移動平均天數})$

α ：平滑常數(smoothing constant)

步驟 3. 計算 DIF_t

$$DIF_t = T1 \text{ 日 EMA} - T2 \text{ 日 EMA} \quad (6)$$

步驟 4. 計算 MACD

$$MACD_t = MACD_{t-1} + \alpha(DIF_t - MACD_{t-1}) \quad (7)$$

步驟 5. 一般買賣點研判標準如下：

- A. 當 DIF 與 MACD 為正值時，表示市場處於上升行情，若 DIF 由下往上突破 MACD 時發出買進訊號；若 DIF 由上往下跌破 MACD 時發出賣出訊號。
- B. 當 DIF 與 MACD 為負值時，表示市場處於下跌行情，若 DIF 由上往下跌破 MACD 時發出賣出訊號；若 DIF 由下往上突破 MACD 時發出買進訊號。

➤ 相對強弱指標(Relative Strength Index, RSI)

RSI 為國內應用最普遍的技術指標，若與大盤指數走勢圖搭配，能有效預測股市的買賣超現象。

- A. T 日 RSI 85 以上或 15 以下，可能出現反轉的信號。
- B. 高 RSI 區域低 RSI 區時會漸形遲鈍，而失去其作用。
- C. RSI 可以頭肩頂，頭肩底，三角型等型態作買賣點的訊號。

- D. RSI 可切線，找阻力及支撐線。
- E. RSI 重要訊號之一即和股價背離。
- F. RSI 50 以下弱勢市場，50 以上強勢市場。
- G. RSI 在 50 以上的準確性較高。

1978 年美國作者華德在其著作『技術交易系統中的新觀念』中所提之交易方法之一。相對強弱指標(RSI)為先行指標的一種，而其是以一定期間內商品(股票或指數)價格的變動關係為基礎，去推敲其未來價位的變動方向。基本原理是利用在正常股市中多空買賣雙方的力道必預取得均衡，股價才會穩定。而 RSI 是計算在一定期間內，股價上漲總幅度平均值佔總漲跌幅總幅度平均值的比例，RSI 值會介於 0~100 之間。

計算公式如下：

$$T \text{ 日 RSI} = \frac{UP_{avg(t)}}{(UP_{avg(t)} + DOWN_{avg(t)})} \times 100, \text{ 其}, \quad (8)$$

$UP_{avg(t)}$ = T 天中收盤價上漲幅度的平均數。

$DOWN_{avg(t)}$ = T 天中收盤價下跌幅度的平均數。

$$T \text{ 日 RSI} = \frac{(T \text{ 日內上漲總幅度平均值})}{(T \text{ 日內上漲和下跌總幅度平均值})} \quad (9)$$

➤ 威廉指標 (Williams%R)

威廉指數是由威廉斯(Larry Williams)於 1973 年在『How I Made A Million Dollars』中首先提出的，原名為威廉斯超買超賣指標，而在

1999 年出版的『Long-term secrets to short-term trading』之中做更詳細的介紹。

威廉指標是運用股市的擺動點來量度其超買超賣現象，可以量度循環期內的高點或低點，提出有效率的投資訊號，其為一有效的短期指標，威廉指標算出來的值介於 0~100 之間，越接近 0 代表收盤價越接近 N 天內的最低價，越接近 100 代表收盤價越接近 N 天內的最高價，其公式如下：

$$\text{威廉指標}(W\%R_t) = \frac{H_t - C_t}{H_t - L_t} \times 100 \quad \text{其,} \quad (10)$$

H_t ：t 日內最高價

L_t ：t 日內最低價

C_t ：第 t 日內收盤價

而其與 KD 隨機指標比較，不同點在於分子部分。

➤ 乖離率(BIAS)

乖離率(BIAS)常被以 72 日(或 60 日)移動平均線算出來，而被做為研判大盤指數漲跌中賣賣點的較佳技術指標。而股市距離移動平均線之距離稱為乖離，將乖離再除以移動平均線得出乖離率，其公式如下：

$$\text{BIAS}_t = \frac{C - \text{MA}_t}{\text{MA}_t} \times 100\% \quad \text{其,} \quad (11)$$

C：當日內收盤價

MA_t ：t 日移動平均線

移動平均線代表投資人平均資本，而乖離率則是平均報酬率。當股價在平均移動線上，乖離率會為正值，正值越大時，代表大多數投資人皆處於獲利狀態且獲利頗豐富，故正乖離率過大時，股價回檔整理的可能越高。反之，當股價在平均移動線下 0，乖離率會為負值，負值越大時，代表大多數投資人虧數越為嚴重，故當負乖離率過大時，股價反彈機率越高。

➤ 停損點轉向操作空間(SAR)

停損點轉向操作空間(SAR)為 1978 年美國作者華德(Welles Wilder Jr.)在其著作『技術交易系統中的新觀念』中所提之交易方法之一。其在明確而快速的趨勢中，由於聽損價位的軌跡類似拋物線，所以又被稱為拋物線交易系統。

而其計算方式如下：

1. 數據：AF 加速因子在交易的第一天起，就以 0.02 為起始值。

未來若在漲勢中，股價美創新高，或是在跌勢中，股價美創新低 AF 便依次遞增 0.02，其最高上限為 0.20。故若 AF 因數次遞增而大於 0.20，概以 0.20 代之。

2. SAR 計算：分為漲勢買進或跌勢賣出時兩種買賣：

➤ 漲時買進：

$$SAR_t = SAR_{t-1} + AF(EPL - SAR_{t-1}) \text{ 其,} \quad (12)$$

EPH：近期相對最高價

當股低點向下穿越 SAR 值時，改為跌勢賣出公式，計算新 SAR。

➤ 跌時賣出

$$SAR_t = SAR_{t-1} + AF(SAR_{t-1} - EPL) \text{ 其,} \quad (13)$$

EPL：近期相對最低價



2.2 技術指標相關文獻回顧

陳逸平、李俊德(2008)研究在股市中、短線操作上，投資者較常用的 KD 隨機指標與 WMS%R 威廉指標兩種技術指標，並結合類神經網路自我訓練、自我學習的優點，來研究這兩種技術指標的預測能力及可靠性。

李良俊(2003)以移動平均線(MA)、隨機指標(KD)、相對強弱指標(RSI)和乖離率(BIAS)組合出 15 種不同的操作策略作為股票買進和賣出之依據，挑選臺灣股市中上市和上櫃共 118 檔個股為研究樣本，研究發現綜合越多的技術指標為操作依據未必有更好的績效，此外這 15 種操作策略中有 12 種策略優於買進持有策略。

蕭政倫(2013)以台灣加權股價指數為例運用移動平均線(MA)、相對強弱指標(RSI)以及趨向指數(DMI)三個技術指標建立技術交易模型，並探討此交易策略對股票市場之擇時能力及獲利能力，得到無論在多頭市場或是空頭市場，技術交易策略的報酬率顯著大於買進持有策略的報酬率。周照偉等人(2015)研究結果發現技術分析十五種操作策略及買進持有策略，皆顯著高於銀行定存，四種基礎指標中，DMI 和 MACD 績效最佳，其次是 KD，最後才是 RSI。

Weia 等人(2017)採用加權企業價值來構建市場的“綜合新聞情緒指數”(ANSI)，並以市場回報(Ret)，交易價值(Value)，周轉率(Turn)

和波動率指數(Vix)，4 項技術指標來確定 ANSI 的超前滯後效應。謝政翰(2019) 技術分析是可讓投資人利用過去的價格資訊而獲得超額的報酬，顯示了技術指標在台灣股票市場上確實有其效用存在。

蕭政倫在 2013 年利用隨機指標(KD)、相對強弱指標(RSI)、指數平滑異同平均線(MACD)及趨向指標(DMI)以台灣加權股價指數日資料為研究樣本，研究期間自 2003 年 1 月 2 日至 2012 年 12 月 28 日止。結果得知無論在多頭市場或是空頭市場，技術交易策略的報酬率顯著大於買進持有策略的報酬率。

綜合上述，因股市技術指標眾多，本研究整理常見技術指標之文獻，如下表 2.1。運用文獻探討法，整理出過去研究中常出現之技術指標，得出 28 個特徵變數以及 1 個預測變數，進行資料前處理，詳細研究過程請參閱第 3 章。

表 2.1 技術指標文獻回顧整理

技術指標	介紹	應用之文獻
KD 隨機指標	融合了動量觀念、移動平均線及相對強弱指標原理之優點發展而成，屬於一種短期而敏感的指標，適合應用於短期波動幅度迅速的商品。	李良俊(2003) 王培宇(2012) 陳玟志(2012) 徐偉成(2016) 謝政翰(2019) 吳睿萍(2018) 曾宇皓(2018)
相對強弱指 標 RSI	相對強弱指標(RSI)為先行指標的一種，而其是以一定期間內商品(股票或指數)價格的變動關係為基礎，去推敲其未來價位的變動方向	李良俊(2003) 陳應慶(2003) 蕭政倫(2013) 周照偉等人 (2015) 謝政翰(2019)
指數平滑異 同移動平均 線 MACD	利用快速及慢速兩條指數平滑移動平均線，計算兩者的差離值 (DIF)，再利用差離值即差離值平均值 (DEM) 的收斂及發散徵兆，因而找出股市或個股的買進或賣出時機。	陳應慶(2003) 周照偉等人 (2015) 吳睿萍(2018)
威廉指數 W%R	運用股市的擺動點來量度其超買超賣現象，可以量度循環期內的高點或低點，提出有效率的投資訊號，其為一有效的短期指標	陳玟志(2012) 徐偉成(2016)
乖離率 (BIAS)	大盤指數收盤價移動平均線的距離。	李良俊(2003)
市場波動率	因各種因素與市場投資者自身因素的變化引起對市場預期的變化而產生的股票價格的變化	葉孝賢(2018)
Beta	評估股市系統風險之工具，用來度量股市或投資組合相對總體市場之波動性。	戴熠星(2016)
保力加通道	結合了移動平均線以及標準差之概念，基本型態為頂端、中間以及下端三部分，而中間為股價的平均成本、上端以及下端分別為壓力線以及支撐點	盧廷宇(2011) Prasetijo et al. (2017)

停損點轉向 操作 SAR	在買賣某個股票之前，先設定一個停損點，來減少投資風險。	吳睿萍(2018)
TR	當日股價真正的波動範圍	Weia et al. (2017)
DIF	快速以及慢速兩條移動平均線相減之值	陳應慶(2003) 王培宇(2012)
ADX	指示市場趨勢的強弱程度，但不能指示趨勢的方向	林素娥(2016) Prasetijo et al. (2017) 曾宇皓(2018)

2.3 分類技術應用之相關文獻

本研究因主要預測主要利用分類技術法來分析預測股票漲跌。下列為本研究所用分類技術之介紹以及相關文獻之應用。

2.3.1 支持向量機 (Support Vector Machine, SVM)

支持向量機 (Support Vector Machine, SVM) 是由 Vapnik 等人於 1995 所發展出來的統計學習理論 (statistical learning theory) 衍生而成的學習演算法，是一種監督式分類演算法，主要用在分類 (Classification) 和迴歸 (Regression) 上。

近年使用 SVM 進行研究的領域相當廣泛，過去文獻常用 SVM 來進行分析，如黃承龍等學者 (2004) 從 UCI 資料庫之信用卡領域取得兩個資料集，使用支援向量機進行分類。其實驗結果，在分類的正確率有不錯的表現。楊維珊 (2009) 以半監督式人臉辨識，藉由 SVM 進行分析，建立辨識規則。實驗結果證明此研究的系統對新資料及新類別擁有較佳的學習效率以及系統彈性。林俐玲等學者 (2011) 以支持向量機探討山坡地土壤沖蝕程度，實驗成果證實支持向量機提升土

壤沖蝕程度等級判定效率。江旻緯(2014)利用改良式 SVM、決策樹以及貝氏分析來進行股票漲跌預測，並以漲跌幅以及反應日期為參數之一來進行預測。根據結果得知在相同環境中改良式 SVM 都會有較高之準確率。以下針對支持向量機進行概念說明。

支持向量機(SVM)主要是可以在多維度中的訓練資料模組 (Train model) 找出一個超平面 (Hyper-plane) 將資料正確切割並分成數個類別。一般而言，訓練資料集 (Training Set)，其中提供最多資訊的資料點稱為 Support Vector，而藉由最靠近超平面的 Support Vector 形成該資料類的邊界，圖 2.1 中 Margin 是指超平面與邊界的距離，Gap 則是指邊界與邊界之間的距離，而如果我們取得 Margin 值最大化，代表著此模型對於未知的資料點能較精確分辨其類別。因此，支持向量機在解決線性、非線性、高維度空間及小樣本問題，其訓練模型擁有特有的優勢。

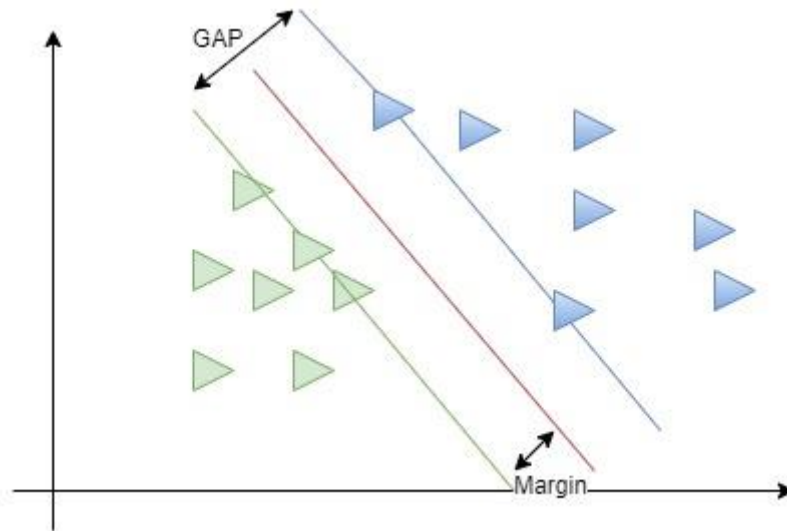


圖 2.1 SVM 概念

➤ 線性支援向量機(Linear Support Vector Machines)

資料本身沒有任何誤差的理想的狀態下，超平面可以將兩類資料完全分成兩群。然而，實際上資料可能會落在空間的任何位置，有時也會很難使兩群不同類別的資料完全分離(Burges, 1998; 吳佳霖, 2016)。

➤ 非線性支援向量機(Nonlinear Support Vector Machines)

當原始資料資料空間中無法以線性函數區分類別時，可透過映射函數，將低維度空間(low dimensional)的線性不可分割資料映射至高維度的特徵空間(high dimensional)，改善不同類別資料分佈重疊的情況，使特徵空間中以線性函數建構出超平面，藉此來區分不同類別的資料(Boser, Guyon, & Vapnik, 1992)。

在非線性支持向量機中，常見的核函數(kernel function)如下

(Burges, 1998 ; Vapnik , 1995) :

一、 線性核心函數 Linear kernel function :

$$K(x_i, x_j) = x_i^T \cdot x_j \quad (16)$$

二、 多項式核心函數 Polynomial kernel function :

$$K(x_i, x_j) = (x_i^T \cdot x_j + r)^d, d \in \mathbb{N} \quad (17)$$

其中 γ 、 r 、 d 為核心參數。

三、 高斯核心函數 Gaussian Kernel(Radial basis function kernel, RBF) :

$$K(x_i, x_j) = \exp\left(-\gamma \left\|x_i - x_j\right\|^2\right), \gamma > 0 \quad (18)$$

其中 γ 為核心參數

本研究採用林智仁博士所開發的 SVM 分析套件 LIBSVM。LIBSVM 是一種整合性的 SVM 分析套件，可應用 SVM 分析、迴歸以及分配估算；傳統 SVM 只能解決二元分類問題，LIBSVM 之套件中擁有多類分類機制，故本研究採用 WEKA 3.8 自由軟體並利用 LIBSVM 之套件進行分析。因本研究資料筆數與變數較多，利用線性較無法準確分離，因此採用非線性函數來做為分析函數。

2.3.2 隨機森林

隨機森林其中的每一棵樹都是根據隨機向量值建立，而隨機向量是依據固定機率分配產生。其用隨機方式根據資料筆數或屬性特徵隨

機挑選建立決策樹，使每棵樹都相互獨立，最後依投票取得最終結果

(Breiman, 2001; 葉子維, 2018)，過去決策樹的缺點如下：

- (1) 分類規則複雜。
- (2) 收斂程度並未達到最佳。
- (3) 容易過度適配而發生 Overfitting 的現象。

隨機森林克服以上缺點，它具有以下特徵：

- (1) 能有效的處理大資料量。
- (2) 能處理較多的特徵變數。
- (3) 能有效分類過程中生成一個內部的泛化誤差的估計值。
- (4) 能有效處理遺失值也不會影響正確率。
- (5) 能處理偏斜的資料。
- (6) 抽樣未使用到的資料能作為驗證資料。

但隨機森林也有缺點，此方法無法清楚說明決策過程，導致模型無法被人所理解，所以被人們稱為黑盒子現象。

2.4 驗證方法

2.4.1 交叉驗證法

變數預測模型建立時，內部驗證(Internal validation)為尋找最佳預測模型的方法，在內部驗證程序中會評估各預測因素中的組合，找出預測最佳之預測因素所建立之模型，並且通過預測效果的評估，預測

效果的評估包含了校準(calibration)與區別(discrimination)兩大類。本研究將利用 SVM 以及隨機森林結合交叉驗證法來找出最佳之預測模型。

在建立預測模型的相關研究中，通常都會將資料歸類為訓練組與測試組，由訓練組建立預測模型，在來將測試組利用訓練組所建立模型而進行測試，來得到整體效能。交叉驗證法可以評估所有的訓練組與測試組的組合方式，並進行平均得到模型整體的真實預測效果，目前常用的交叉驗證法可分為 k 組的交叉驗證法(k-fold cross-validation)與留一交叉驗證法(leave-one-out cross-validation)兩種方法。

本研究所使用之方法為 10 組的交叉驗證(10-folds)，將樣本資料分為 10 組，每一次都取 9 組為訓練組，剩餘 1 組為測試組，由訓練組先建立模型在由測試組進行預測結果，最後則是將 10 次結果平均，得到模型的準確率，也可以避免過度擬合 (overfitting)的問題。

2.4.2 精確率與召回率 (Precision and Recall)

精確率與召回率 (Precision and Recall) 是一種判斷分類工具優劣勢的方法，精確率(Precision)代表分類結果之正確率；召回率(Recall)代表所有正確資料有多少被分類出來。

表 2.2 Precision and Recall

	實際 T 值	實際 F 值
判斷 T 值	True Positive (TP)	False Positive (FP)
判斷 F 值	False Negative (FN)	True Negative (TN)

Precision and Recall 公式為如下：

$$\text{Precision} = \frac{TP}{TP+FP} \quad (19)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (20)$$

2.5 模糊理論介紹以及相關文獻

模糊集合理論 (Fuzzy Set Theory) 又稱可能理論，是由美國加州大學柏克萊分校的自動控制學家 Lotif Zadeh 在 1965 年提出，其為一種定義表達工具，目的是為了表現一些無法明確定義的模糊概念。利用數值的方式來解釋生活中一些模糊現象，來處理不確定資訊，而更重要的是模糊集合理論讓我們能處理含糊不確定的事實。

在生活中，普遍存在各種無法明確定義的現象，其所表現的特徵比電腦計算進位所採用的二元邏輯 (Binary Logic) 只有 0 或 1 的表示較為複雜，但在日常生活中，我們常常因為無法簡單明確表達某些概念，使語言存在著模糊性，而模糊性也會因為人的主觀性而有所不同。目前有許多的領域以模糊集合理論為基礎的延伸(Chen, 1996；

謝政翰, 2019), 如: 模糊時間序列(Fuzzy time series)、模糊控制(Fuzzy control)、模糊邏輯(Fuzzy logic), 例如 Chen (1996) 利用均等模糊理論將入學人數利用模糊時間序列來預測入學率。根據結果得知, 在入學率方面有著顯著的提升。許睿華(2012) 結合多個技術指標與加權模糊時間序列所預測的股票漲跌訊號, 來決定股票的買賣時間點, 根據結果得知, 利用"持續買進, 賣出時的價格必須大於過去平均持有價格"的策略, 能穩定獲利。謝宗男(2012) 根據投資者的情緒分別利用遺傳演算法以及模糊三角歸屬函數來建立最佳模型與增加模型彈性度。根據結果得知, 在選取多股時, 應用模糊理論選股能有效提升其投資效率。陳振東(2018)利用模糊理論結合技術指標來進行漲跌幅預測, 其結果得知模糊理論應用在股票漲跌預測能提高模型正確率。

以下就模糊集合理論進行介紹, 在傳統集合理論中有個重要的概念, 即特徵函數數(Characteristic Function)。對其為一個領域的子集合 A 而言, 定義 $S_A(X)$ 為 A 集合的特徵函數:

$$S_A(X) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \quad (21)$$

如上公式(21)所述, 若 x 為 A 的元素, 則其特徵函數 $S_A(X)$ 為 1; 反之, 則其特徵函數 $S_A(X)$ 為 0。其值域可寫成 $S_A(X) = \{1, 0\}$ 。

Zadeh 將傳統集合理論特徵函數從非此即後的二選一，推廣為可從 0 到 1 之間的任何值作選擇，此種特徵函數稱為歸屬函數 (Membership Function)。藉由歸屬函數對模糊集合進行量化，才能利用明確的數學方法分析和處理模糊性資訊。一般常用的歸屬函數有三角形、梯形、高斯函數…等 (許睿華, 2012)。而模糊值內的歸屬值能有無限種可能，藉由歸屬度的調整，可以更準確地來表達數值的實際情況。而模糊理論被廣泛應用，例如：醫療、自然語言、各種分析等等。

2.5.1 模糊化方法

本小節以模糊歸屬函數進行說明，主要針對三角歸屬函數以及梯形歸屬函數，公式如下(陳耀茂, 1999; 游士億, 2008; 周琬嬪, 2018)

(1) 三角歸屬函數：

$$\mu_{\tilde{A}_i}(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b \leq x \leq c \\ 0, & x \geq c \end{cases} \quad (22)$$

其中 a 代表歸屬函數下限、b 代表中點、c 代表上限。

(2) 梯形歸屬函數：

$$\mu_{\tilde{B}_i}(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & x \geq d \end{cases} \quad (23)$$

其中 a 代表左端點，b 代表左肩點，c 代表右肩點，d 代表右端點

而本研究所使用公式(22)以及(23)來推算均等模糊化，並使其各區間為一致，下圖 2.3 以及圖 2.4 分別是利用三角模糊數與梯形模糊數，

呈現五個語意的均等模糊化的概念圖。

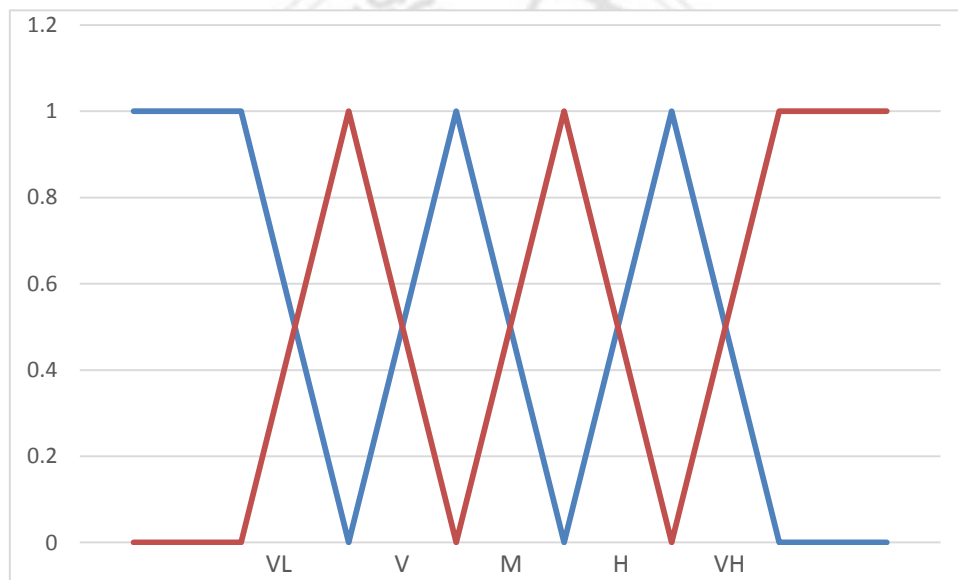


圖 2.2 均等模糊函數(以 5 個語意為例)

第三章 研究方法

本研究以漲跌幅度為預測變數，並以分類技術方法來進行比較，結合模糊化方法建立股票漲跌預測模型，最後將原始明確值與模糊值分析比較，本研究有以下優點：(1)模糊化方法能更有規律地去運算區間值並根據數值去做分布歸類，使預測效果更好。(2)CPDA 能有效地進行屬性切割，並得到更合理的歸屬值。(3)陳振東(2017)在漲跌幅預測中以當日進行預測，但技術指標為事後反應值，應要次日較為合理，因此本研究以次日漲跌為預測目標。(4)實際採用 2014-2018 實際股市資料來進行實例驗證。因此本研究為了上述幾點而利用 CPDA 模糊化與均等模糊化結合各種分類技術方法來探討預測效果並進行分析比較。

3.1 研究架構

首先本研究先將全曜財經資訊股份有限公司所開發的投資決策支援系統(簡稱 CMoney)中所得的技術指標共 97 項，並透過文獻探討法篩選 28 個常見之技術指標再利用模糊化與分類技術方法(SVM、隨機森林)來做分析工具，最後利用交叉驗證(10-Folds)來結果評估，以期提供後續研究參考依據。以下則是研究架構之步驟說明。

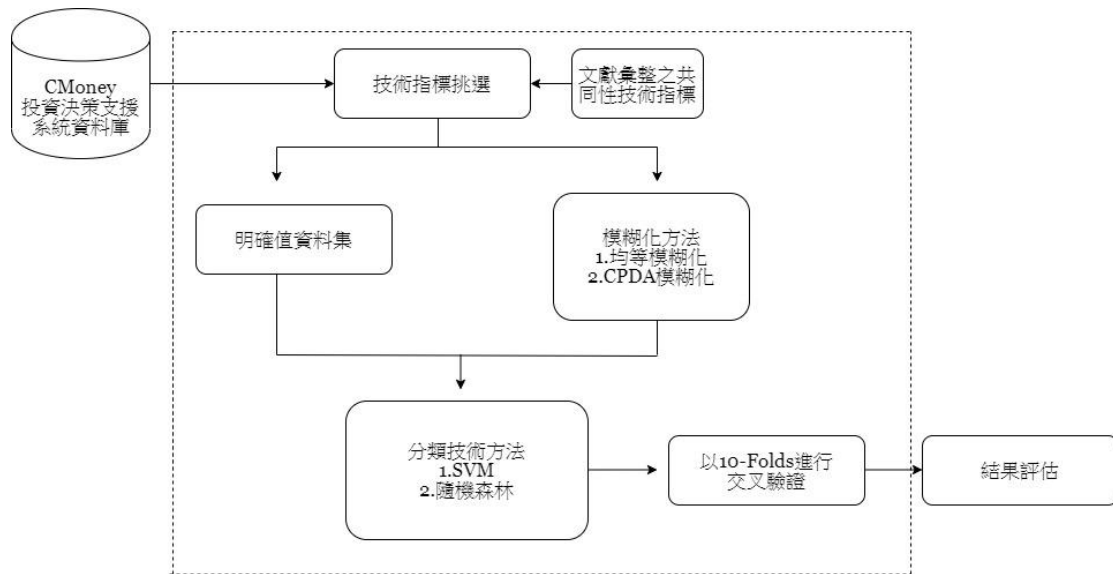


圖 3.1 研究架構

步驟 1. 資料來源

本研究實驗對象為鴻海進行測試，並從 CMoney 中蒐集該公司所需個股資料，並進行資料前處理，技術指標經由 Cmoney 系統中擷取資料。

步驟 2. 特徵變數篩選

本研究依文獻選取 28 個常見技術指標 (T1-T28) 作為做為特徵變數，以次日漲跌服做為預測變數(O1)，如下表 3.1。

表 3.1 各欄位資料

變數	代表變量	變數	代表變量
T1	隨機指標 K (9)	T2	隨機指標 D (9)
T3	相對強弱指標 RSI (5)	T4	相對強弱指標 RSI (10)
T5	指數平滑異同移動平均線 MACD	T6	DIF-MACD
T7	威廉指數 W%R (5)	T8	威廉指數 W%R (10)
T9	Beta (21D)	T10	Beta (65D)
T11	Beta (250D)	T12	市場波動率 (21D)
T13	市場波動率 (25D)	T14	乖離率 (20D)
T15	乖離率 (60D)	T16	乖離率 (250D)
T17	相對強弱指標 RSI (D)	T18	相對強弱指標 RSI (W)
T19	EWMA (%)	T20	保力加通道_頂端 (20)
T21	保力加通道_中端 (20)	T22	保力加通道_底端 (20)
T23	CM-VIX (%)	T24	停損點轉向操作 SAR
T25	TR (1)	T26	DIF
T27	ADX (14)	T28	當日收盤價
O1	次日漲跌幅		

步驟 3. 特徵變數模糊化

此步驟將特徵變數進行模糊化，本研究模糊化方法有兩種，

(1)均等模糊化(2)累積機率分配方法 (Cumulative Probability Distribution Approaches, CPDA)。本研究將針對明確值、均等模糊化以及 CPDA 進行後續實驗分析，下表 3.2 以及表 3.3 為明確值以及模糊值資料樣態。

表 3.2 明確值資料表

	T1	T2	T3	T28	漲跌
2014/1/3	56.8	68.61	34.38	102.5	S
2014/1/6	42.03	59.75	34.38	102.5	S
2014/1/7	28.02	49.17	30.29	102	P
2014/1/8	35.35	44.56	56.3	104	N
2014/1/9	26.6	38.57	33.12	101	S
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2018/12/27	41.49	35.36	58.37	223	P
2018/12/28	60.99	43.9	65.51	225.5	N

表 3.3 模糊化資料

	T1	T2	T3	T28	漲跌
2014/1/3	L3	L4	L2	L1	S
2014/1/6	L2	L3	L2	L1	S
2014/1/7	L1	L2	L1	L1	P
2014/1/8	L2	L2	L3	L1	N
2014/1/9	L1	L2	L1	L1	S
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2018/12/27	L2	L2	L4	L5	P
2018/12/28	L4	L2	L4	L5	N

步驟 4. 進行預測分類

本研究採用 SVM 以及隨機森林兩種常見的分類技術，針對漲跌幅來進行股票漲跌之預測，而次日漲跌幅公式如下：

$$\text{次日漲跌幅度 (\%)} = \frac{\text{收盤價}_{t+1} - \text{收盤價}_t}{\text{收盤價}_t}, \text{ 其 } t \text{ 為當日。} \quad (24)$$

參照過去方法當以預測漲跌幅上下 1% 為預測目標，因此本研究將漲跌幅度大於預設漲幅 (1%) 則標記為 P；若小於預設漲跌幅度 (-1%) 則標記為 N，其餘為持平則標記為 S(請參閱上表)。

本實驗將上述 28 項參數由明確值、均等模糊值以及 CPDA 模糊值經由不同的分析方法 (SVM、隨機森林) 以及漲跌幅度 O1 進行股價漲跌預測。在本研究最後則是採用交叉驗證 10-Folds 來做為評估，並比較 Precision、Recall 之結果。

3.2 累積機率分配方法 (Cumulative Probability Distribution Approaches, CPDA)

CPDA 有以下優點：(1) 產生屬性的歸屬函數值以得到更主觀的歸屬值；(2) 為了不同相似度函數門檻來產生規則，能使 CPDA 更有效的處理分類問題。本研究模糊化方法是根據 Chung-Ho Su 等人於 2010 年提出之累積機率分配 (Cumulative Probability Distribution Approaches, CPDA)，其根據累積分佈函數(CDF)的算法(公式 26)進行改良，CPDA 模糊化方法步驟說明如下：

$$P=F(x|\mu, \sigma)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} \quad (25)$$

P 表示具有參數 μ 和 σ 的正態分佈的觀測值落入區間。

步驟 1. 定義模糊區間 U

將 U 定義為目標數據集 $U[D_{min} - \sigma, D_{max} + \sigma]$ ，其中 D_{min} 表示最小值， D_{max} 表示最大值，而 σ 表示目標數據集中的標準偏差。以特徵變數 K(9) 為例，其 D_{min} 為 6.92、 D_{max} 為 96.55、 σ 為 23.37，故 $U[D_{min} - \sigma, D_{max} + \sigma]$ 為 $U(-16, 120)$ 。

步驟 2. 定義累積機率分配的下限和上限。

對於每個定的語言值，累積機率的下限(P_{LB})以及累積機率上限(P_{UB})由下方公式(27)~(29)所定義：

$$P_{LB}^1 = 1 - \sum_{i=2}^n P_{LB}^i \quad (26)$$

$$P_{LB}^i = (2i - 3) / 2n, (2 \leq i \leq n) \quad (27)$$

$$P_{UB} = i/n, (1 \leq i \leq n) \quad (28)$$

其中 i 為語言值的順序， n 為已定義語言值的數量。

以 K(9) 為例，其下限值 (P_{LB}) 和上限值 (P_{UB}) 範例如下。

表 3.4 五個語言值的語言區間值

語言值	語言間隔	
	P_{LB}	P_{UB}
L1	-16	38
L2	28	52
L3	45	64
L4	58	77
L5	70	120

步驟 3. 定義模糊集並建立歸屬函數

利用公式(22)來推斷其模糊歸屬程度，其中 $\mu_{\tilde{A}_i}(x)$ 表示 X 在模糊集的歸屬函數，而 a、b、c 分別為 $\mu_{\tilde{A}_i}$ 的下限、中點以及上限，如果該值滿足兩個或多個歸屬度函數，則選擇具有最大歸屬度值的語言值並將其標記在該值上，下表 3.3 展示了 K(9)之三角模糊數。

表 3.5 K(9)模糊集

語言值	三角模糊數		
	<i>a(lower bound)</i>	<i>b(midpoint)</i>	<i>c(upper bound)</i>
L1	-16	11	38
L2	28	40	52
L3	45	57	64
L4	58	64	77
L5	70	95	120

步驟 4. 模糊化歷史數據

對於價格波動的語言變量，使用 CDF 和參數化的三角模糊數，可以將數據中所有結果模糊化為語言值。

綜合上述，本研究利用技術指標進行模糊化，將數值轉化成五語意區間。模糊化方法採用方法主因 CPDA 模糊化優勢為分割後分布較合理，其分割會依資料散佈進行調整，而均等模糊化值不會因為數值而去改變，其採用固定區間，所以本研究採用 CPDA 進行來建立模型。

3.3 例子說明

本小節以鴻海為例，首先依照 3.1 之步驟 1 資料來源進行資料蒐集，主要蒐集鴻海 2014 年 4 月到 2015 年 4 月共 267 天，共 267 筆資料。

步驟 2 對於表 3.1 常見技術指標來做為特徵變數。

步驟 3 進行特徵模糊化，此步驟會產生 3 種資料集(1)明確值資料集，將上述 28 個原始明確值進行整理歸納；(2)均等模糊化，依照均等模糊化方法將上述 28 個參數模糊化得到均等模糊值；(3)CPDA 模糊化，利用累積機率分配方法(CPDA)來進行模糊化得到 CPDA 模糊值。

步驟 4 利用利用分類技術法來進行分析，在例子部分以支持向量機($c : 0.8$ ， $\text{Gamma} : 5$)與隨機森林為例。

步驟 5 則是利用 10-Folds 以及 Recall、Precision 來評估模型。其結果呈現如下

- 在漲跌幅 1%時，分析方法正確率比較
- 實驗對象：鴻海
 - 樣本數：267 筆
 - 特徵維度：28
 - 分析方法：SVM (參數設定： $\text{Gamma} : 5$ ， $c : 0.8$ ，多項

式核心函數)

表 3. 6 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	146	14	13	173	0.843931
P	32	6	7	45	0.133333
N	40	5	4	49	0.081633
total	218	25	24	正確率 58.42	
Precision	0.669725	0.24	0.166667		

表 3. 7 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	136	18	19	173	0.786127
P	29	9	7	45	0.2
N	40	4	5	49	0.102041
total	205	31	31	正確率 56.17	
Precision	0.663415	0.290323	0.16129		

表 3. 8 SVM 之明確值混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	110	28	35	173	0.635838
P	29	7	9	45	0.155556
N	35	7	7	49	0.142857
total	174	42	51	正確率 46.44	
Precision	0.632184	0.166667	0.1372549		

- 實驗對象：鴻海
- 樣本數：267 筆
 - 特徵維度：28
 - 分析方法：隨機森林

表 3.9 隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	123	22	28	173	0.710983
P	28	8	9	45	0.177778
N	35	8	6	49	0.122449
total	186	38	43	正確率 51.31	
Precision	0.66129	0.210526	0.139535		

表 3.10 隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	117	30	26	173	0.676301
P	31	7	7	45	0.155556
N	36	4	9	49	0.183673
total	184	41	42	正確率 49.81	
Precision	0.63587	0.170732	0.214286		

表 3.11 隨機森林之明確值混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	106	28	39	173	0.612717
P	25	8	12	45	0.177778
N	36	6	7	49	0.142857
total	167	42	58	正確率 43.51	
Precision	0.634731	0.190476	0.1206897		

表 3.12 在 1%漲跌幅中鴻海 SVM 以及隨機森林正確率比較

分析工具	方法		
	明確值	均等模糊化	CPDA 模糊化
SVM	46.44	56.17	58.42
隨機森林	45.31	49.81	51.31

在表 3.6~3.12 能發現 CPDA 在兩種分析方法皆高於明確值，而在模糊森林中，並能看到其模糊值正確率皆高於明確值，在第 4 章本研究將擴張資料集筆數來探討其正確率以及穩定度。

3.4 小結

在過去研究中，股票分析分為基本面分析與技術面分析，而以往研究中技術面分析常利用部分技術指標，本研究利用股價漲跌幅進行

預測預測，並且以 1%漲跌幅來進行觀察，在何種漲跌幅以及，另外以往研究中，股票預測分為長期趨勢與短期預測，本研究屬於短期預測，所以本研究希望透過觀察漲跌幅並以當天參數預測隔天收盤漲跌來建立模型。過往技術分析常以簡單貝氏分類或決策樹對於股票股價進行預測，有鑑於過去支持向量機之研究發現，支持向量機對於解決線性或非線性問題，都有較好的效能與較高的正確率，所以本研究利用 SVM、隨機森林於股票股價漲跌進行預測，針對上述之方法來找出對於此模型最好之選擇，在第四章會進行 5 年實例驗證，詳細細節請參閱第 4 章。

第四章實證比較

經由 3.3，本章節擴增資料筆數測試方法穩定度，實驗對象為台灣證券上市公司，並對其中挑選四家上市公司(台積電、鴻海、南亞與台塑)來進行實證，且利用金曜財金資訊股份有限公司(簡稱 CMoney)研發的 CMoney 投資決策資源系統來進行資料蒐集，資料期間為 2014 年 1 月至 2018 年 12 月底 5 年，共 1227 筆資料進行實證比較。本研究採用 3 種資料集，(1)明確資料集；(2)均等模糊值；(3)CPDA 模糊值，並以 SVM 以及隨機森林來進行分析。在實驗中，SVM 參數利用逼近方法採較佳結果參數進行說明($\text{Gamma} : 5$ 、 $c : 0.8$ ，多項式核心函數 (Polynomial kernel function) 為本研究之參數)，最後以 10-Folds、Recall 以及 Precision 來進行交叉驗證。

4.1. 數據分析

以下針對上述所提四家公司，實際分析案例依序呈現如下：實驗對象 A：鴻海、實驗對象 B：台塑、實驗對象 C：台積電、實驗對象 D：南亞

- 實驗對象 A：鴻海
- 樣本數：1227 筆
- 特徵維度：28
- 分析方法：SVM (參數設定： $\text{Gamma} : 5$ ， $c : 0.8$ ，多項式核心函

數)

從表 4.1~4.7 可以發現模糊化方法用於分類技術其正確率都較優於明確值如表 4.7，CPDA 模糊化方法其 SVM 正確率為 53.7 優於明確值 46.54，在隨機森林方面，均等模糊化與 CPDA 模糊化分別為 49.06 與 49.47 均高於明確值 47.51。從分析中可得知，在分類技術方便，不論是 SVM 以及隨機森林，均等模糊化以及 CPDA 模糊化都明顯優於明確值，並且 CPDA 更優於均等模糊化方法。

表 4.1 鴻海之 SVM 之明確值混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	452	173	151	776	0.582474
P	118	60	50	228	0.263158
N	105	65	53	223	0.237668
total	675	298	254	正確率 46.04	
Precision	0.66963	0.201342	0.208661		

表 4.2 鴻海之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	545	126	105	776	0.70232
P	155	35	38	228	0.153509
N	141	44	38	223	0.170404
total	841	205	181	正確率 50.36	
Precision	0.648038	0.170732	0.209945		

表 4.3 鴻海之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	585	102	90	777	0.752896
P	166	32	30	228	0.140351
N	144	37	42	223	0.188341
total	895	171	162	正確率 53.70	
Precision	0.653631	0.187135	0.259259		

➤ 實驗對象 A：鴻海

- 樣本數：1227 筆
- 特徵維度：28
- 分析方法：隨機森林

表 4.4 鴻海之隨機森林之明確值混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	491	159	126	776	0.632732
P	132	53	43	228	0.232456
N	124	60	39	223	0.174888
total	747	272	208	正確率 47.51	
Precision	0.657296	0.194853	0.1875		

表 4.5 鴻海之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	531	123	122	776	0.684278
P	159	33	36	228	0.144737
N	144	41	38	223	0.170404
total	834	197	196	正確率 49.06	
Precision	0.636691	0.167513	0.193878		

表 4.6 鴻海之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	526	133	117	776	0.677835
P	142	38	48	228	0.166667
N	142	38	43	223	0.192825
total	810	209	208	正確率 49.47	
Precision	0.649383	0.181818	0.206731		

表 4.7 鴻海 SVM 以及隨機森林正確率比較

分析工具	方法		
	明確值	均等模糊化	CPDA 模糊化
SVM	46.04	50.36	53.70
隨機森林	47.51	49.06	49.47

➤ 實驗對象 B：台塑

- 樣本數：1227 筆
- 分析方法：SVM
- 特徵維度：28

從表 4.8~4.14 可以發現可以發現模糊化方法用於分類技術其正確率較優於明確值如表 4.14，CPDA 模糊化方法其 SVM 正確率為 64.14 優於明確值 50.69，在隨機森林方面，均等模糊化與 CPDA 模糊化分別為 56.96 高於明確值 55.05。從分析中可得知，在分類技術方面，在 SVM 中，均等模糊化以及 CPDA 模糊化都明顯優於明確值，而在隨機森林中均等模糊化明顯優於明確。

表 4.8 台塑之 SVM 之明確值混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	551	141	158	850	0.648235
P	129	33	27	189	0.174603
N	109	41	38	188	0.202128
total	789	215	223	正確率 50.69	
Precision	0.698352	0.153488	0.170404		

表 4.9 台塑之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	675	101	74	850	0.794118
P	149	27	13	189	0.142857
N	147	13	28	188	0.148936
total	971	141	115	正確率 59.49	
Precision	0.69516	0.191489	0.243478		

表 4.10 台塑之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	768	46	36	850	0.903529
P	176	9	4	189	0.047619
N	169	9	10	188	0.053191
total	1113	64	50	正確率 64.14	
Precision	0.690027	0.140625	0.2		

➤ 實驗對象 B：台塑

- 樣本數：1227 筆
- 分析方法：隨機森林
- 特徵維度：28

表 4.11 台塑之隨機森林之明確值混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	610	130	110	850	0.717647
P	128	38	23	189	0.201058
N	129	26	33	188	0.175532
total	867	194	166	正確率 55.05	
Precision	0.703576	0.195876	0.198795		

表 4.12 台塑之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	641	113	96	850	0.754118
P	134	30	25	189	0.15873
N	140	20	28	188	0.148936
total	915	163	149	正確率 56.96	
Precision	0.700546	0.184049	0.187919		

表 4.13 台塑之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	612	124	114	850	0.72
P	145	21	23	189	0.111111
N	124	34	30	188	0.159574
total	881	179	167	正確率 54.03	
Precision	0.694665	0.117318	0.179641		

表 4.14 台塑 SVM 以及隨機森林正確率比較

分析工具	方法		
	明確值	均等模糊化	CPDA 模糊化
SVM	50.69	59.49	64.14
隨機森林	55.05	56.96	54.03

➤ 實驗對象 C：台積電

- 樣本數：1227 筆
- 特徵維度：28
- 分析方法：SVM (參數設定：Gamma：5，c：0.8，多項式核心數)

從表 4.15~4.21 可以發現模糊化方法用於分類技術其正確率都較優於明確值如表 4.21，CPDA 模糊化方法其 SVM 正確率為 45.15 優於明確值 34.63，在隨機森林方面，均等模糊化與 CPDA 模糊化分別為 44.9 與 43.84 均高於明確值 43.03。從分析中可得知，在分類技術方便，不論是 SVM 以及隨機森林，均等模糊化以及 CPDA 模糊化都明顯優於明確值，並且 CPDA 更優於均等模糊化方法。

表 4.15 台積電之 SVM 之明確值混淆矩陣(以 10-Folds 為例)

	P	N	S	total	Recall
P	73	96	101	270	0.27037
N	85	82	79	246	0.333333
S	173	268	270	711	0.379747
total	331	446	450	正確率 34.63	
Precision	0.220544	0.183857	0.6		

表 4.16 台積電之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	438	144	130	712	0.615169
P	173	43	54	270	0.159259
N	140	61	44	245	0.179592
total	751	248	228	正確率 42.78	
Precision	0.583222	0.173387	0.192982		

表 4.17 台積電之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	466	130	116	712	0.654494
P	179	35	56	270	0.12963
N	141	51	53	245	0.216327
total	786	216	225	正確率 45.15	
Precision	0.592875	0.162037	0.235556		

➤ 實驗對象 C：台積電

- 樣本數：1227 筆
- 分析方法：隨機森林
- 特徵維度：28

表 4.18 台積電之隨機森林之明確值混淆矩陣(以 10-Folds 為例)

	P	N	S	total	Recall
P	58	56	156	270	0.214815
N	58	46	142	246	0.186992
S	148	139	424	711	0.596343
total	264	241	722	正確率 43.03	
Precision	0.219697	0.190871	0.587258		

表 4.19 台積電之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
S	452	142	118	712	0.634831
P	162	54	54	270	0.2
N	144	56	45	245	0.183673
total	758	252	217	正確率 44.90	
Precision	0.596306	0.214286	0.207373		

表 4.20 台積電之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	S	P	N	total	Recall
s	443	160	109	712	0.622191
p	178	43	49	270	0.159259
n	141	52	52	245	0.212245
total	762	255	210	正確率 43.84	
Precision	0.581365	0.168627	0.247619		

表 4.21 台積電 SVM 以及隨機森林正確率

分析工具	方法		
	明確值	均等模糊化	CPDA 模糊化
SVM	34.63	42.78	45.15
隨機森林	43.03	44.90	43.84

➤ 實驗對象 D：南亞

- 樣本數：1227 筆
- 特徵維度：28
- 分析方法：SVM (參數設定：Gamma：5，c：0.8，多項式核心函數)

從表 4.22~4.28 可以發現可以發現模糊化方法用於分類技術其正確率較優於明確值如表 4.28，CPDA 模糊化方法其 SVM 正確率為 56.31 優於明確值 47.91，在隨機森林方面，均等模糊化與 CPDA 模糊化分別為 53.54 高於明確值 51.42。從分析中可得知，在分類技術方面，在 SVM 中，均等模糊化以及 CPDA 模糊化都明顯優於明確值，而在隨機森林中均等模糊化明顯優於明確。

表 4.22 南亞之 SVM 之明確值混淆矩陣(以 10-Folds 為例)

	S	N	P	total	Recall
S	503	130	156	789	0.637516
N	106	42	52	200	0.21
P	126	57	45	228	0.197368
total	735	229	253	正確率 49.71	
Precision	0.684354	0.183406	0.177866		

表 4.23 南亞之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)

	S	N	P	total	Recall
S	612	88	99	799	0.765957
N	126	31	43	200	0.155
P	143	47	38	228	0.166667
total	881	166	180	正確率 55.50	
Precision	0.694665	0.186747	0.211111		

表 4.24 南亞之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	S	N	P	total	Recall
S	617	81	101	799	0.772215
N	128	32	40	200	0.16
P	144	42	42	228	0.184211
total	889	155	183	正確率 56.31	
Precision	0.694038	0.206452	0.229508		

➤ 實驗對象 D：南亞

- 樣本數：1227 筆
- 分析方法：隨機森林
- 特徵維度：28

表 4.25 南亞之隨機森林之明確值混淆矩陣(以 10-Folds 為例)

	S	N	P	total	Recall
S	534	126	139	799	0.668335
N	113	48	39	200	0.24
P	128	51	49	228	0.214912
total	775	225	227	正確率 51.42	
Precision	0.689032	0.213333	0.215859		

表 4.26 南亞之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)

	S	N	P	total	Recall
S	583	108	108	799	0.729662
N	126	34	40	200	0.17
P	144	44	40	228	0.175439
total	853	186	188	正確率 53.54	
Precision	0.68347	0.182796	0.212766		

表 4.27 南亞之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	S	N	P	total	Recall
S	555	121	123	799	0.694618
N	125	28	47	200	0.14
P	148	39	41	228	0.179825
total	828	188	211	正確率 50.85	
Precision	0.67029	0.148936	0.194313		

表 4.28 南亞 SVM 以及隨機森林正確率比

分析工具	方法		
	明確值	均等模糊化	CPDA 模糊化
SVM	47.91	55.50	56.31
隨機森林	51.42	53.54	50.85

4.2.研究結果分析

■ 模糊化方法在分類正確率上優於明確值

在 SVM 方面，本研究針對上述表 4.1~表 4.28 以及下圖 4.1 四家公司收盤價波動來做探討，並發現在分析中，模糊資料集分類正確率皆高於明確資料集。從圖 4.1 中可看出，台積電 2014~2018 年間收盤價波動幅度從 100~250 之間，波動幅度極大，從研究分析結果發現模糊化分類技術在此現象中也能得到較高之結果。而在台塑、南亞穩定波動以及鴻海小幅度波動中都擁有較高的正確率，由此得知 SVM 結合模糊化都能擁有較高的正確率，並且在 CPDA 模糊化更優於均等模糊化。

在隨機森林方面，表 4.7 實驗對象鴻海中可以發現模糊值皆高於明確值，而 CPDA 模糊化高於均等模糊化。表 4.14 實驗對象台塑發現明確值高於 CPDA 模糊化並且均等模糊化則皆高於其於兩者。表 4.21 實驗對象台積電發現模糊值皆高於明確值，並且均等模糊化高於 CPDA 模糊化。表 4.28 實驗對象南亞發現明確值高於 CPDA 值並且均等模糊化則優於其兩者。綜合上述，隨機森林在 CPDA 於均等模糊化方法正確率優劣上雖有跳動，但差異

不大，而且可以發現模糊化結果還是比明確值較高。

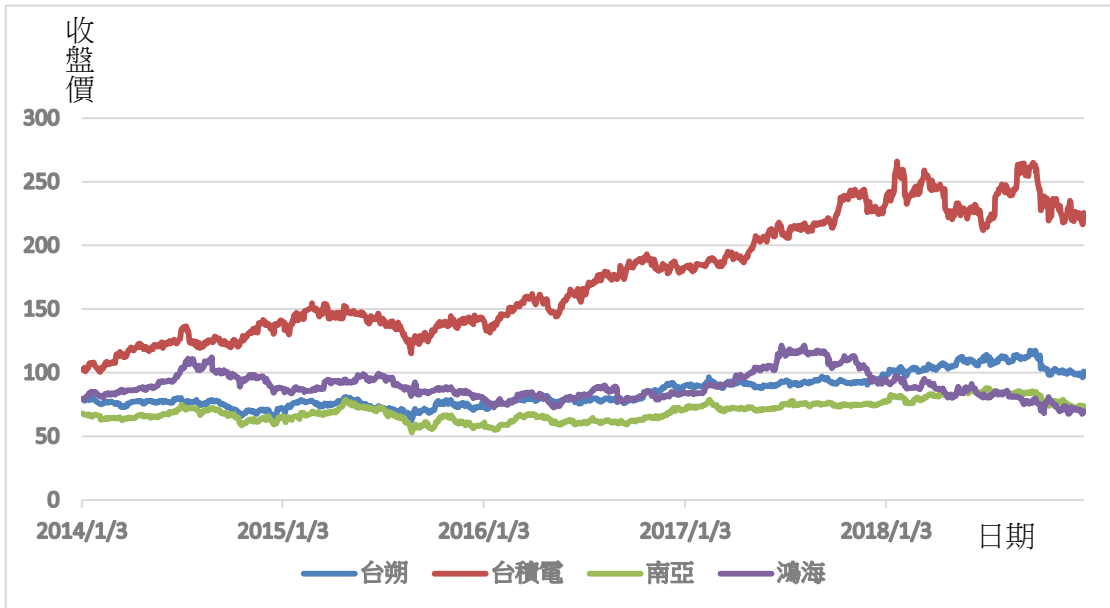


圖 4.1 四家公司收盤價波動之折線圖

■ 混淆矩陣分析

在此部分本研究將混淆矩陣整理成在持平、上漲跟下跌中所預測之 Recall 值與 Precision 值以鴻海為例來進行呈現。在表 4.29~4.34 發現，在持平、上漲以及下跌之中，都擁有不穩定之波動，但差異幅度不大。本研究將其餘三家上漲、下跌與持平的 Recall 值與 Precision 值比較表請參閱附錄二。

表 4.29 鴻海在持平之中的明確值與模糊值之 Recall 比較

	持平	上漲	下跌	total	Recall
持平/SVM 明確值	452	173	151	776	0.582474
持平/SVM 均等模糊化	545	126	105	776	0.70232
持平/SVM CPDA 模糊化	585	102	90	777	0.752896
持平/隨機森林 明確值	491	126	159	776	0.632732
持平/隨機森林 均等模糊化	531	123	122	776	0.684278
持平/隨機森林 CPDA 模糊化	526	133	117	776	0.677835

表 4.30 鴻海在持平之中的明確值與模糊值之 Precision 比較

	持平	上漲	下跌	total	Precision
持平/SVM 明確值	452	118	105	675	0.66963
持平/SVM 均等模糊化	545	155	141	841	0.648038
持平/SVM CPDA 模糊化	585	166	144	895	0.653631
持平/隨機森林 明確值	491	132	124	747	0.657296
持平/隨機森林 均等模糊化	531	159	144	834	0.636691
持平/隨機森林 CPDA 模糊化	526	142	142	810	0.649383

表 4.31 鴻海在下跌之中的明確值與模糊值之 Recall 比較

	持平	上漲	下跌	total	Recall
下跌/ SVM 明確值	105	65	53	223	0.237668
下跌/SVM 均等模糊化	141	44	38	223	0.170404
下跌/SVM CPDA 模糊化	144	37	42	223	0.188341
下跌/隨機森林 明確值	124	60	39	223	0.174888
下跌/隨機森林 均等模糊化	144	41	38	223	0.170404
下跌/隨機森林 CPDA 模糊化	142	38	43	223	0.192825

表 4.32 鴻海在下跌之中的明確值與模糊值之 Precision 比較

	持平	上漲	下跌	total	Precision
下跌/ SVM 明確值	151	50	53	254	0.208661
下跌/SVM 均等模糊化	105	38	38	181	0.209945
下跌/SVM CPDA 模糊化	90	30	42	162	0.259259
下跌/隨機森林 明確值	126	43	39	208	0.1875
下跌/隨機森林 均等模糊化	122	136	38	196	0.193878
下跌/隨機森林 CPDA 模糊化	117	48	43	208	0.206731

表 4.33 鴻海在上漲之中的明確值與模糊值之 Recall 比較

	持平	上漲	下跌	total	Recall
上漲/ SVM 明確值	118	60	50	228	0.263158
上漲/SVM 均等模糊化	155	35	38	228	0.153509
上漲/SVM CPDA 模糊化	161	31	36	228	0.135965
上漲/隨機森林 明確值	132	53	43	228	0.232456
上漲/隨機森林 均等模糊化	159	33	36	228	0.144737
上漲/隨機森林 CPDA 模糊化	142	38	48	228	0.166667

表 4.34 鴻海在上漲之中的明確值與模糊值之 Precision 比較

	持平	上漲	下跌	total	Precision
上漲/ SVM 明確值	173	60	65	298	0.201342
上漲/SVM 均等模糊化	126	35	44	205	0.170732
上漲/SVM CPDA 模糊化	102	31	34	101	0.187135
上漲/隨機森林 明確值	159	53	60	272	0.194853
上漲/隨機森林 均等模糊化	123	33	41	197	0.167513
上漲/隨機森林 CPDA 模糊化	133	38	38	209	0.181818

■ 其他發現與討論

過去文獻(吳嘉川，2008；江旻偉，2014；謝政翰，2019)

探討重點皆以正確率來進行分析，較少以混淆矩陣來做探討。

故本研究以鴻海 2014 年~2018 年 5 年資料為例，來進行 3% 漲跌幅之分析預測。

由表 4.33、表 4.34 之中發現，SVM 利用明確值進行預測時，原始在 S 且被歸類在 S 的值有 1085 筆，而 CPDA 模糊化方法有 1148 筆，在 P 的部分有 1 筆，而 CPDA 模糊化方法有 0 筆，而在 N 的部分有 3 筆，而 CPDA 模糊化方法有 2 筆。正確率方面分別為 88.75 以及 93.72。可以發現其正確率皆在 8 成以上，而 CPDA 模糊化更高達 93.72%，但從表 4.33 以及 4.34 可發現，正確率貢獻大部分都集中在 S 上，此為資料偏斜在分類技術方法上之問題，後續研究應注意此問題。其餘三家在 3% 漲跌幅預測之混淆矩陣請參閱附錄。

表 4.35 鴻海之 SVM 之明確值混淆矩陣(以 10-Folds 為例)

	s	p	n	total	Recall
s	1085	47	40	1172	0.925768
p	25	1	2	28	0.035714
n	22	2	3	27	0.111111
total	1132	50	45	正確率 88.75	
Precision	0.958481	0.02	0.066667		

表 4.36 鴻海之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	s	p	n	total	Recall
s	1148	10	14	1172	0.979522
p	28	0	0	28	0
n	24	1	2	27	0.074074
total	1200	11	16	正確率 93.72	
Precision	0.956667	0	0.125		

上述實驗中，發現在漲跌幅 1% 中利用 SVM 與隨機森林在模糊化方法中皆能有較高的正確率，並比一般明確值擁有明顯較高的提升。而在 3% 漲跌幅中，模糊化方法正確率皆優於明確值，並且正確率達至 9 成，但可發現預測準確率的貢獻大部分都落在持平的筆數上。所以在 SVM、決策樹、隨機森林等分類技術，能發現資料偏斜會影響正確性，此為本研究發現與限制。

第五章結論

模糊理論由許多前人的研究證實，不論是應用任何領域都有著良好的表現。而股票投資分面存在著高報酬以及高風險，所以為了讓投資者能有依據以及降低風險的目的下，過去文獻則會利用股市的歷史資料來分析預測漲跌，並利用開盤價換算各種新指標，我們把這些指標稱技術分析指標，以利於各種股市環境之預測，本研究利用分類技術方法以及模糊化方法來進行股價漲跌之預測。

本研究針對四家上市公司進行實驗，並以隔日漲跌為預測變數，並以 28 個常見技術指標做為研究的特徵變數，將特徵變數進行模糊化得到兩種模糊值(1)均等模糊值；(2)CPDA 模糊值。並以上述兩種模糊值與明確值進行分類技術(SVM、隨機森林)，最後則以 10-Folds、Recall 以及 Precision 來進行結果之驗證。由實證分析得知，在 1%漲跌幅中，SVM 方法較隨機森林穩定，並 CPDA 模糊化則在 SVM 之中擁有較穩定之結果。均等模糊化在隨機森林中，擁有較穩定的正確率，並且皆比明確值擁有較高的提升。

在 4.2 討論中，在 3%預測環境，上漲與下跌筆數都較少，在分類方法技術上，較不容易進行預測，並且將大部分值都歸類筆數較多的類別上，而導致正確率高。因此得知在資料過度偏斜的問題上，後續研究應小心注意。綜合上述，本研究貢獻如下

- 應用均等模糊化以及 CPDA 模糊化方法建構分類模型，進行實例股市分析比較。
- 採用 2014~2018 年實際四家股市資料進行實例分析。
- 利用模糊分類技術與股市技術指標結合 10-Folds 交叉驗證，來探討模糊值與明確值之關聯性。



參考文獻

一. 中文文獻

1. 王威能(2015)，依據技術指標投資臺灣 50 ETF 及其成分股，國立臺北大學經濟學研究所碩士論文。
2. 王培宇(2012)，DIF_KD 技術指標應用於台灣股票市場之實證研究，東海大學管理學院財務金融研究所碩士在職專班論文。
3. 江旻緯(2014)，以改良式支持向量機建立股票股價漲跌預測模型，高雄應用科技大學資訊管理系碩士班碩士論文。
4. 吳嘉川(2008)，應用資料採礦技術於股市加權指數預測之研究，輔仁大學資訊管理系碩士論文。
5. 吳睿萍(2018)，技術分析在商品期貨市場獲利能力-以 MACD、KD、SAR 為例，國立成功大學碩士論文。
6. 宋念遠(2014)，影響乳癌放射線治療後引起放射性肺炎併發症預測因子之研究，國立高雄應用科技大學電子工程系碩士班碩士論文。
7. 李良俊(2003)，臺灣股票市場技術分析有效性之研究，實踐大學企業管理研究所碩士論文。
8. 李逢嘉(2010)，特徵選取為基礎之複合分類預測模式以信用資料為例，國立清華大學工業工程與工程管理研究所博士論文。
9. 周琬禎(2018)，結合模糊理論與 Apriori 演算法於數值型資料關聯規則探勘之研究，國立雲林科技大學工業工程與管理系工業工程組碩士論文。
10. 周照偉，鄭榮祿，蔡賢亮，楊崇宏，牟聖遠(2015)，臺灣股市技術分析實證：以隨機指標，相對強弱指標，指數平滑異同平均線指標及趨向指標為例，高雄應用科技大學人文與社會科學學刊。
11. 周聖鈞(2012)，趨勢策略應用於台灣股票市場之研究國立臺北科技大學工商管理研究所博士論文。
12. 林宇恆(2015)，決策樹結合複迴歸模型預測氣溫與雨量，國立臺灣師範大學電機工程學系碩士論文。
13. 林俐玲，王兆文，沈哲緯，陳品岡，翁志成(2011)，應用支持向量機探討山坡地土壤沖蝕程度之研究，水土保持學報，43 卷，1 期，頁 49-68，3 月。
14. 林素娥(2016)，技術分析交易策略實證研究 -以台灣 50 ETF 為例，國立高雄應用科技大學金融資訊碩士論文。
15. 林智仁，取自 <http://www.csie.ntu.edu.tw/~cjlin/>。
16. 凌士雄(2004)，非對稱性分類分析解決策略之效能比較，國立中山大學資訊管理學系碩士論文。
17. 徐偉城(2016)，運用威廉指標於台指期貨之當沖研究，開南大學財務金融學系碩士論文。

18. 高敬恆(2015)，技術指標在台指期貨市場之應用，國立臺北大學經濟學研究所，碩士論文。
19. 謝睿帆(2008)，CPDA 化模糊規則評估學生成績等等，國立雲林科技大學資訊管理系碩士班碩士論文。
20. 陳玟志(2012)，RSI 與威廉指標於台灣加權指數操作績效實證，輔仁大學金融與國際企業學金融碩士班碩士論文。
21. 陳振東、謝政翰(2019)，應用機器學習與模糊推論於股價漲跌預測之研究，資訊管理學報。
22. 陳逸平，李俊德. (2008)，結合類神經網路及隨機與威廉指標應用於股市預測之研究。
23. 陳應慶(2003)，應用技術分析指標於台灣股票市場加權指數買進時機切入之實證研究—以 RSI、MACD 及 DIF 為技術指標，佛光人文社會學院管理學研究所碩士論文。
24. 陳耀茂(1999)，模糊理論，五圖書出版有限公司。
25. 曾宇皓(2018)，結合基本分析與技術分析程序化交易之實證研究—以臺灣股票市場為例，國立中央大學財務金融學系碩士論文。
26. 游士億(2008)，以 FCM 分群、指標模糊化與 SVR 探討多國股市的動態，中華大學資訊管理學碩士論文。
27. 賀安平(2016)，從新聞文章預測股票走：使用 SVM 與 LDA 演算法，國立高雄應用科技大學資訊管理系碩士班碩士論文。
28. 黃承龍，陳穆臻，王界人(2004)，支援向量機於信用評等之應用，計量管理期刊，1 卷，2 期，頁 155~172，12 月。
29. 楊維珊(2009)，入侵者偵測之半監督式人臉辨識系統，國立成功大學碩士論文。
30. 葉子維(2018)，顧客消費行為分析及行動銀行使用預測 - 決策樹、隨機森林與判別分析之比較，國立臺北大學統計學系(所)碩士論文。
31. 葉孝賢(2018)，台灣股票市場波動率之研究，靜宜大學財務工程學系碩士論文。
32. 葉鎮源 (2012)，見解探勘。
33. 盧廷宇(2011)，利用倒傳遞類神經網路結合保力加通道支援外匯交易決策，大同大學資訊經營管理學系碩士論文。
34. 戴熠星(2017)，台灣 50 指數成份股的高頻資料系統風險係數 beta 之研究，輔仁大學金融與國際企業學系金融碩士班碩士論文。
35. 謝和昆(1999)，疾病分類支援系統之建構-以神經外科與放射科為例，東海大學工業工程與經營資訊研究所碩士論文。
36. 蕭政倫(2012)，技術分析之實證研究-以台灣加權股市指數為例，國立高雄第一科技大學金融研究所碩士論文。

37. 龔健生 (2016)，分類技術於類別不平衡資料集之研究，國立中央大學資訊管理學系碩士論文。



二. 英文文獻

- Agung B. Prasetijo & Takdir A. Saputro & Ike P. Windasari & Yudi E. Windarto (2017). Buy/sell signal detection in stock trading with bollinger bands and parabolic SAR: With web application for proofing trading strategy. 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE) .
- Allahviranloo, M., & Recker, W. (2013). Daily Activity Pattern Recognition By Using Support Vector Machines With Multiple Classes. *Transportation Research Part B: Methodological*, 58, 16-43.
- Appel, G. & Hitschler, W. F. (1979) .Stock Market Trading System. South Carolina: Traders Press.
- Breiman, L. Random forests(2001). *Random forests .Machine Learning*, 45, 5-32
- Breiman, L. (1996). Bagging predictors, *Machine Learning*, 24, 123-140.
- Burges, C. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*,2, 121-167.
- Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273-297.
- Chen, S. M. (1996). Forecasting enrollments based on fuzzy time series.. *Fuzzy sets and systems*, 3, 311-319.
- Cortes, C. & Vapnik, V. (1995). *Support-Vector Networks*. Kluwer Academic Publishers.
- Fernandez-Gavilance, M. & Alvarez-Lopez, T. & Juncal-Martinez, J. & Costa-Montenegro, E. & Gonzalez-Castano, F. J.,(2016). Unsupervised method for sentiment analysis in online texts.*Expert System With Application*,58,57-75.
- George, L. (1957).Research Stochastic lineis also one of the very popular technical indicators.
- Gerlein, E. A. & McGinnity, M. & Belatreche, A., & Coleman, S. (2016). Evaluating machine learning classification for financial trading: An empirical approach. *Expert Systems with Applications*, 54, 193-207.
- Quinlan, J. R. (1993). *Programs for Machine Learning*. Morgan kaufmann.
- Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1, 81-106.
- Su, C. H. & Chen, T. L. & Cheng, C. H. & Chen , Y. C. (2010). Forecasting the Stock Market with Linguistic Rules Generated from the Minimize Entropy Principle and the Cumulative Probability Distribution Approaches. *Entropy* 2010, 10, 2397-2417.
- Weia, Y. C. & Lub, Y. C. & Chenb , J. N. & Hsuc, Y. J. (2017). Informativeness of the market news sentiment in the Taiwan stock market. *North American Journal of*

- Economics and Finance ,39, 158–181.
- Wu, G. G. R. & Hou, T. C. T. & Lin, J. L. (2018).Can economic news predict Taiwan stock market returns? *Asia Pacific Management Review* ,24 ,54-59.
- Ravi, K. & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14-46.
- Williams, L. (1973). *How I made a Million Dollars*. USA: Investor, FLIX.
- Xu, C. J. & A. van der Schaaf A. & Van't Veld A. A. & Langendijk J. A. & Schilstra C. (2012).Statistical validation of normal tissue complication probability models. *International Journal of Radiation Oncology* Biology* Physics*,1,e123-e129.
- Zadeh, L. A. (1965).Fuzzy Set. *information and Control*,8, 333-353.



附 錄 一

在1%漲跌幅時，持平、上漲以及下值之 Recall、Precision 比較。

表 6.1 台塑在持平之中的明確值與模糊值之 Recall 比較

	持平	上漲	下跌	total	Recall
持平/SVM 明確值	551	141	158	850	0.648235
持平/SVM 均等模糊化	675	101	74	850	0.794118
持平/SVM CPDA 模糊化	768	46	36	850	0.903529
持平/隨機森 林明確值	610	130	110	850	0.717647
持平/隨機森 林 均等模糊化	641	113	96	850	0.754118
持平/隨機森 林 CPDA 模 糊化	612	124	114	850	0.72

表 6.2 台塑在持平之中的明確值與模糊值之 Precision 比較

	持平	上漲	下跌	total	Precision
持平/SVM 明確值	551	129	109	789	0.698352
持平/SVM 均等模糊化	675	149	147	971	0.69516
持平/SVM CPDA 模糊化	768	176	169	1113	0.690027
持平/隨機森林 明確值	610	128	129	867	0.703576
持平/隨機森林 均等模糊化	641	134	140	915	0.700546
持平/隨機森林 CPDA 模糊化	612	145	124	881	0.694665

表 6.3 台塑在上漲之中的明確值與模糊值之 Recall 比較

	持平	上漲	下跌	total	Recall
上漲/ SVM 明確值	129	33	27	189	0.174603
上漲/SVM 均等模糊化	149	27	13	189	0.142857
上漲/SVM CPDA 模糊 化	176	9	4	189	0.047619
上漲/隨機森 林明確值	128	38	23	189	0.201058
上漲/隨機森 林 均等模糊化	134	30	25	189	0.15873
上漲/隨機森 林 CPDA 模 糊化	145	21	23	189	0.111111

表 6.4 台塑在上漲之中的明確值與模糊值之 Precision 比較

	持平	上漲	下跌	total	Precision
上漲/ SVM 明確值	141	33	41	215	0.153488
上漲/SVM 均等模糊化	101	27	13	141	0.191489
上漲/SVM CPDA 模糊 化	46	9	9	64	0.140625
上漲/隨機森 林明確值	130	38	26	194	0.195876
上漲/隨機森 林 均等模糊化	113	30	20	163	0.184049
上漲/隨機森 林 CPDA 模 糊化	124	21	34	179	0.117318

表 6.5 台塑在下跌之中的明確值與模糊值之 Recall 比較

	持平	上漲	下跌	total	Recall
下跌/ SVM 明確值	109	41	38	188	0.202128
下跌/SVM 均等模糊化	147	13	28	188	0.148936
下跌/SVM CPDA 模糊 化	169	9	10	188	0.053191
下跌/隨機森 林明確值	129	26	33	188	0.175532
下跌/隨機森 林 均等模 糊化	140	20	28	188	0.148936
下跌/隨機森 林 CPDA 模 糊化	124	34	30	188	0.159574

表 6.6 台塑在下跌之中的明確值與模糊值之 Precision 比較

	持平	上漲	下跌	total	Precision
下跌/ SVM 明確值	158	27	38	223	0.170404
下跌/SVM 均等模糊化	74	13	28	115	0.243478
下跌/SVM CPDA 模糊 化	36	4	10	50	0.2
下跌/隨機森 林明確值	110	23	33	116	0.198795
下跌/隨機森 林 均等模 糊化	96	25	28	149	0.187919
下跌/隨機森 林 CPDA 模 糊化	114	23	30	167	0.179641

表 6.7 南亞在下跌之中的明確值與模糊值之 Recall 比較

	持平	下跌	上漲	total	Recall
持平/SVM 明確值	503	130	156	789	0.637516
持平/SVM 均等模糊化	612	88	99	799	0.765957
持平/SVM CPDA 模糊 化	617	81	101	799	0.772215
持平/隨機森 林明確值	534	126	139	799	0.668335
持平/隨機森 林 均等模 糊化	583	108	108	799	0.729662
持平/隨機森 林 CPDA 模 糊化	555	121	123	799	0.694618

表 6.8 南亞在上漲之中的明確值與模糊值之 Precision 比較

	持平	下跌	上漲	total	Precision
持平/SVM 明確值	503	106	126	735	0.68435
持平/SVM 均等模糊化	612	126	143	881	0.694665
持平/SVM CPDA 模糊 化	617	128	144	889	0.694038
持平/隨機森 林明確值	534	113	128	775	0.689032
持平/隨機森 林均等模 糊化	583	126	144	853	0.68347
持平/隨機森 林 CPDA 模 糊化	555	125	148	828	0.67029

表 6.9 南亞在上漲之中的明確值與模糊值之 Recall 比較

	持平	下跌	上漲	total	Recall
上漲/ SVM 明確值	126	57	45	228	0.197368
上漲/SVM 均等模糊化	143	47	38	228	0.166667
上漲/SVM CPDA 模糊 化	144	42	42	228	0.184211
上漲/隨機森 林明確值	128	51	49	228	0.214912
上漲/隨機森 林 均等模 糊化	144	44	40	228	0.175439
上漲/隨機森 林 CPDA 模 糊化	148	39	41	228	0.179825

表 6.10 南亞在上漲之中的明確值與模糊值之 Precision 比較

	持平	下跌	上漲	total	Precision
上漲/ SVM 明確值	156	52	45	253	0.177866
上漲/SVM 均等模糊化	99	43	38	180	0.211111
上漲/SVM CPDA 模糊 化	101	40	42	183	0.229508
上漲/隨機森 林明確值	139	39	49	227	0.215859
上漲/隨機森 林 均等模 糊化	108	40	40	188	0.212766
上漲/隨機森 林 CPDA 模 糊化	123	47	41	211	0.19431

表 6.11 南亞在下跌之中的明確值與模糊值之 Recall 比較

	持平	下跌	上漲	total	Recall
下跌/ SVM 明確值	126	57	45	228	0.197368
下跌/SVM 均等模糊化	143	47	38	228	0.166667
下跌/SVM CPDA 模糊 化	144	42	42	228	0.184211
下跌/隨機森 林明確值	128	51	49	228	0.214912
下跌/ 隨機森林 均等模糊化	144	44	40	228	0.175439
下跌/ 隨機森林 CPDA 模糊 化	148	39	41	228	0.179825

表 6.12 南亞在上漲之中的明確值與模糊值之 Precision 比較

	持平	下跌	上漲	total	Precision
下跌/ SVM 明確值	156	52	45	253	0.177866
下跌/SVM 均等模糊化	99	43	38	180	0.211111
下跌/SVM CPDA 模糊 化	101	40	42	183	0.229508
下跌/隨機森 林明確值	139	39	49	227	0.215859
下跌/隨機森 林 均等模 糊化	108	40	40	188	0.212766
下跌/隨機森 林 CPDA 模 糊化	123	47	41	211	0.19431

附 錄 二

以下皆為3%漲跌幅時之混淆矩陣：

表 7.1 南亞之 SVM 之明確值混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1133	36	21	1190	0.952101
n	18	2	1	21	0.095238
p	15	1	0	16	0
total	1166	39	22	正確率 92.5	
Precision	0.971698	0.051282	0		

表 7.2 南亞之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1175	6	9	1190	0.987395
n	20	0	1	21	0
p	15	1	0	16	0
total	1210	7	10	正確率 95.76	
Precision	0.971074	0	0		

表 7.3 南亞之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1174	12	4	1190	0.986555
n	20	1	0	21	0.047619
p	16	0	0	16	0
total	1210	13	4	正確率 95.76	
Precision	0.970248	0.076923	0		

表 7.4 南亞之隨機森林之明確值混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1157	20	13	1190	0.972269
n	18	1	2	21	0.047619
p	15	1	0	16	0
total	1190	22	15	正確率 94.37	
Precision	0.972269	0.045455	0		

表 7.5 南亞之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1166	15	9	1190	0.979832
n	21	0	0	21	0
p	15	1	0	16	0
total	1202	16	9	正確率 95.02	
Precision	0.97005	0	0		

表 7.6 南亞之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1153	20	17	1190	0.968908
n	21	0	0	21	0
p	16	0	0	16	0
total	1190	20	17	正確率 93.96	
Precision	0.968908	0	0		

表 7.7 台積電之 SVM 之明確值混淆矩陣(以 10-Folds 為例)

	S	n	p	total	Recall
s	1073	42	58	1173	0.914749
n	21	1	4	26	0.038462
p	25	3		28	0
total	1119	46	62	正確率 87.53	
Precision	0.958892	0.021739	0		

表 7.8 台積電之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1149	13	11	1173	0.97954
n	23	1	2	26	0.038462
p	24	4	0	28	0
total	1196	18	13	正確率 93.72	
Precision	0.960702	0.055556	0		

表 7.9 台積電之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1154	11	5	1170	0.986325
n	24	0	2	26	0
p	27	1	0	28	0
total	1205	12	7	正確率 94.29	
Precision	0.957676	0	0		

表 7.10 台積電之隨機森林之明確值混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1111	26	36	1173	0.947144
n	22	1	3	26	0.038462
p	24	4	0	28	0
total	1157	31	39	正確率 90.62	
Precision	0.960242	0.032258	0		

表 7.11 台積電之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1136	16	21	1173	0.968457
n	25	0	1	26	0
p	27	1	0	28	0
total	1188	17	22	正確率 92.58	
Precision	0.956229	0	0		

表 7.12 台積電之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1138	14	21	1173	0.970162
n	24	0	2	26	0
p	26	2	0	28	0
total	1188	16	23	正確率 92.74	
Precision	0.957912	0	0		

表 7.13 台塑之 SVM 之明確值混淆矩陣(以 10-Folds 為例)

	S	n	p	total	Recall
s	1132	24	26	1182	0.957699
n	15	0	1	16	0
p	16	2	1	19	0.052632
total	1163	26	28	正確率 92.33	
Precision	0.973345	0	0.035714		

表 7.14 台塑之 SVM 之均等模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1172	11	9	1192	0.983221
n	15	0	1	16	0
p	17	1	1	19	0.052632
total	1204	12	11	正確率 95.59	
Precision	0.973422	0	0.090909		

表 7.15 台塑之 SVM 之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1182	8	2	1192	0.991611
n	16	0	0	16	0
p	18	0	1	19	0.052632
total	1216	8	3	正確率 94.61	
Precision	0.972039	0	0.333333		

表 7.16 台塑之隨機森林之明確值混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1162	15	15	1192	0.974832
n	12	1	3	16	0.0625
p	19	0	0	19	0
total	1193	16	18	正確率 94.78	
Precision	0.974015	0.0625	0		

表 7.17 台塑之隨機森林之均等模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1169	15	8	1192	0.980705
n	16	0	0	16	0
p	19	0	0	19	0
total	1204	15	8	正確率 95.27	
Precision	0.97093	0	0		

表 7.18 台塑之隨機森林之 CPDA 模糊化混淆矩陣(以 10-Folds 為例)

	s	n	p	total	Recall
s	1160	16	16	1192	0.973154
n	16	0	0	16	0
p	17	0	2	19	0.105263
total	1193	16	18	正確率 94.70	
Precision	0.972339	0	0.111111		