



國立中山大學財務管理學系

碩士論文

Department of Finance

National Sun Yat-sen University

Master's Thesis

使用價量資訊及券商分點買賣日報表

來預測台股交易訊號：機器學習的應用

Forecasting Trading Signals for Taiwanese Stocks using  
Price-Volume Information and Brokerage Branch Buy-Sell

Daily Reports: Application of Machine Learning

研究生：劉承彥

Cheng-Yen Liu

指導教授：翁培師

Dr. Pei-Shih Weng

中華民國 112 年 6 月

June 2023

# 論文審定書

## 國立中山大學研究生學位論文審定書

本校財務管理學系碩士在職專班

研究生劉承彥（學號：N104030001）所提論文

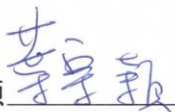
使用價量資訊及券商分點買賣日報表來預測台股交易訊號：機器學習的應用

Forecasting Trading Signals for Taiwanese Stocks using Price-Volume Information and Brokerage Branch Buy-Sell Daily Reports: Application of Machine Learning

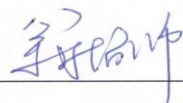
於中華民國 112 年 3 月 25 日經本委員會審查並舉行口試，符合碩士學位論文標準。

學位考試委員簽章：

召集人 葉宗穎



委員 翁培師



委員 柯冠成



委員 蔡明宏



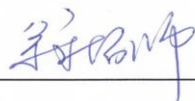
委員

\_\_\_\_\_

委員

\_\_\_\_\_

指導教授(翁培師)



(簽名)

## 摘要

本研究旨在探討使用價量資訊以及券商分點買賣日報表，應用於預測台股交易訊號的可行性。研究背景為 2013 年至 2020 年的台灣股票市場，挑選了 2013 年初每個產業別市值最高的標的來進行台灣股市整體市場的分析，檢視價量以及券商分點日報表的資料是否能有效率的對於全體台灣股票市場進行未來的交易訊號預測。本研究採用 lightGBM 機器學習演算法的方法，對台灣股票市場中的價格、成交量和券商分點買賣日報表等資訊，進行交易訊號的訓練和測試。其中，本研究的交易訊號與常見的預測股價漲跌不同，本研究預測的交易訊號是買進後的出場訊號（到期出場、停損、停利），透過這樣的方式，改善以往研究框架對於潛在獲利、風險的不確定性。從研究結果發現，不同產業屬性的股票標的，對於價格面、籌碼面的資料特徵有不同的反映。本研究進一步比較了不同特徵和參數對模型的影響，結果指出，券商分點日報表所衍生出的資料特徵與價量的資料特徵一同進行本模型的訓練預測時，券商分點日報表的特徵影響模型的結果重要性不低於價量特徵，並對模型的表現進行了詳細的分析和討論。根據結果指出，本模型最有影響力的特徵普遍為過去一段期間的價格「標準差」，同時也代表過去一段時間的價格震盪程度，對於未來交易訊號的預測判斷有巨大的影響力，另外，而主力買賣超集中度也是對於該模型有一定影響力。

關鍵字：lightGBM 機器學習演算法、券商分點買賣日報表、預測台股交易訊號、潛在獲利風險模型、技術面

# Abstract

This study aims to explore the feasibility of using price and volume information, as well as daily reports of broker branch, to predict trading signals in the Taiwan stock market. The research focuses on the period between 2013 and 2020, selecting the highest market capitalization stocks in each industry at the beginning of 2013 to analyze the overall Taiwan stock market. The study examines whether data on price, volume, and daily reports of broker branch can efficiently predict future trading signals for the entire Taiwan stock market. The study uses the lightGBM machine learning algorithm to train and test trading signals based on information such as price, volume, and daily reports of broker branch in the Taiwan stock market. Unlike typical predictions of stock price movements, the trading signals predicted in this study refer to exit signals after buying stocks (expiration, stop loss, and stop profit). This approach improves the uncertainty in previous research frameworks regarding potential profits and risks. The research findings indicate that stocks with different industry attributes exhibit different reactions to data features related to price. The study compares the effects of different features and parameters on the model. It indicates that data features derived from daily reports of broker branches are as important as price and volume data features when combined in model training and prediction. The study provides a detailed analysis and discussion of the model's performance based on these features. According to the results, the most influential feature in the model is typically the standard deviation of prices over a certain period, which represents the volatility of stock during that time and has a significant impact on the prediction of future trading signals. Additionally, the concentration of net buying and selling by major broker branch investors also has a certain impact on the model.

Keywords : lightGBM machine learning algorithm 、 daily reports of broker branch in the Taiwan stock market 、 trading signals 、 potential profits and risks 、 price and volume

# 目錄

論文審定書	i
摘要	ii
Abstract	iii
目錄	iv
圖次	vi
表次	vii
緒論	1
1-1. 研究背景簡介 .....	1
1-2. 研究動機與目的 .....	1
1-3. 研究架構 .....	2
1. 文獻探討	3
2-1. 投資人類別與交易分析 .....	3
2-2. 機器學習與股市預測 .....	3
2. 研究資料與方法	5
3-1. 研究資料 .....	5
3-2. 研究模型介紹 .....	5
3-2-1. 梯度提升機 (Gradient Boosting Machine) .....	5
3-2-2. 輕量梯度提升模型 (Light Gradient Boosting) .....	5
3-3. 機器學習標籤制定規則 (LABELS) .....	6
3-4. 機器學習所用到的特徵 (FEATURES) .....	6
3-4-1. 收盤價 .....	6
3-4-2. 成交值 .....	6
3-4-3. 累計成交量 .....	7
3-4-4. 主力券商淨買賣超 .....	7
3-4-5. 主力買賣超集中度 .....	7
3-4-6. 買賣家數差 .....	7
3-4-7. 過去一段時間的累積報酬率 (動能) .....	8
3-4-8. 過去一段時間的最大回檔 .....	8
3-4-9. 移動平均線 .....	8
3-4-10. 股價標準差 .....	9
3-4-11. 股價當前位階 .....	9
3-4-12. 價格突破策略訊號 .....	9
3-5. 研究方法設計 .....	10
3-5-1. 樣本內樣本外測試 .....	10
3-6. 機器學習績效評估 .....	10
3-6-1. 混淆矩陣 .....	10

3.	研究資料與方法	12
4-1.	模型參數處理 .....	12
4-1-1.	資料預處理.....	12
4-1-2.	特徵值敘述統計.....	12
4-1-3.	Label 值敘述統計.....	13
4-1-4.	特徵資料標準化.....	13
4-1-5.	特徵相關性檢定.....	14
4-2.	模型預測結果分析 .....	14
4-2-1.	Feature 重要性 .....	14
4-2-3.	混沌矩陣.....	15
4.	結論	18
5.	未來研究建議	18
6.	參考文獻	19
	中文部分.....	19
	英文部分.....	19

## 圖次

圖 1.	預測 LABEL 差異圖.....	32
圖 2.	交易訊號與收盤價共同繪製圖.....	33
圖 3.	累計成交量與收盤價共同繪製圖.....	34
圖 4.	主力券商淨買超與收盤價共同繪製圖.....	35
圖 5.	主力買賣超集中度與收盤價共同繪製圖.....	36
圖 6.	買賣家數差與收盤價共同繪製圖.....	37
圖 7.	累積報酬率與收盤價共同繪製圖.....	38
圖 8.	最大回檔與收盤價共同繪製圖.....	39
圖 9.	移動平均線與收盤價共同繪製圖.....	40
圖 10.	股價標準差與收盤價共同繪製圖.....	41
圖 11.	股價當前位階與收盤價共同繪製圖.....	42
圖 12.	價格突破策略訊號與收盤價共同繪製圖.....	43
圖 13.	模型預測與實際的結果繪製.....	44
圖 14.	台積電特徵相關性.....	45
圖 15.	台積電特徵重要性.....	46
圖 16.	重要特徵前 10 名出現排行.....	47
圖 17.	台積電混沌矩陣圖.....	48

## 表次

表 1－研究標的清單 .....	22
表 2－各產業特徵敘述統計 1 .....	23
表 3－各產業特徵敘述統計 2 .....	24
表 4－各產業特徵敘述統計 3 .....	25
表 5－各產業特徵敘述統計 4 .....	26
表 6－各產業特徵敘述統計 5 .....	27
表 7－各產業類別樣本次數 .....	28
表 8－台積電混沌矩陣評價指標 .....	29
表 9－各產業混沌矩陣評價指標 .....	30
表 10－全樣本混沌矩陣敘述統計 .....	31



# 緒論

## 1-1. 研究背景簡介

在這個資訊傳遞快速的年代，每天都有各式各樣的多空因子在市場流動，影響著投資人情緒，且當前數位科技發展迅速，更方便的下單環境、更深、更廣的資訊揭示程度、更低的交易手續費，這些原因已經大幅影響當前國內股票投資人，導致投資人交易頻率提升。

筆者觀察到近幾年來台灣程式交易、投資量化分析迅速崛起，多數的投資人有更多可以方便觀察到的數據，包含「技術面」、「基本面」、「籌碼面」，由於影響市場的因子零瑯滿目，導致一般投資人更容易受到不同種類資訊的刺激，沒有穩固的交易計畫，因此激勵筆者以「主力」券商分點追蹤分析的方法來制定交易策略。

並且由於近年來由於機器學習在財務金融領域已經成為一項新趨勢，優點是針對股價分析擺脫了以往的分析框架，並且不事先假設具體的金融分析模型，然後透過機器學習找出各種資料之間的關聯。

## 1-2. 研究動機與目的

投資人進入股市交易，目的就是賺取超額利潤，而一般散戶與專業投資者的差異就在於，是否有確切的進出場依據，其進出場依據可以大致上分為「技術面」、「基本面」、「籌碼面」。台灣投資人結構可以分為三大法人以及其他投資戶，其中三大法人中又以外資的資金量居多，買賣動向會影響整體大盤市場漲跌、個股漲跌，但除了三大法人以外，其他投資戶的市場總成交量將近七成，更是台灣股市活絡的來源。

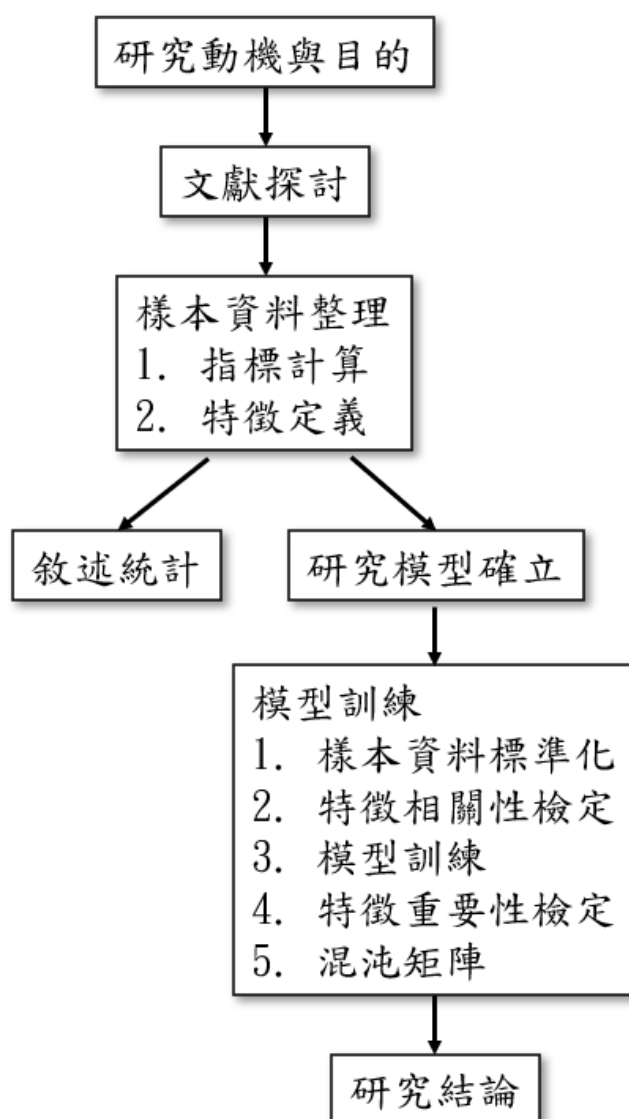
不過三大法人的盤後公開資訊，並沒有辦法得知「個別法人」的進出動向，提供給資訊分析者更大的優勢，所以本研究會將投資人區分為「主力」以及「散戶」。特定的交易人族群中，其大量買進賣出的行為影響著特定股票標的的價格走勢，我們稱為市場「主力」，主力於個股進出的動作，其中成交張數、成交價位，可以透過各個券商分點每日在每檔股票的交易統資料中查詢得知。證交所在每天的收盤後都會揭示每檔個股每日的券商分點買賣資訊。

我們可以透過券商分點，分析「特定期間」某檔各股之累計買入、累計賣出最多的券商，稱之為「主力券商」。透過主力券商的延伸計算，得知特

定個股每天市場主力券商動向，即可追蹤分析該主力券商對各股的影響。

過去文獻中，多探討機構法人交易對於股市的影響，鮮少提到主力投資者的交易行為對股市報酬率的影響，本研究會透過機器學習中的監督式學習來進行預測，透過每檔不同的股票，針對不同面向的資訊來組合進行分析，將多種市場因素當作特徵，預測個股的「交易訊號」，而非個股未來的漲跌，預測股市未來的報酬並沒有辦法非常明確的衡量真正要投資的「潛在風險」以及「潛在報酬」，因此筆者會透過交易策略的結果（停損、停利、固定持有天數出場）來進行預測，較能符合投資人真正交易的情況。

### 1-3. 研究架構



# 1. 文獻探討

因券商分點日報表資料屬於台灣特有的資料類型，以及分點資料不易分析的特性，有關市場主力研究的論文探討較少，有關分點日報表資料類型的論文，以下將過去研究文獻分為「投資人類別與交易分析」以及「機器學習與股市預測」這兩部份分別進行探討。

## 2-1. 投資人類別與交易分析

台灣股市投資人類別可以分為三大法人與散戶，其中散戶該投資人類別佔全市場之成交量其中的七成。

戴欣頻(2009) 分點券商熱門度與股價報酬關係之實證研究，發現到台灣股市在短時間的交易過程中，個股「前一日分點券商熱門度」越高，代表關注該個股的投資人多已參與其中，對短期內股價走勢有利。

廖子瑄(2014) 股市主力交易行為與委託簿資訊內涵之互動，其中的累積異常報酬實證結論當中，發現到主力券商在特定群體(市值中)股票具有資訊優勢，其中又以國內券商總公司的主力最有資訊價值。

戴欣頻(2019)提到券商分點熱門度增加時，對於當日股價會產生顯著的影響，此外也提到各類型的台灣股票籌碼資料對於蘋果概念股的影響程度，其中包含融資融券、當沖幅度、三大法人等籌碼資料。

根據 Lee, Lin, and Liu (1999)依照投資者的成交量，將投資人類別區分為「機構投資者」、「大型散戶」、「小型散戶」，並運用 VAR 模型來檢視哪種類型的投資者是最具有資訊的投資者，其中發現到大型散戶是最具有資訊的投資人，其次為機構投資人，最後為小型散戶。

根據 Bozcuk and Lasfer (2005)以倫敦交易所之法人交易資料中發現，基金經理人具有大量的資訊內涵，並且得知投資人類別、成交量、庫存部位是對價格影響的重要因素。

根據 Boehmer and Kelley (2009)研究紐約證交所的上市股票，發現機構投資人持股量對於價格效率有明顯的相關性。

## 2-2. 機器學習與股市預測

本研究製作當下已經有許多利用機器學習、深度學習來從事金融相關的預測了。

根據 Ke et al. (2017)發表的文章顯示 LightGBM 相較於 XGBoost、pGBRT 是一個更有效率的梯度提升的機器學習模型，LightGBM 相較於傳統的 GBDT 提高了將近 20 倍的訓練速度，並提供相同的精準度。

根據 Dash and Dash (2016)透過類神經網路模型(CEFLNN)去預測股票未來走勢，模型的輸入特徵是透過技術分析、趨勢追蹤等指標，輸出類別是將結果分為買入、持有、賣出三種類別，結果發現相較於其他傳統的模型預測具有更高的利潤。

根據 Li, Xu, and Li (2002)研究中國股市，使用隨機森林、XGBoost、LightGBM 分析選股模型，並證實 LightGBM 的效果是最好的。

根據 Gunduz(2021)研究八大銀行股每 8 小時的價格走勢，根據 SVM、深度學習 LSTM、機器學習 LightGBM 來預測，並且透過不同方法選擇(所有特徵、VAE 減少特徵等)出來的特徵來進行預測並得到不同的預測結果。

Kimoto et al. (1990)以多個類神經學習演算法來預測東京證交所價格指數，在回測中獲得更豐厚的報酬率。

Lee et al. (2007)研究如何透過機器學習中的強化學習框架，來建構真實市場的交易系統，提供買賣建議給投資者。Chang et al. (2009)研究提出建構學習的交易系統的方法，根據 PLP、ANN 模型，建構一個動態閾值決策交易系統。Teixeira et al. (2010)研究透過 PLP、ANN 模型的檢測，來提供歷史資料內非線性的關係，來捕捉隱藏在歷史資料的交易秘密，並將系統用於實際交易中。Rodríguez-Gonz et al. (2011)研究透過基本面與技術面如何改善投資的技術，透過 ABC、KNN 模型去實現。Kara et al. (2011)研究透過適應預期模型與其他傳統模型相比，並且用於改善投資建議系統。Yang et al. (2021)的研究中提到，XGBoost 與 LightGBM 兩個模型針對股價價格預測的方法中最有效率的方式是透過權重 1:1 結合的方式，會得到比單一模型預測更好的結果。

根據林尚毅(2022)使用羅吉斯回歸分類的機器學習模型，研究指出將技術分析策略結合至動能策略會對於策略模型當中的風險承受程度有相當程度的幫助。

劉馨文、王正豪 (2020)，使用長短期記憶(LSTM)來進行股票趨勢預測，並指出相較於僅單獨使用新聞或是股價資訊的模型，加入時間序列的新聞文章訊息使模型更能精準的預測股價市場趨勢。張力元 (2018) 透過神經網路的簡單序列模型來預測股價，模型預測上漲機率較高的股票，並且進行回測，得到相較於大盤更佳的回測績效。

## 2. 研究資料與方法

### 3-1. 研究資料

本研究的目的是分析券商分點日報表與交易訊號的預測，資料頻率為日頻率，因此使用到的分為以下兩個部分：第一種為「還原股價資料」，用來進行技術面的資料特徵建構(包含價格指標計算)，還原價格資料可以解決除權息以後的回測價差失真；第二種資料為「每日券商分點資料」，用來計算主力相關指標。樣本期間為：2013/01/01 至 2020/12/31 年。

台灣證交所截至 2013 年初，上市櫃共有 1636 檔股票掛牌，但因實證解釋篇幅限制，因此本研究選取 TEJ 資料庫當中 2013 年初時各產業最高市值的標的股票，並排除 KY、DR 股，共選出 31 間公司，標的如表 1 所示。

[ 表 1 – 研究標的清單 ]

### 3-2. 研究模型介紹

#### 3-2-1. 梯度提升機 (Gradient Boosting Machine)

梯度提升機是一種集成學習法。梯度提升機(GBM)的演算法是 Boosting 演算法的一種。常見的梯度提升機主要有 XGBoost、LightGBM。

梯度提升機概念是基於之前建立的基底學習器的損失函數的梯度下降方向來建立下一個新的基底學習器目的就是希望通過集成這些基底學習器使得模型總體的損失函數不斷下降，將模型不斷改進。

#### 3-2-2. 輕量梯度提升模型 (Light Gradient Boosting)

本研究採用的模型為輕量梯度提升模型，英文為 Light Gradient Boosting，簡寫為 LightGBM。

LightGBM 是由微軟在 2017 年 4 月所提出。LightGBM 與 XGBoost 都是 Gradient Boosting Decision Tree (GBDT 梯度提升機) 的工具，GBDT 是透過迭代方式產生多棵決策數，並且由所有的決策數的結論累加起來做為最終答案的方法。

LightGBM 是輕量化的梯度提升機的模型，相對 XGBoost 來說 LightGBM 有訓練速度、記憶體占用低的優勢，也是近幾年常見的機器學習模型。

### 3-3. 機器學習標籤制定規則 (Labels)

以往傳統的機器學習，都是透過未來 N 天的報酬率是上漲、下跌、持平來決定標籤值 (label) 為 1、0 或 -1，這種做法所產生出來的模型，可以預測未來 N 天的股價是上漲或下跌，不過無法得知確切的潛在獲利以及風險，本研究預測 label 與傳統常見預測 label 相比，如圖 1 所示。

[ 圖 1. 預測 label 差異圖 ]

如果要明確知道每次的投資潛在報酬以及風險，就必須明確的定義每個時間點投資的出場情況，其中包含「停利」、「停損」、「固定時間出場」，並且透過這三種出場結果去制定標籤，設計的詳細規則如下：

以往預測股價類別	本研究預測類別
上漲 label = 1	停利 label = 1
下跌 label = -1	停損 label = -1
持平 label = 0	固定時間出場 label = 0

其中 labels 制定時，需要輸入三種參數，本研究的參數、參數值分別如下。

1. 持有天數：20 天，預設進場價格為隔日開盤價，固定時間出場為進場後 N 日，收盤價出場。
2. 停損比例：6%，進場後的收盤價若低於停損價，則出場
3. 停利比例：10%，進場後的收盤價若高於停利價，則出場

本文以台積電為例，繪製收盤價與 label(交易訊號)的資料來檢視資料，如圖 2 所示。

[ 圖 2. 交易訊號與收盤價共同繪製圖 ]

### 3-4. 機器學習所用到的特徵 (Features)

#### 3-4-1. 收盤價

收盤價，研究標的的收盤前最後一筆成交價格。

#### 3-4-2. 成交值

成交值，研究標的所有當日所有成交金額總和。

### 3-4-3. 累計成交量

累計成交量是由每日的成交量累計而成，而本特徵將會計算不同時間週期的累計成交量，本研究當中以季成交量（連續 60 日累計）、月成交量（連續 20 日累計）、週成交量為主（連續 5 日累計），分別偵測個股因應不同時間週期的累計成交量所帶來的反應。

本文以台積電為例，繪製收盤價與累計成交量的資料來檢視資料，如圖 3 所示。

[ 圖 3. 累計成交量與收盤價共同繪製圖 ]

### 3-4-4. 主力券商淨買賣超

計算過去一段時間，主力券商買賣超的淨張數，將主力券商的淨買賣進行加總，本研究當中指標頻率以季（連續 60 日累計）、月（連續 20 日累計）、週（連續 5 日累計）為主。

本研究主力券商的定義為過去一段時間淨買超最多的前 10 間券商。

本文以台積電為例，繪製收盤價與主力券商淨買賣超的資料來檢視資料，如圖 4 所示。

[ 圖 4. 主力券商淨買賣超與收盤價共同繪製圖 ]

### 3-4-5. 主力買賣超集中度

計算過去一段時間，本研究當中指標頻率以季（連續 60 日）、月（連續 20 日）、週（連續 5 日）為主，主力券商買賣超的淨張數除以時間段內的成交總量，主力淨買賣超的定義如 3-4-4 小節，去進行主力買賣超集中度的計算。

本研究主力券商的定義為過去一段時間淨買超最多的前 10 間券商。

本文以台積電為例，繪製收盤價與主力買賣超集中度的資料來檢視資料，如圖 5 所示。

[ 圖 5. 主力買賣超集中度與收盤價共同繪製圖 ]

### 3-4-6. 買賣家數差

當天所有台灣國內的券商分點進行統計，如果是該分點當日結算為淨買超，則被歸類為買進分點，反之，如果是該分點當日結算為淨賣超，則被歸類為賣出分點，並將所有買賣超分點分別進行加總，如果當天買家分點總計為 300 間、賣家分點總計為 150 間，則買賣家數差為 150 (300-150)，買賣家數差該值過大，則代表有特定主力券商正在大量賣出股票，買賣家數差該

值過小，則代表有特定主力券商整在大量買進該股票。

本文以台積電為例，繪製收盤價與買賣家數差的資料來檢視資料，如圖 6 所示。

[ 圖 6. 買賣家數差與收盤價共同繪製圖 ]

#### 3-4-7. 過去一段時間的累積報酬率（動能）

計算不同時間周期的動能，本研究動能的定義為特定周期內的累計報酬率。

本研究當中以季動能（連續 60 日累計）、月動能（連續 20 日累計）、週動能（連續 5 日累計）為主，分別偵測個股因應不同時間週期的動能所帶來的反應。

本文以台積電為例，繪製收盤價與累積報酬率的資料來檢視資料，如圖 7 所示。

[ 圖 7. 累積報酬率與收盤價共同繪製圖 ]

#### 3-4-8. 過去一段時間的最大回檔

計算不同時間週期內的回檔，回檔的定義是在一段期間內，是指在特定一段時間內任一歷史時點往後推，報酬率從頂峰跌至低谷，並恢復至原先頂峰的最大跌幅。

計算方法為每個時間點的最高累計報酬率減去當前累計報酬率，可以得到每個時間點的回檔，再將一段期間的回檔取得最大值，則為最大回檔。

本研究當中以季回檔（60 日）、月回檔（20 日）、週回檔（5 日）為主

本文以台積電為例，繪製收盤價與最大回檔的資料來檢視資料，如圖 8 所示。

[ 圖 8. 最大回檔與收盤價共同繪製圖 ]

#### 3-4-9. 移動平均線

移動平均是將一定時期內的價格加以平均並將不同時間的平均值連接起來，便可以得到一條移動平均線。研究週期為季（60 日）、月（20 日）、週（5 日），分別計算短中長期趨勢。

本文以台積電為例，繪製收盤價與移動平均線的資料來檢視資料，如圖 9 所示。

[ 圖 9. 移動平均線與收盤價共同繪製圖 ]



#### 3-4-10. 股價標準差

標準差在統計上，是衡量資料的離散程度，在本研究中是採用股票收盤價的標準差，也就是股價的波動程度，也就意味著波動風險。本研究的研究週期為季（60 日）、月（20 日）、週（5 日），分別計算短中長期風險。

本文以台積電為例，繪製收盤價與股價標準差的資料來檢視資料，如圖 10 所示。

[ 圖 10. 股價標準差與收盤價共同繪製圖 ]

#### 3-4-11. 股價當前位階

透過移動平均線以及 N 倍標準差，可以算出當前股價的位階，位階可以衡量目前股價的價格是偏高還是偏低。本研究的研究週期為季（60 日）、月（20 日）、週（5 日），分別計算短中長期位階風險。

該指標值如下。

值	意義
3	股價位於「移動平均值」加上「三倍標準差」之上
2	股價介於「移動平均值」加上「兩倍標準差」以及「移動平均值」加上「三倍標準差」之間
1	股價介於「移動平均值」加上「一倍標準差」以及「移動平均值」加上「兩倍標準差」之間
0	股價介於「移動平均值」加上「負一倍標準差」以及「移動平均值」加上「一倍標準差」之間
-1	股價介於「移動平均值」加上「負一倍標準差」以及「移動平均值」加上「負兩倍標準差」之間
-2	股價介於「移動平均值」加上「負兩倍標準差」以及「移動平均值」加上「負三倍標準差」之間
-3	股價位於「移動平均值」加上「負三倍標準差」之下

本文以台積電為例，繪製收盤價與股價當前位階的資料來檢視資料，如圖 11 所示。

[ 圖 11. 股價當前位階與收盤價共同繪製圖 ]

#### 3-4-12. 價格突破策略訊號

價格突破訊號，當股價向上突破指定週期的股價最高點時，則為向上突

破，當股價向下突破指定周期的股價最低點。本研究的研究週期為季（60 日）、月（20 日）、週（5 日），分別計算短中長期突破訊號。

該指標值如下。

值	意義
1	最近一次收盤價突破區間最高點
-1	最近一次收盤價突破區間最低點

本文以台積電為例，繪製收盤價與價格突破策略訊號的資料來檢視資料，如圖 12 所示。

[ 圖 12. 價格突破策略訊號與收盤價共同繪製圖 ]

### 3-5. 研究方法設計

機器學習基於讓機器模型找出資料的特徵或規則將其記錄起來，並在未來遇到相似的情況將結果輸出，理論上來說，輸入模型的訓練資料越多，越能夠訓練出好的模型，提升預測準確率。

#### 3-5-1. 樣本內樣本外測試

樣本內樣本外測試（In-Sample Out-of-Sample Testing）是一種統計學方法，用於比較樣本內和樣本外兩種情況，用於檢驗時間序列的數據是否具有趨勢或是季節性，通常用於股票或其他金融市場，檢測在不同期間的變化。

樣本內資料是指用來進行模型訓練、統計分析的資料，樣本外資料指的是用來驗證訓練後的模型是否可行的資料。

### 3-6. 機器學習績效評估

#### 3-6-1. 混淆矩陣

面對分類問題時，混淆矩陣式一種很實用的方法，而本研究預測分類共有「停利」、「停損」、「到期出場」三種，所以必須使用三分類的混淆矩陣，如下表所示，其中表中的 A、E、I 區塊，是預測與實際情形相同的區塊。至於 B、C、D、F、G、H 欄呈現的則是預測類別與實際類別不符合的區塊。

預測狀況\實際情形	實際停利 1	實際到期 0	實際停損 -1
預測停利 1	A	B	C

預測到期出場 0	D	E	F
預測停損 -1	G	H	I

從混淆矩陣中，可以看出預測類別數與實際類別數分不的狀況，還可以擴增計算出下列評估指標。

精準度 (Precision)：預測正確的數量除該類別的預測總數，例如  $A/(A+B+C)$ 、 $E/(D+E+F)$ 、 $I/(G+H+I)$ ，代表每個 Label 的準確度，當某個類別的準確度較高，則代表該類別的預測可能性相對較高。

召回率 (Recall)：預測正確的數量除與預測的所有數量，例如  $A/(A+D+G)$ 、 $E/(B+E+H)$ 、 $I/(C+F+I)$ ，召回率的意思是指在實際的情況下，模型預測成功的比率。當召回率越高，代表模型越有能力找出該類別的真實資料。

F1-Score：精確度乘上召回率的兩倍再除上精確度加召回率，F1-Score 是用來綜合考量召回率以及準確度的指標。

準確度 (Accuracy)：預測正確的總數除該樣本總數，例如： $(A+E+I) / \text{所有樣本加總}$ ，代表整體樣本當中，總共有多少比例是被準確預測。



### 3. 研究資料與方法

#### 4-1. 模型參數處理

##### 4-1-1. 資料預處理

資料集切分，本研究會將全部資料切分為樣本內、樣本外兩種資料集，較大的訓練集可以有效預防機器學習過度擬和，缺點是測試集過小，用於計算績效的時間也會變少。

本研究會將樣本內訓練集資料期間設為總樣本 75%，樣本外訓練集測試時間設為總樣本 25%。

##### 4-1-2. 特徵值敘述統計

本研究中的資料特徵主要分為兩類，第一種為股票價量資訊，包含累積報酬率、價格最大回檔、移動平均線、股價標準差、股價位階、價格突破訊號；第二種為主力資訊，主力淨買超、買賣超集中度、賣賣家數差。

以下將依照不同產業來區分研究標的，首先針對價量的特徵進行敘述統計，特徵分別有「收盤價」、「成交值」、「周成交量」、「月成交量」、「季成交量」，由此表可以觀察出研究期間中平均股價最高的標的為「和泰車」，平均成交值最高的標的為「台積電」，如表 2。

[表 2 – 各產業特徵敘述統計 1]

針對價量衍生出來的「回落」、「季均線」來進行敘述統計，可以從不同時間週期的回落平均值觀察出哪一些標的在研究期間價格下跌的平均幅度，均線的平均值觀察到的則與成交價平均值類似，如表 3。

[表 3 – 各產業特徵敘述統計 2]

針對價量衍生出的「累計報酬」、「價格突破策略訊號」指標進行敘述統計，平均累計報酬可以觀察出樣本在這一段期間的平均一段期間上漲或下跌的幅度，平均價格突破策略訊號則可以觀察出，股價位於多頭的時間比較長（該值大於 0）還是位於空頭的時間比較長（該值小於 0），如表 4。

[表 4 – 各產業特徵敘述統計 3]

針對價量衍生出的「相對位階」指標進行敘述統計，從相對位階的平均值可以觀察到，樣本在研究期間平均位於哪個相對高位還是相對低位，如表 5。

[ 表 5 - 各產業特徵敘述統計 4 ]

將針對券商分點日報表資料所衍生出的主力指標進行敘述統計，分別為「主力集中度」、「買賣家數差」，從主力集中度的平均值中可以觀察到該標的在研究期間的主力集中狀態，集中度越高的個股，代表籌碼被主力操控的機會越大，反之集中度越低的個股，代表籌碼被主力操控的機會越小，其中季主力集中度最高的是「農業科技的惠光」，最低的則是「文化創意的鈺象」、「電機機械的上銀」；買賣家數差平均值則是可以觀察到各標的平均被主力券商影響的程度，當買賣家數差平均越高時，代表都被特定的主力券商售出，平均越低時，代表都被特定的主力券商買進，買賣家數差平均最低的標的則為「文化創意的鈺象」、「電機機械的上銀」，平均最高的則為「光電業的群創」，如表 6。

[ 表 6 - 各產業特徵敘述統計 5 ]

#### 4-1-3. Label 值敘述統計

依照各股的股價資料去進行 label 計算，計算出每個交易日的 label，舉例來說，當日為 T 日，則抓取之後 20 天進行回測，可能產生三種：到期出場、停損、停利。

為了讓 label 樣本分佈在停損、停利的次數較平均，讓模型能夠不要因為樣本比例偏差造成預測不準確，所以 label 制定的結果盡量將三種結果發生次數相當，參數：最多持有 20 天出場、停利 10%、停損 6%。

將統計在樣本期間內，如表 7，每一檔樣本 label 出現的次數，透過發生的次數，可以發現停利的次數普遍低於停損的次數，從樣本的統計當中可以發現，研究期間當中最常發生停損的公司為「網家」，最常發生停利的公司也是「網家」，代表該公司在研究樣本期間的價格波動性高。而最少發生停損停利的公司為「中華電」，代表該公司在研究樣本期間的價格波動性低。

[ 表 7 - 各產業類別樣本次數 ]

#### 4-1-4. 特徵資料標準化

資料標準化是指在機器學習中，常見的資料預處理方法，目的是將數據轉換為標準分布，使各個特徵的數據範圍類似。

本研究所採用的是最小最大值標準化 (MinMaxScaler)，MinMaxScaler 是一種數據模型，用於將數據轉換成 0 到 1 之間的範圍。它的基本原理是將數據的最小值轉換為 0，最大值轉換為 1，中間的值則相應轉換為 0 到 1 之間的值。

MinMaxScaler 的公式如下：

$$X' = (X - Xmin) / (Xmax - Xmin)$$

其中，X'是轉換後的值，X 是原始值，Xmin 是數據中的最小值，Xmax 是數據中的最大值。

MinMaxScaler 常用於在機器學習中對數據進行預處理，以便使用各種算法對數據進行分析。它的優點在於簡單易用，能夠將數據轉換成同一範圍內的值，使得模型更容易收斂。

但是，MinMaxScaler 也有一些缺點，例如它不能處理離群值，會使得離群值對結果產生較大影響。因此，在使用 MinMaxScaler 之前，通常需要對數據進行離群值處理，以避免對結果產生不利影響。

#### 4-1-5. 特徵相關性檢定

本研究的標的分別有 31 間公司，本研究以相關性熱力圖展示特徵相關性，本文以台灣龍頭台積電樣本為例(如圖 13)，圖中的橫軸以及縱軸分別為各個特徵，交會的方塊中顯示的數值則是兩個特徵之間的相關性，如果是比較深色的方塊，代表相關性越高，而方塊越淺色，代表相關性越低。

[ 圖 13. 台積電特徵相關性 ]

從台積電特徵相關性檢定來看，由於特徵多數是由價量所衍生出來，所以收盤價與均線展現高度相關性，突破訊號與報酬率也呈現高度正相關。

另外可以注意到的是，主力集中度與累計報酬率也是高度正相關，代表主力集中買進時報酬率也是同步成長，而買賣家數差則是與多數的價格特徵呈現負相關。

## 4-2. 模型預測結果分析

#### 4-2-1. Feature 重要性

重要度排行表展示特徵在模型中重要的程度，本文以台積電當做範例，如圖 14，圖中的特徵重要性由上而下進行排序，縱軸為特徵排行，橫軸為重要性分數。

[ 圖 14. 台積電特徵重要性 ]

由台積電的訓練結果導出特徵重要性，以該檢定來看，以前十名分別為季標準差、季均線、季回落、月標準差、週主力、季報酬、季成交量、成交值(千元)、月回落、買賣家數差。可以從該結果得知，訓練的 label 為持有 20 天到期的交易訊號，相對應重要的特徵多數是以「季」為單位，以季為

單位的特徵在前 10 名當中有 5 名，而該五個特徵均為價量衍生出來的特徵。其中可以發現主力資訊在前 10 名特徵中佔有 2 名，分別為週主力、買賣家數差，代表過去 5 天的主力買賣超以及過去 1 天的券商買賣超，會對於未來 20 天的交易訊號的預測有明確影響力。

本研究除了觀察單一標的特徵重要性以外，為了檢視特定特徵是否廣泛有效於整體股票市場，本研究將特徵重要性前十名進行統計，也就是統計本研究內的所有樣本的前 10 名特徵重要性，檢視哪類特徵出現在重要性前十名的次數最高，以此判斷該特徵對於整體市場是有影響力，而非單一標的，統計結果如圖 15。

[ 圖 15. 重要特徵前 10 名出現排行 ]

由圖 3 中，可以看到該模型重要程度最高，最頻率出現的特徵是「季標準差」，出現的頻率為 30 次，所有樣本為 31 個，也就代表幾乎包辦所有樣本的前 10 名重要特徵，也就代表對於預測 20 天後的交易訊號，季標準差是有明顯影響力的。除此之外，主力的特徵中最常出現在重要排行榜前 10 名的是「季主力集中度」，出現的頻率為 12 次，而「月主力集中度」，出現的頻率為 9 次，代表主力資訊在本模型當中，對於特定產業的樣本具有影響力，並非所有樣本均有影響力。

#### 4-2-3. 混沌矩陣

本研究將透過混沌矩陣來進行機器學習的結果評價。在進行混沌矩陣的結果介紹前，首先檢視一下學習預測結果與實際的結果比較圖，如圖 16。

[ 圖 16. 模型預測與實際的結果繪製 ]

本文將透過龍頭股台積電來進行範例介紹，繪製混沌矩陣圖來進行實際、預測樣本解讀介紹，以圖 4 為例，縱軸代表為實際發生的次數，橫軸代表預測發生的次數，各分為三個值，分別為-1（停損）、0（到期出場）、1（停利），所以必須使用三分類的混淆矩陣，而混沌矩陣當中有九格，由左至右由上至下的進行排序的話，會由「A」、「B」、「C」、「D」、「E」、「F」、「G」、「H」、「I」，分別對應到的數字為「35」、「27」、「0」、「166」、「127」、「9」、「65」、「46」、「1」。

台積電樣本中，測試期間的天數為 476 天，可以從混沌矩陣中的九格進行加總得來。測試期間，所有實際停損訊號的 62 天，可以從「A」、「B」、「C」三格中加總得來，實際到期出場訊號的 302 天，可以從「C」、「D」、「E」三格中加總得來，實際停損訊號的 112 天，可以從「G」、「H」、「I」三格中加總得來；另外，所有預測停損訊號的 266 天，可以從「A」、「D」、「G」三格中加總得來，預測到期出場訊號的 200 天，可以從「B」、「E」、「H」三格中加總得來，預測停損訊號的 10 天，可以從「C」、「F」、「I」三

格中加總得來。

預測與實際都為停損的樣本數為 35，預測與實際都為到期出場的樣本數為 127，預測與實際都為停利的樣本數為 1。

[ 圖 4. 台積電混沌矩陣圖 ]

接下來透過台積電的混沌矩陣，去進行評價函數的計算，台積電的混沌矩陣評價函數結果如表 8。

評價函數所採用的計算方法是將所有結果加權平均計算，接下來將依序介紹各評價函數的計算過程。精準度的計算過程，停損的精準度(precision)計算為預測成功的停損樣本天數「A」除上全部預測停損的樣本天數「A」、「D」、「G」，到期出場の精準度(precision)計算為預測成功的到期出場樣本天數「E」除上全部預測到期出場的樣本天數「B」、「E」、「H」，停利的精準度(precision)計算為預測成功的停利樣本天數「I」除上全部預測停利的樣本天數「C」、「F」、「I」。

召回率的計算過程，停損的召回率(recall)計算為預測成功的停損樣本天數「A」除上全部實際停損的樣本天數「A」、「B」、「C」，到期出場的召回率(recall)計算為預測成功的到期出場樣本天數「E」除上全部實際到期出場的樣本天數「C」、「D」、「E」，停利的召回率(recall)計算為預測成功的停利樣本天數「I」除上全部實際的樣本天數「G」、「H」、「I」。

F1-Score 的計算過程為，精準度(precision)乘上召回率(recall)的兩倍再除上精準度(precision)加召回率(recall)。

準確度(accuracy)則為所有預測成功的樣本天數「A」、「E」、「I」去除上所有樣本。

[ 表 8 - 台積電混沌矩陣評價指標 ]

而本研究當中，將所有的標的回測評價結果統整至一張圖表中，如表 9，欄位有準確度(accuracy)、精準度(precision)、召回率(recall)、F1-score，其中準確度(accuracy)、精準度(precision)、召回率(recall)三個欄位所呈現的結果是停損、到期出場、停利三個類別所計算出來的平均值。

表 9 中，特別將準確率高達 7 成的樣本行結果用粗體標註出來，本模型準確率高達 7 成的樣本有六組，約莫為全體樣本的 22%，產業別包含「食品工業」、「塑膠工業」、「鋼鐵工業」、「金融保險」、「通信網路業」、「農業科技」，從這些產業別可以發現到，價格波動起伏較低的公司往往會有比較高的預測力。

[ 表 9 - 各產業混沌矩陣評價指標 ]



將全樣本的混沌矩陣評價函數進行統計計算，並列出敘述統計表，如表 10。

從表 10 當中可以觀察到，最高的，整體的平均預測準確率為 5 成 2，整體的平均預測精準度為 5 成 6，整體的平均預測召回率為 5 成 2，整體的平均 F1-score 為 5 成。

[ 表 10 - 全樣本混沌矩陣敘述統計 ]

## 4. 結論

本研究使用 2013 年 1 月至 2021 年 12 月的台股日頻率資料，以分點日報表所計算出的主力指標以及價量衍伸的指標來進行交易訊號的預測，使用的機器學習模型為 LightGBM，在 Ke et al. (2017)發表的文章中證實，LightGBM 可以大幅提升運算效率並得到與其他機器學習相同精確度的結果。

本論文改變以往預測股價漲跌的財務模型預測，預測股價漲跌的模型並沒有辦法提供潛在獲利、潛在風險的評估，可能導致實作上無法良好控制風險的可能。

就特徵相關性而言，可以發現主力指標相較於同週期的累計報酬率有高度正相關，代表主力的買進、賣出動作與價格變動是有相關性的，並且依照特徵重要性前 10 名排行榜顯示出，主力相關指標也經常出現在排行榜上，尤其為主力集中度指標，排行上的主力集中度的週期重要程度是季>月>周。

本研究是透過價量、籌碼特徵來進行交易策略未來出場訊號的判斷，而對於該方法統計出的最重要特徵是「標準差」，代表過去的標準差，對於該模型未來的預測有非常重要的資訊意義，而主力買賣超集中度也是對於該模型有一定影響力。

以混沌矩陣的結果來說，效果最好的產業分別為「食品工業」、「塑膠工業」、「鋼鐵工業」、「金融保險」、「通信網路業」、「農業科技」，從結果可以推論，價格起伏較穩定的產業龍頭股較容易在本次研究中取得比較高的評分。

## 5. 未來研究建議

本研究所採用的研究方法是透過預測未來策略訊號，而並非預測股價本身，主要的用意在於可以衡量潛在風險與報酬。

透過機器學習演算法，意義在於我們可以偵測到潛藏在資料當中的微小特徵關聯，讓機器自動分類 label。

在未來後續的研究當中，還可以針對以下項目進行進一步的研究與探討：基於本研究的交易訊號，來進行回測覆盤演練，檢查交易訊號與實際回測報酬率之間的關聯，判斷預測的交易訊號是否對於策略績效有明顯的影響力

## 6. 參考文獻

### 中文部分

1. 廖子瑄(2014)，「股市主力交易行為與委託簿資訊內涵之互動」，國立中山大學財務管理學系碩士班
2. 張力元 (2018)，「深度學習應用於股價走勢之研究：以大陸市場為例」，國立政治大學風險管理與保險學系研究所
3. 戴欣頻(2019)，「分點券商熱門度與股價報酬關係之實證研究 以台灣股市蘋果概念股為例」，國立中央大學產業經濟研究所
4. 戴欣頻(2019)，「分點券商熱門度與股價報酬關係之實證研究 以台灣股市蘋果概念股為例」，國立中央大學產業經濟研究所
5. 劉馨文、王正豪 (2020)，「基於混合注意力機制與長短期記憶之股票趨勢預測」，The 32nd Conference on Computational Linguistics and Speech Processing The Association for Computational Linguistics and Chinese Language Processing
6. 林尚毅(2022)，「以機器學習建構交易策略」，國立臺北大學統計學系

### 英文部分

1. Bozcuk A. and Lasfer A. M. (2005). The Information Content of Institutional Trades on the London Stock Exchange. The Journal of Financial and Quantitative Analysis, pp.621-644
2. Boehmer E. and Kelley E. K. (2016). Institutional Investors and the Informational Efficiency of Prices. The Society for Financial Studies , Oxford University Press. Sponsor: The Society for Financial Studies. , 3563-3594
3. Chang P. C., Fan C.Y, , Liu C. H. (2009).Integrating a piecewise linear representation method and a neural network model for stock trading points prediction. IEEE Trans Syst Man Cybern Part C Appl Rev.;39:80e92.
4. Dash R. and Dash P. K. (2016). A hybrid stock trading framework

integrating technical analysis with machine learning techniques . The Journal of Finance and Data Science , 42-57

5. Gunduz H.(2021). An efficient stock market prediction model using hybrid feature reduction method based on variational autoencoders and recursive feature elimination. Financial innovation , 7:28
6. Ke G. and Meng Q. and Finley T. and Wang T. and Chen W. and Ma W. and Ye Q. and Liu T. Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Long Beach Convention Center
7. Kimoto T. and Asakawa K. and Yoda M. and Takeoka M. (1990). Stock Market Prediction System with Modular Neural Networks. IEEE 1990 IJCNN International Joint Conference on Neural Networks, San Diego, CA, USA, 1990, pp. 1-6 vol.1, doi: 10.1109/IJCNN.1990.137535.
8. Kara Y., Boyacioglu M.A., Baykan O.K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the Istanbul stock exchange. Expert Syst Appl. .38:5311e5319.
9. Li Z. and Xu W. and Li A. (2022). Research on multi factor stock selection model based on LightGBM and Bayesian Optimization. Procedia Computer Science , 1234-1240
10. Lee Y. T. and Lin J. C. and Liu Y. J. (1999). Trading patterns of big versus small players in an emerging market: An empirical analysis . Journal of Banking & Finance , 701-725
11. Lee J. W. and Park J. and Lee J. and Hong E.(2007). A multiagent approach to Q-learning for daily stock trading. IEEE Trans Syst Man Cybern Part A Syst Humans. 37:864e877.
12. Rodríguez-González A., García-Crespo A. , Colomo-Palacios R., Iglesias F. G., GomezBerbís J. M. (2011). CAST: using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator. Expert Syst Appl. 38:11489e11500.
13. Teixeira L. A., De O. A. (2010). A method for automatic stock trading combining technical analysis and nearest neighbor classification. Expert Syst Appl. 37:6885e6890.

14. Yang Y. ,Wu Y., Wang P., Xujiali. (2021). Stock Price Prediction Based on XGBoost and LightGBM . E3S Web of Conferences 275, 01040

表 1－研究標的清單

證券代碼	簡稱	TSE 產業別	證券代碼	簡稱	TSE 產業別
1101	台泥	水泥工業	2317	鴻海	其他電子業
1216	統一	食品工業	1802	台玻	玻璃陶瓷
2707	晶華	觀光事業	3481	群創	光電業
1301	台塑	塑膠工業	1907	永豐餘	造紙工業
5522	遠雄	建材營造	2002	中鋼	鋼鐵工業
2207	和泰車	汽車工業	2105	正新	橡膠工業
2308	台達電	電子零組件業	2603	長榮	航運業
1402	遠東新	紡織纖維	2330	台積電	半導體業
9904	寶成	其他	2412	中華電	通信網路業
2912	統一超	貿易百貨	2347	聯強	電子通路業
2049	上銀	電機機械	4994	傳奇	資訊服務業
1789	神隆	生技醫療	6505	台塑化	油電燃氣業
2382	廣達	電腦及週邊設備業	2882	國泰金	金融業
1605	華新	電器電纜	3293	鈺象	文化創意
1722	台肥	化學工業	8044	網家	電子商務
1101	台泥	水泥工業			

表格備註：證券代碼為台灣證交所所制定的各上市公司股票代碼，公司簡稱是上市公司的簡稱，TSE 產業別則是證交所分類各上市公司的產業別標籤。

表 2－各產業特徵敘述統計 1

產業別	標的	收盤價	成交值	週成交量	月成交量	季成交量
水泥工業	台泥	22.78	448384.85	55757.32	222537.04	663138.07
食品工業	統一	49.11	586212.68	47273.10	189256.01	568462.82
塑膠工業	台塑	65.61	553497.72	31726.22	126813.31	378777.14
紡織纖維	遠東新	21.85	197756.79	34077.82	136210.29	407881.45
電機機械	上銀	196.20	758479.45	14859.83	59346.33	178236.10
電器電纜	華新	10.14	172792.75	61523.15	243702.79	722795.65
玻璃陶瓷	台玻	14.99	50219.21	15519.99	60474.24	169822.66
造紙工業	永豐餘	9.91	35496.91	12594.78	49858.90	148538.42
鋼鐵工業	中鋼	18.61	402643.33	85844.63	340988.99	1012092.75
橡膠工業	正新	46.28	322377.58	26217.86	104977.32	314697.48
汽車工業	和泰車	340.29	148396.14	1813.51	7257.15	21775.03
建材營造	遠雄	23.54	48742.65	6080.18	24078.71	71679.61
航運業	長榮	26.74	404905.03	102545.85	382991.89	1013859.53
觀光事業	晶華	143.56	47753.69	1196.05	4781.62	14224.65
金融業	國泰金	33.04	859251.82	95520.94	383289.88	1163077.95
貿易百貨	統一超	209.89	346269.87	6755.32	27017.46	80690.00
其他	寶成	29.66	258603.93	34812.65	139109.49	410528.10
化學工業	台肥	37.58	130044.07	12800.35	51268.96	154503.66
生技醫療	神隆	37.52	89614.27	9593.90	38447.09	116673.53
油電燃氣業	台塑化	76.82	293052.56	15771.00	62956.11	188162.24
半導體業	台積電	178.35	7607389.68	180533.57	722555.44	2163389.82
電腦及週邊設備業	廣達	45.62	486434.09	37030.52	148050.70	443594.82
光電業	群創	9.36	993517.17	426788.27	1682364.70	4907788.73
通信網路業	中華電	82.35	856342.03	41075.32	164173.51	492756.25
電子零組件業	台達電	132.63	924751.82	29407.35	117709.89	350628.01
電子通路業	聯強	27.88	190737.83	23139.02	92513.42	278063.99
資訊服務業	傳奇	70.26	28870.65	1341.16	5359.76	16050.11
其他電子業	鴻海	71.52	3221274.71	186422.66	744101.59	2224474.16
文化創意	鈞象	92.50	433287.21	7836.31	31277.16	93253.19
農業科技	惠光	20.38	4619.43	716.86	2867.09	8633.60
電子商務	網家	185.37	203325.57	4686.50	18692.32	55553.27

表格備註：本表顯示各產業別中指標性的公司企業，並將樣本期間內的資料進行敘述統計，統計的欄位有收盤價、成交值、周成交量（連五日累積成交量）、月成交量（連十日累積成交量）、月成交量（連二十日累積成交量）。

表 3－各產業特徵敘述統計 2

產業別	標的	季回落	月回落	週回落	季均線	月均線	週均線
水泥工業	台泥	0.10	0.05	0.02	22.50	22.69	22.76
食品工業	統一	0.08	0.05	0.02	48.69	48.98	49.08
塑膠工業	台塑	0.09	0.05	0.02	65.10	65.42	65.56
紡織纖維	遠東新	0.10	0.05	0.02	21.76	21.82	21.85
電機機械	上銀	0.18	0.09	0.03	193.81	195.21	195.98
電器電纜	華新	0.13	0.07	0.02	9.98	10.09	10.13
玻璃陶瓷	台玻	0.15	0.07	0.02	15.13	15.03	15.00
造紙工業	永豐餘	0.10	0.05	0.02	9.72	9.84	9.89
鋼鐵工業	中鋼	0.07	0.04	0.01	18.57	18.59	18.60
橡膠工業	正新	0.11	0.05	0.02	46.43	46.33	46.29
汽車工業	和泰車	0.14	0.07	0.03	333.60	338.23	339.85
建材營造	遠雄	0.10	0.05	0.02	23.23	23.43	23.52
航運業	長榮	0.13	0.07	0.02	26.43	26.58	26.69
觀光事業	晶華	0.09	0.04	0.01	144.27	143.76	143.60
金融業	國泰金	0.09	0.05	0.02	32.82	32.97	33.02
貿易百貨	統一超	0.08	0.05	0.02	207.81	209.23	209.76
其他	寶成	0.12	0.06	0.02	29.54	29.62	29.66
化學工業	台肥	0.08	0.04	0.01	37.54	37.56	37.57
生技醫療	神隆	0.15	0.08	0.03	37.89	37.64	37.55
油電燃氣業	台塑化	0.10	0.05	0.02	76.38	76.66	76.79
半導體業	台積電	0.09	0.05	0.02	172.15	176.25	177.90
電腦及週邊設備業	廣達	0.10	0.06	0.02	45.13	45.45	45.58
光電業	群創	0.17	0.08	0.03	9.39	9.36	9.36
通信網路業	中華電	0.04	0.02	0.01	81.71	82.15	82.31
電子零組件業	台達電	0.11	0.06	0.02	130.48	131.87	132.46
電子通路業	聯強	0.10	0.05	0.02	27.76	27.83	27.87
資訊服務業	傳奇	0.18	0.09	0.03	70.35	70.28	70.26
其他電子業	鴻海	0.10	0.05	0.02	71.01	71.32	71.48
文化創意	鈦象	0.19	0.10	0.03	88.14	91.08	92.18
農業科技	惠光	0.07	0.03	0.01	20.19	20.32	20.36
電子商務	網家	0.21	0.10	0.04	185.45	185.36	185.37

表格備註：本表顯示各產業別中指標性的公司企業，並將樣本期間內的資料進行敘述統計，統計的欄位有季回落（60 日內的最大回落）、月回落（20 日內的最大回落）、周回落（5 日內的最大回落）、季均線（60 日內的平均收盤價）、月均線（20 日內的平均收盤價）、周均線（5 日內的平均收盤價）



表 4 - 各產業特徵敘述統計 3

產業別	標的	季報酬	月報酬	週報酬	季突破	月突破	週突破
水泥工業	台泥	0.028	0.010	0.002	0.316	0.17	0.09
食品工業	統一	0.022	0.007	0.002	0.174	0.08	0.05
塑膠工業	台塑	0.018	0.007	0.002	0.318	0.22	0.08
紡織纖維	遠東新	0.011	0.005	0.001	0.048	0.11	0.06
電機機械	上銀	0.039	0.015	0.004	0.046	0.15	0.08
電器電纜	華新	0.037	0.014	0.004	0.155	0.15	0.05
玻璃陶瓷	台玻	(0.006)	0.001	0.000	(0.066)	(0.06)	0.01
造紙工業	永豐餘	0.029	0.011	0.003	0.110	0.01	0.00
鋼鐵工業	中鋼	0.005	0.003	0.001	0.058	0.06	0.05
橡膠工業	正新	(0.001)	(0.000)	(0.000)	(0.056)	(0.10)	0.00
汽車工業	和泰車	0.056	0.018	0.004	0.067	0.06	0.06
建材營造	遠雄	0.023	0.009	0.002	0.167	0.07	0.03
航運業	長榮	0.025	0.012	0.003	(0.085)	(0.04)	(0.01)
觀光事業	晶華	(0.006)	(0.001)	(0.000)	0.035	(0.02)	(0.09)
金融業	國泰金	0.022	0.006	0.002	0.137	0.11	0.04
貿易百貨	統一超	0.028	0.009	0.002	0.034	0.15	0.04
其他	寶成	0.016	0.006	0.001	0.005	0.00	0.04
化學工業	台肥	0.004	0.002	0.001	0.199	(0.01)	(0.03)
生技醫療	神隆	(0.008)	(0.003)	(0.001)	(0.279)	(0.15)	(0.06)
油電燃氣業	台塑化	0.014	0.006	0.002	0.118	0.08	0.08
半導體業	台積電	0.064	0.022	0.006	0.530	0.31	0.13
電腦及週邊設備業	廣達	0.024	0.009	0.002	0.062	0.14	0.09
光電業	群創	0.010	0.006	0.001	0.075	(0.03)	(0.01)
通信網路業	中華電	0.017	0.006	0.001	0.403	0.21	0.10
電子零組件業	台達電	0.039	0.013	0.003	0.193	0.06	0.07
電子通路業	聯強	0.012	0.005	0.001	0.160	0.05	0.03
資訊服務業	傳奇	0.015	0.007	0.002	(0.190)	(0.18)	(0.06)
其他電子業	鴻海	0.022	0.008	0.002	0.137	0.18	0.06
文化創意	鈺象	0.123	0.037	0.009	(0.059)	0.04	0.04
農業科技	惠光	0.020	0.007	0.002	0.289	0.17	0.11
電子商務	網家	0.015	0.008	0.002	(0.097)	(0.03)	(0.02)

表格備註：本表顯示各產業別中指標性的公司企業，並將樣本期間內的資料進行敘述統計，統計的欄位有季報酬（60 日累積報酬率）、月報酬（20 日累積報酬率）、週報酬（5 日累積報酬率）、季突破（收盤價突破 60 日的收盤價最高則為 1 直到反向突破 60 日內最低價則為 0）、月突破（收盤價突破 20 日的收盤價最高則為 1 直到反向突破 20 日內最低價則為 0）、週突破（收盤價突破 5 日的收盤價最高則為 1 直到反向突破 5 日內最低價則為 0）

表 5 - 各產業特徵敘述統計 4

產業別	標的	季相對位階	月相對位階	週相對位階
水泥工業	台泥	0.25	0.14	0.03
食品工業	統一	0.19	0.08	0.03
塑膠工業	台塑	0.23	0.11	0.04
紡織纖維	遠東新	0.16	0.09	0.02
電機機械	上銀	0.16	0.11	0.02
電器電纜	華新	0.20	0.07	(0.00)
玻璃陶瓷	台玻	(0.06)	(0.06)	(0.02)
造紙工業	永豐餘	0.05	0.02	(0.02)
鋼鐵工業	中鋼	0.04	0.05	0.00
橡膠工業	正新	(0.05)	(0.06)	(0.01)
汽車工業	和泰車	0.15	0.06	0.01
建材營造	遠雄	0.15	0.05	0.01
航運業	長榮	0.04	0.02	(0.01)
觀光事業	晶華	(0.09)	(0.10)	(0.04)
金融業	國泰金	0.13	0.07	0.01
貿易百貨	統一超	0.13	0.06	0.02
其他	寶成	0.06	0.04	0.02
化學工業	台肥	0.04	0.01	(0.00)
生技醫療	神隆	(0.16)	(0.08)	(0.06)
油電燃氣業	台塑化	0.12	0.05	0.02
半導體業	台積電	0.45	0.23	0.07
電腦及週邊設備業	廣達	0.14	0.10	0.03
光電業	群創	0.04	(0.01)	(0.01)
通信網路業	中華電	0.35	0.19	0.10
電子零組件業	台達電	0.13	0.06	0.02
電子通路業	聯強	0.09	0.04	0.01
資訊服務業	傳奇	(0.11)	(0.09)	(0.06)
其他電子業	鴻海	0.18	0.12	0.05
文化創意	鈞象	0.09	0.09	0.01
農業科技	惠光	0.24	0.12	0.05
電子商務	網家	0.00	0.02	(0.00)

表格備註：本表顯示各產業別中指標性的公司企業，並將樣本期間內的資料進行敘述統計，統計的欄位有季相對位階、月相對位階、周相對位階

表 6 - 各產業特徵敘述統計 5

產業別	標的	周集中度	月集中度	季集中度	買賣家數差
水泥工業	台泥	0.32	0.21	0.14	23.56
食品工業	統一	0.32	0.21	0.15	2.89
塑膠工業	台塑	0.33	0.22	0.15	(13.78)
紡織纖維	遠東新	0.35	0.23	0.15	7.84
電機機械	上銀	0.17	0.11	0.07	(69.81)
電器電纜	華新	0.27	0.19	0.14	(13.07)
玻璃陶瓷	台玻	0.31	0.19	0.12	5.25
造紙工業	永豐餘	0.31	0.19	0.12	(5.39)
鋼鐵工業	中鋼	0.31	0.22	0.16	(11.78)
橡膠工業	正新	0.29	0.19	0.13	36.80
汽車工業	和泰車	0.31	0.21	0.14	(12.56)
建材營造	遠雄	0.34	0.23	0.15	7.89
航運業	長榮	0.29	0.18	0.12	13.77
觀光事業	晶華	0.36	0.23	0.13	8.56
金融業	國泰金	0.25	0.16	0.11	5.82
貿易百貨	統一超	0.36	0.24	0.16	14.82
其他	寶成	0.32	0.22	0.15	(1.44)
化學工業	台肥	0.29	0.19	0.11	27.35
生技醫療	神隆	0.21	0.11	0.07	(7.59)
油電燃氣業	台塑化	0.34	0.22	0.14	0.21
半導體業	台積電	0.26	0.16	0.10	(8.95)
電腦及週邊設備業	廣達	0.33	0.22	0.14	(5.80)
光電業	群創	0.16	0.09	0.06	51.19
通信網路業	中華電	0.34	0.24	0.17	31.08
電子零組件業	台達電	0.31	0.20	0.13	(24.12)
電子通路業	聯強	0.32	0.21	0.14	12.47
資訊服務業	傳奇	0.31	0.19	0.12	(13.10)
其他電子業	鴻海	0.20	0.13	0.09	19.57
文化創意	鈦象	0.13	0.07	0.04	(65.53)
農業科技	惠光	0.46	0.26	0.14	0.86
電子商務	網家	0.22	0.14	0.09	(40.73)

表格備註：本表顯示各產業別中指標性的公司企業，並將樣本期間內的資料進行敘述統計，統計的欄位有季集中度（60 日主力買賣超除上 60 日累計成交量）、月集中度（20 日主力買賣超除上 20 日累計成交量）、周集中度（5 日主力買賣超除上 5 日累計成交量）

表 7 - 各產業類別樣本次數

	停損(-1)	到期出場(0)	停利(1)
台泥	462	1748	205
統一	305	1956	154
台塑	405	1864	146
遠東新	462	1772	181
上銀	1050	713	652
華新	813	1049	553
台玻	785	1142	488
永豐餘	492	1594	329
中鋼	282	1972	161
正新	534	1645	236
和泰車	701	1177	537
遠雄	531	1614	270
長榮	695	1166	554
晶華	459	1742	214
國泰金	394	1794	227
統一超	301	1951	163
寶成	601	1489	325
台肥	362	1888	165
神隆	827	1213	375
台塑化	521	1657	237
台積電	395	1762	258
廣達	597	1538	280
群創	1008	778	629
中華電	18	2391	6
台達電	703	1267	445
聯強	504	1689	222
傳奇	1032	799	584
鴻海	622	1541	252
鈺象	1078	687	650
惠光	282	2006	127
網家	1146	562	707

表格備註：本表顯示各研究標的中，各種 Label 值所發生的次數。

表 8－台積電混沌矩陣評價指標

	準確度 (accuracy)	精準度 (precision)	召回率 (recall)	F1-score
停損		0.13	0.56	0.21
到期出場		0.64	0.42	0.51
停利		0.1	0.01	0.02
整體	0.34	0.44	0.34	0.35

表格備註：本表顯示台積電的混沌矩陣的表格值。

表 9－各產業混沌矩陣評價指標

產業別	名稱	準確度(accuracy)	精準度(precision)	召回率(recall)	F1-score
水泥工業	台泥	0.37	0.79	0.37	0.47
食品工業	統一	<b>0.78</b>	<b>0.70</b>	<b>0.78</b>	<b>0.74</b>
塑膠工業	台塑	<b>0.75</b>	<b>0.62</b>	<b>0.75</b>	<b>0.66</b>
紡織纖維	遠東新	0.59	0.57	0.59	0.58
電機機械	上銀	0.33	0.33	0.33	0.32
電器電纜	華新	0.42	0.52	0.42	0.42
玻璃陶瓷	台玻	0.32	0.48	0.32	0.32
造紙工業	永豐餘	0.47	0.65	0.47	0.50
鋼鐵工業	中鋼	<b>0.90</b>	<b>0.83</b>	<b>0.90</b>	<b>0.86</b>
橡膠工業	正新	0.54	0.52	0.54	0.53
汽車工業	和泰車	0.29	0.58	0.29	0.21
建材營造	遠雄	0.56	0.62	0.56	0.57
汽車工業	長榮	0.36	0.39	0.36	0.37
觀光事業	晶華	0.58	0.36	0.58	0.43
金融保險	國泰金	<b>0.81</b>	<b>0.79</b>	<b>0.81</b>	<b>0.79</b>
貿易百貨	統一超	0.58	0.76	0.58	0.65
其他	寶成	0.55	0.54	0.55	0.54
生物科技類	台肥	0.55	0.61	0.55	0.58
通訊網路業	神隆	0.46	0.44	0.46	0.44
軟體業	台塑化	0.53	0.52	0.53	0.52
證券業	台積電	0.34	0.44	0.34	0.35
電腦及週邊設備業	廣達	0.60	0.51	0.60	0.55
光電業	群創	0.43	0.46	0.43	0.42
通信網路業	中華電	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
電子零組件業	台達電	0.43	0.44	0.43	0.34
電子通路業	聯強	0.21	0.56	0.21	0.26
資訊服務業	傳奇	0.42	0.39	0.42	0.34
其他電子業	鴻海	0.53	0.41	0.53	0.46
文化創意	鈞象	0.35	0.43	0.35	0.25
農業科技	惠光	<b>0.72</b>	<b>0.76</b>	<b>0.72</b>	<b>0.74</b>
電子商務	網家	0.27	0.40	0.27	0.29

表格備註：本表顯示各個研究標的混沌矩陣評價函數值，分別有四個欄位：準確度(accuracy)、精準度(precision)、召回率(recall)、F1-score。

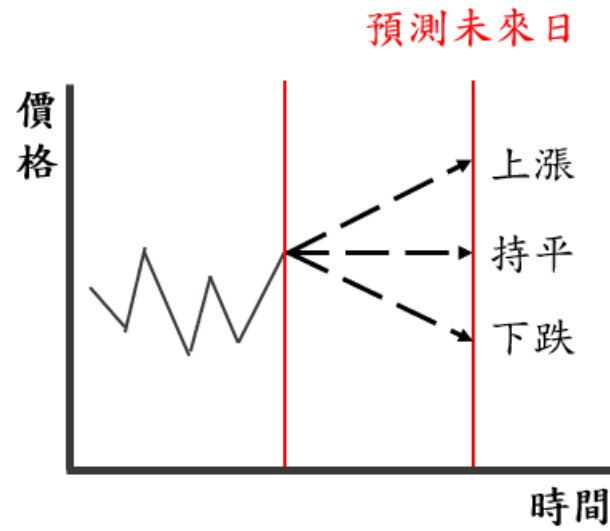
表 10 – 全樣本混沌矩陣敘述統計

	準確度(accuracy)	精準度 (precision)	召回率(recall)	F1-score
count	31.00	31.00	31.00	31.00
mean	0.52	0.56	0.52	0.50
std	0.19	0.16	0.19	0.19
min	0.21	0.33	0.21	0.21
25%	0.37	0.44	0.37	0.35
50%	0.53	0.52	0.53	0.47
75%	0.58	0.64	0.58	0.58
max	1.00	1.00	1.00	1.00

表格備註：本表顯示各種研究標的混沌矩陣評價函數敘述統計。

圖1. 預測 Label 差異圖

傳統預測label



本研究預測label

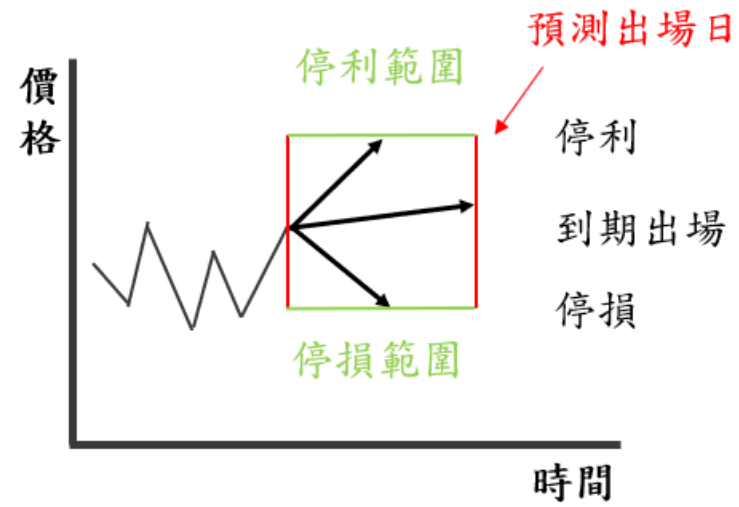




圖2. 交易訊號與收盤價共同繪製圖

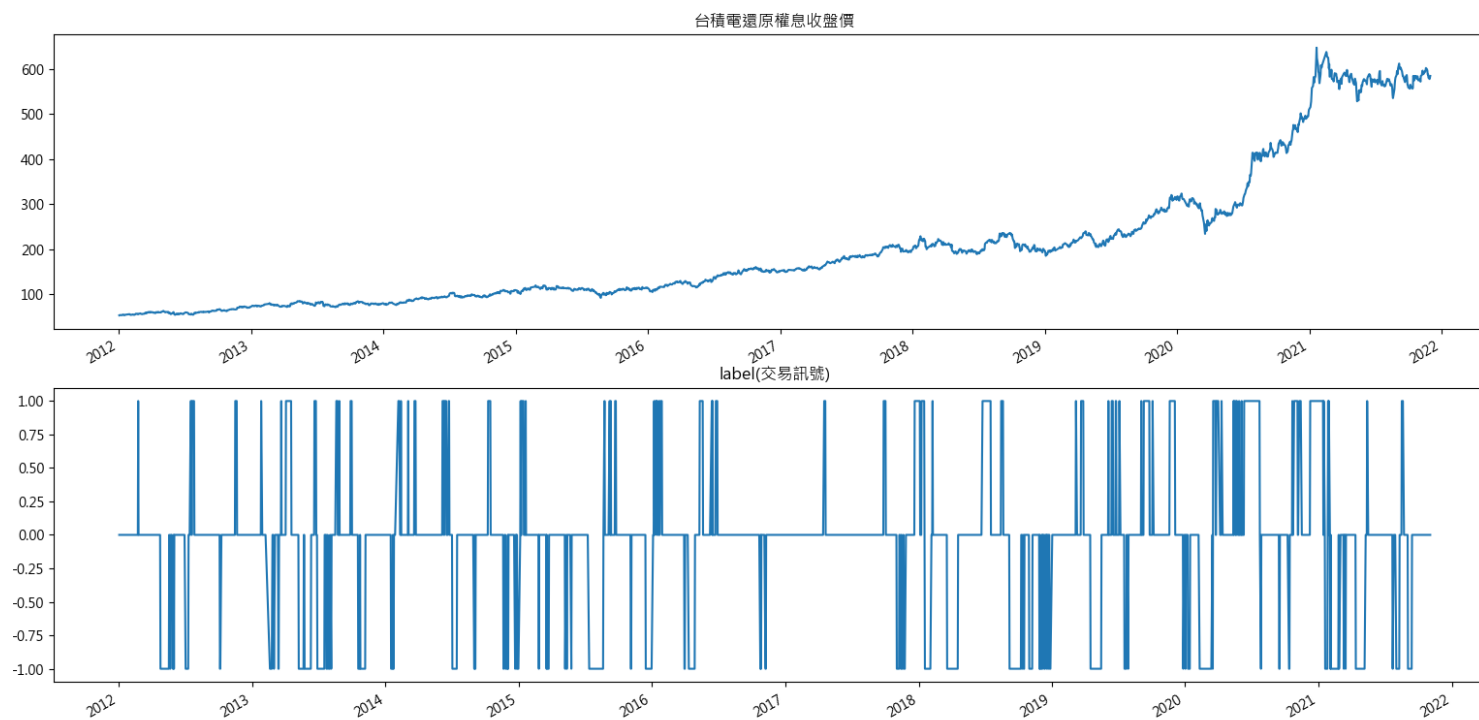


圖3. 累計成交量與收盤價共同繪製圖

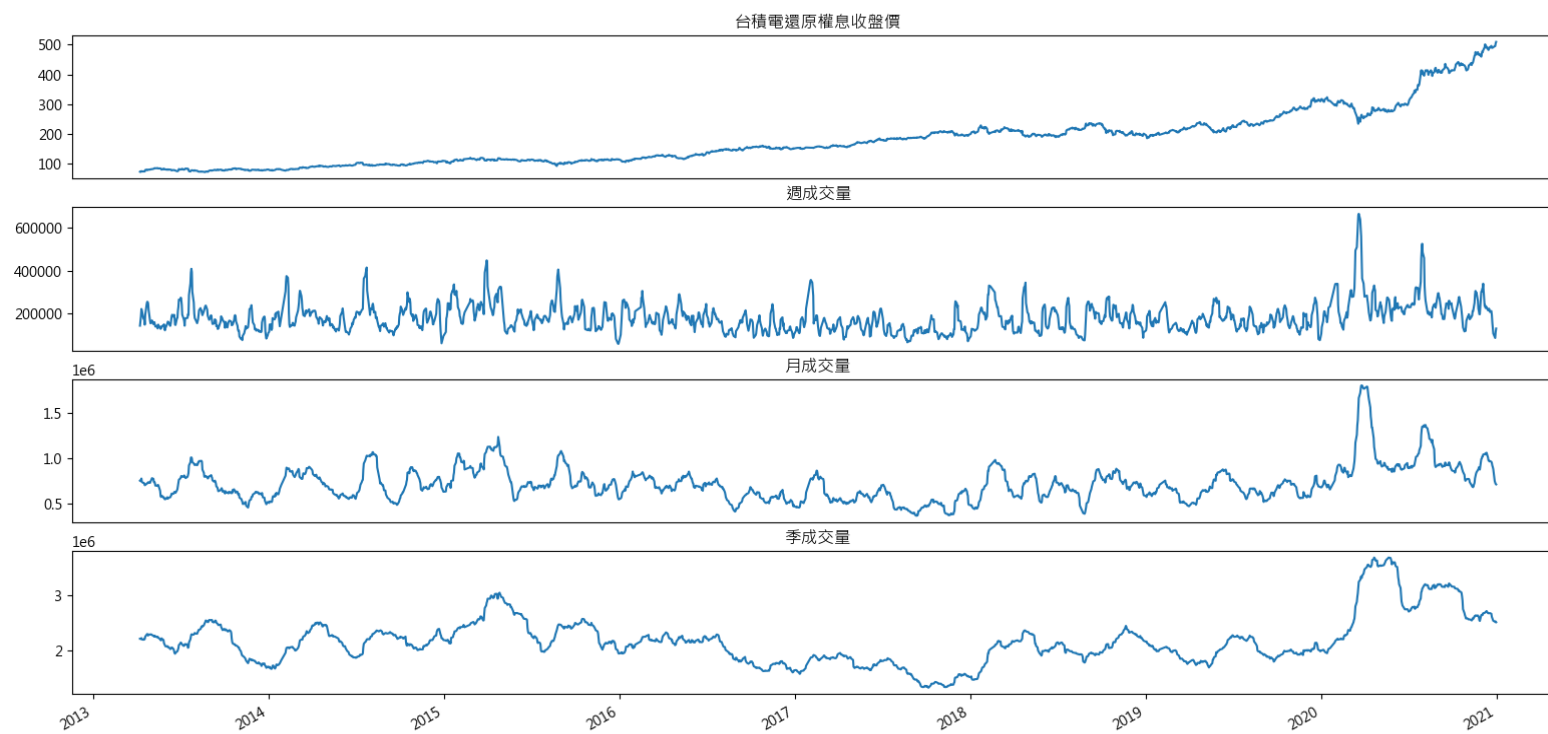


圖4. 主力券商淨買超與收盤價共同繪製圖

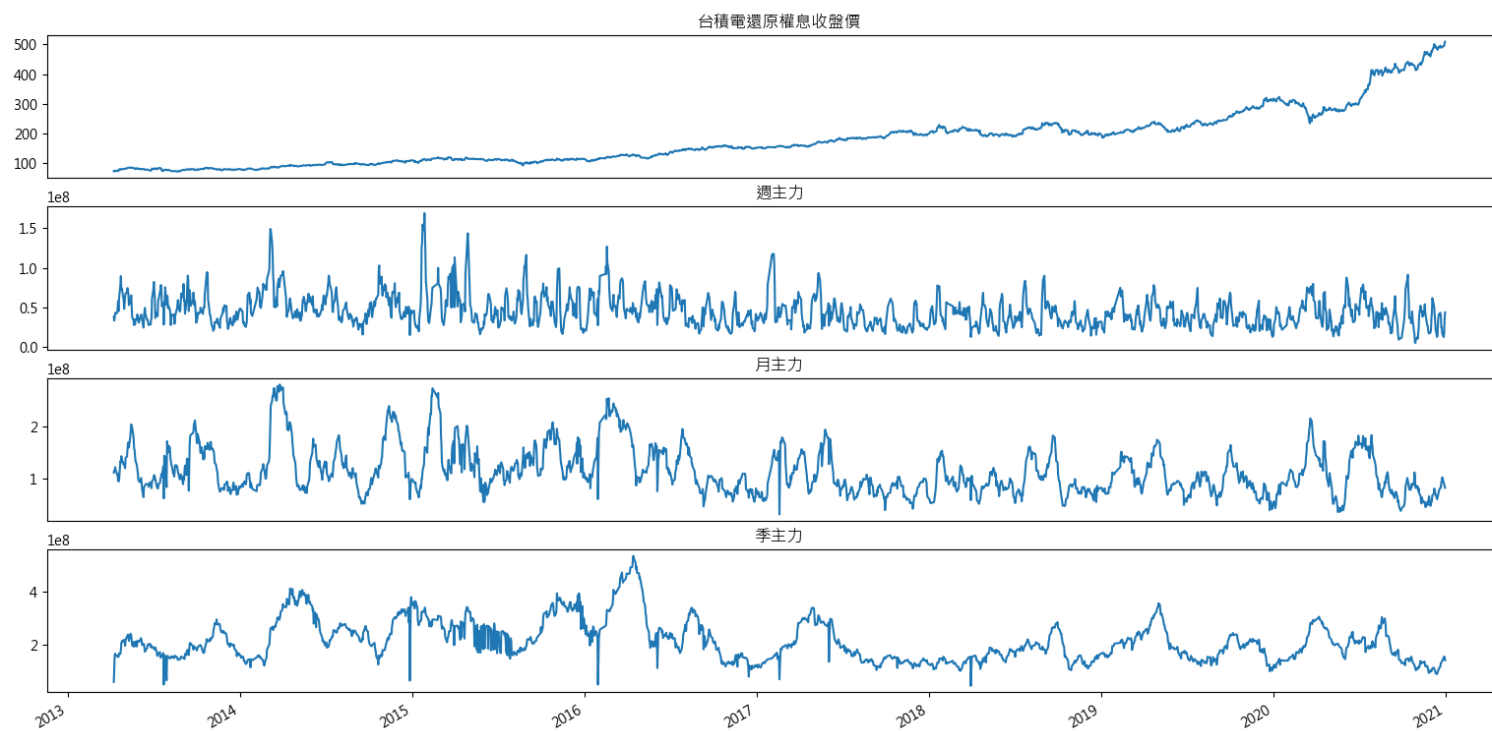


圖5. 主力買賣超集中度與收盤價共同繪製圖

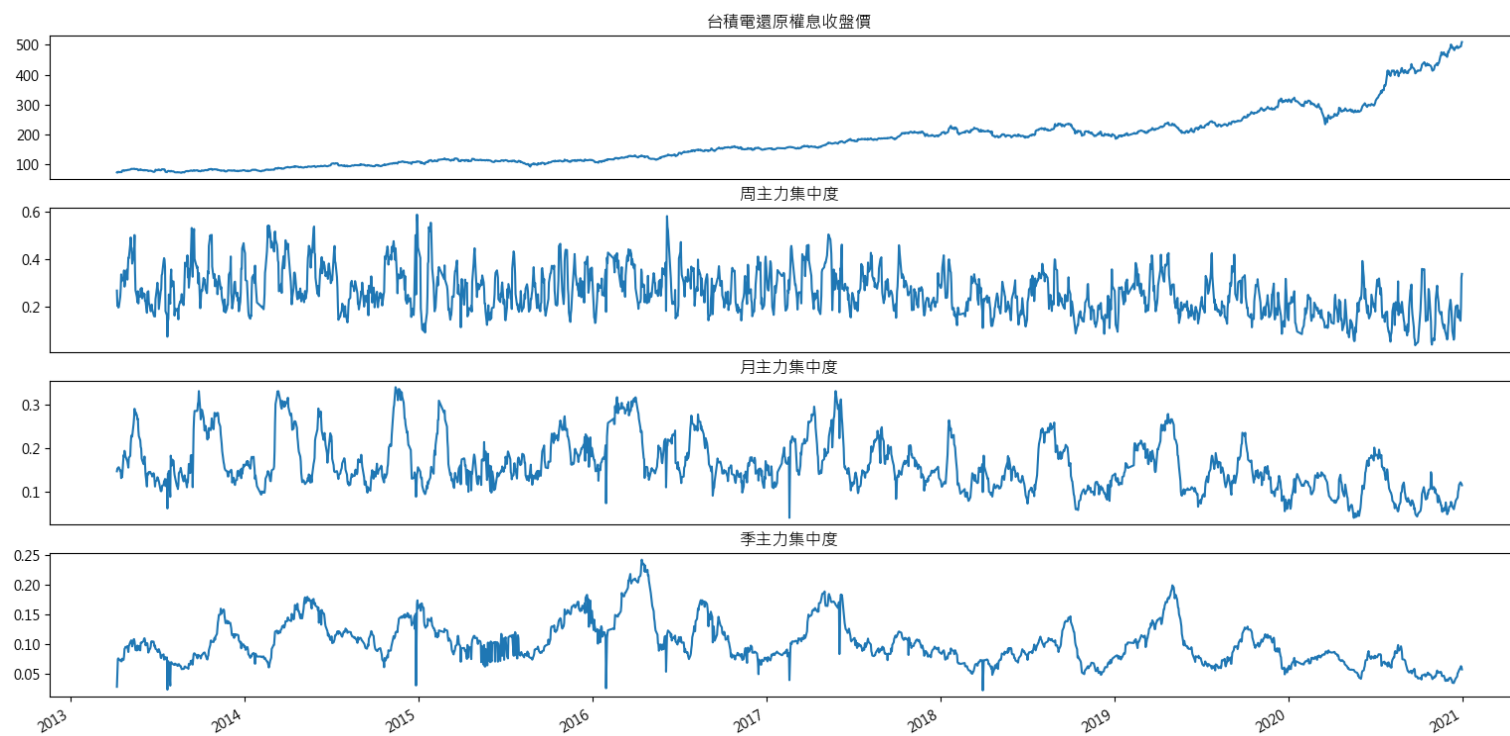


圖6. 買賣家數差與收盤價共同繪製圖

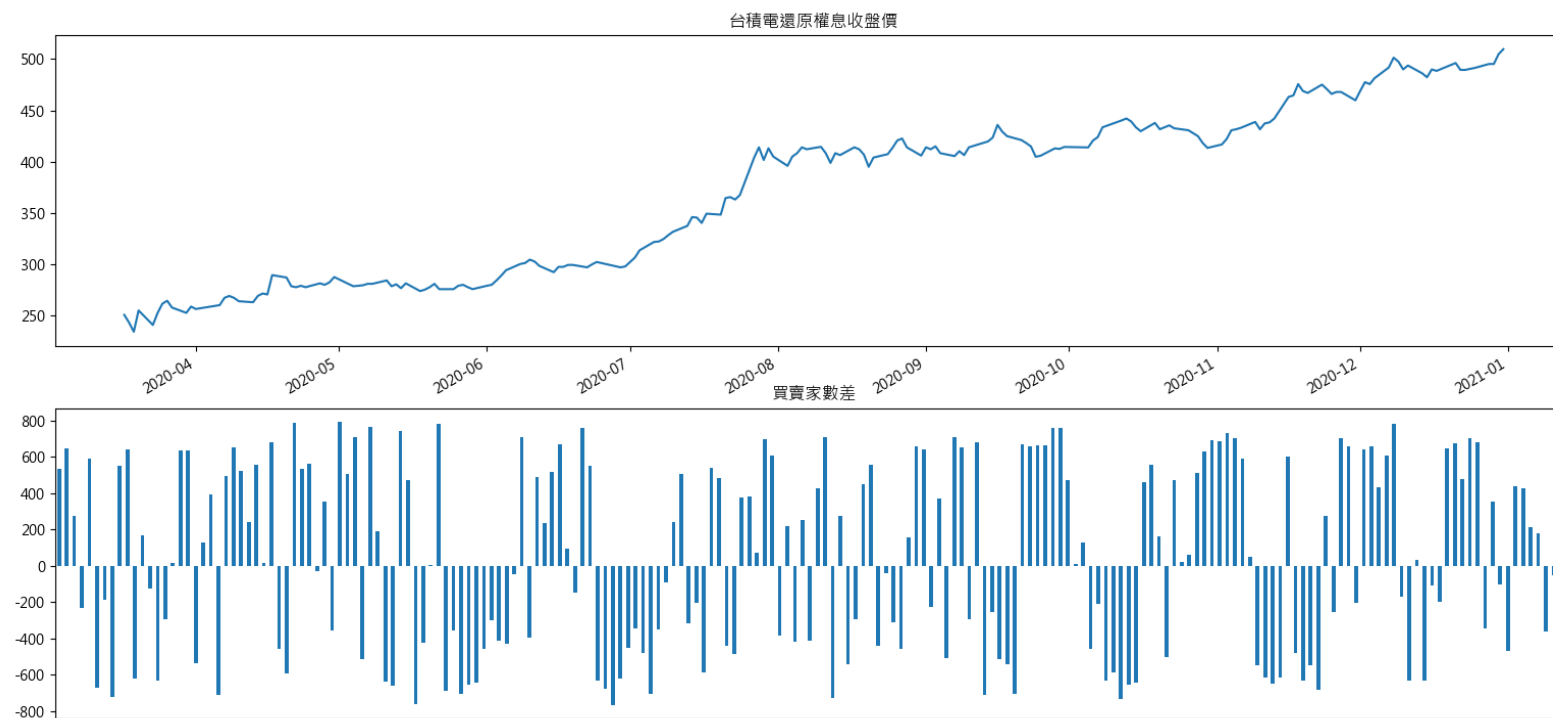


圖7. 累積報酬率與收盤價共同繪製圖

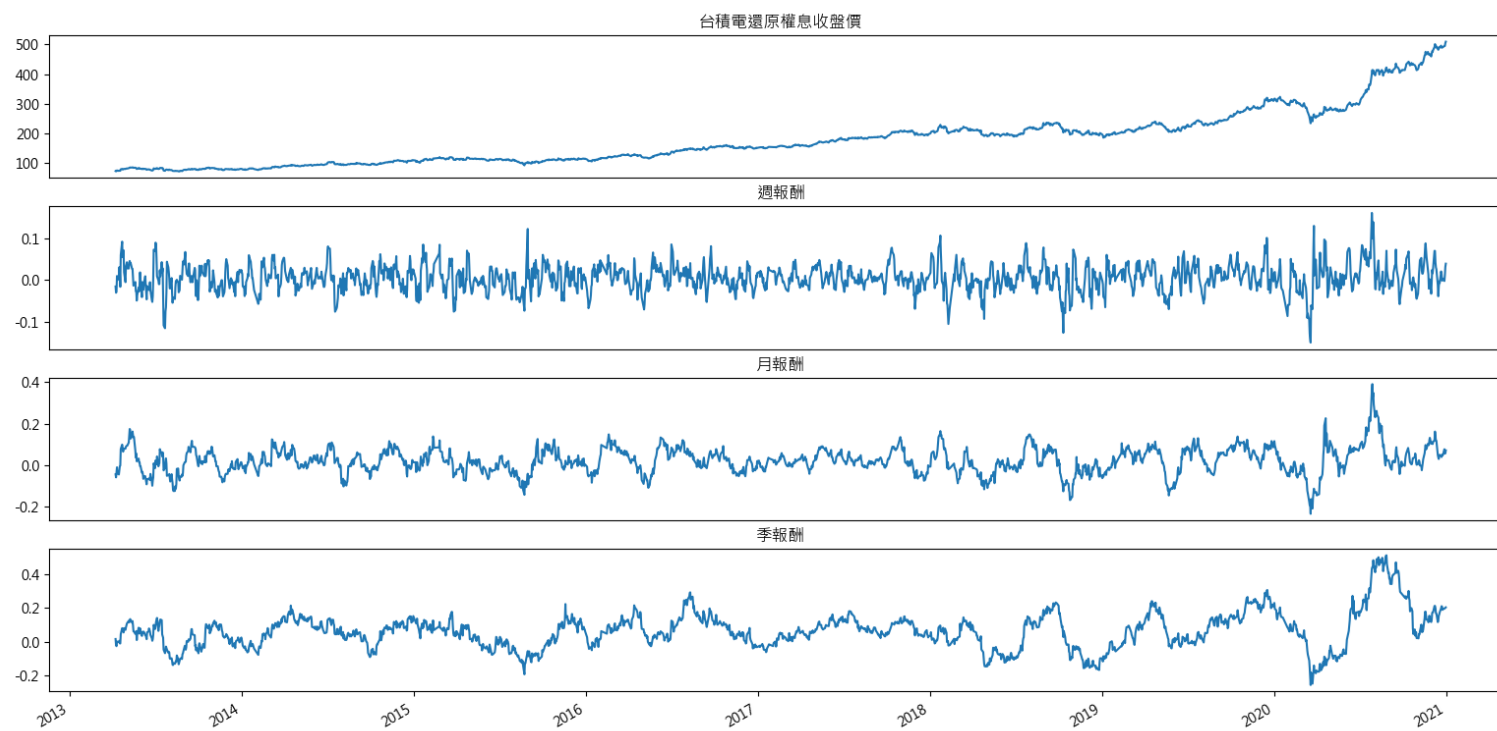


圖8. 最大回檔與收盤價共同繪製圖

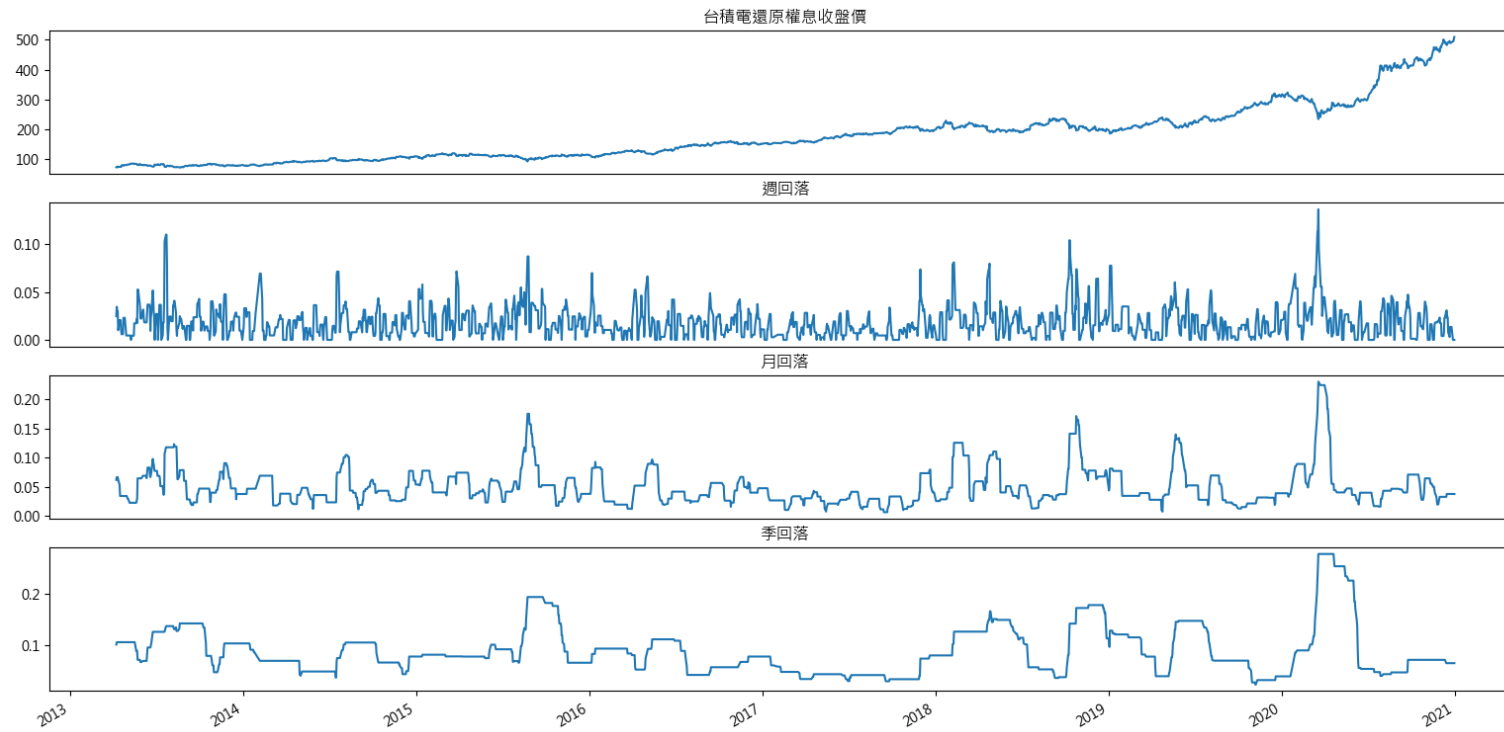


圖9. 移動平均線與收盤價共同繪製圖

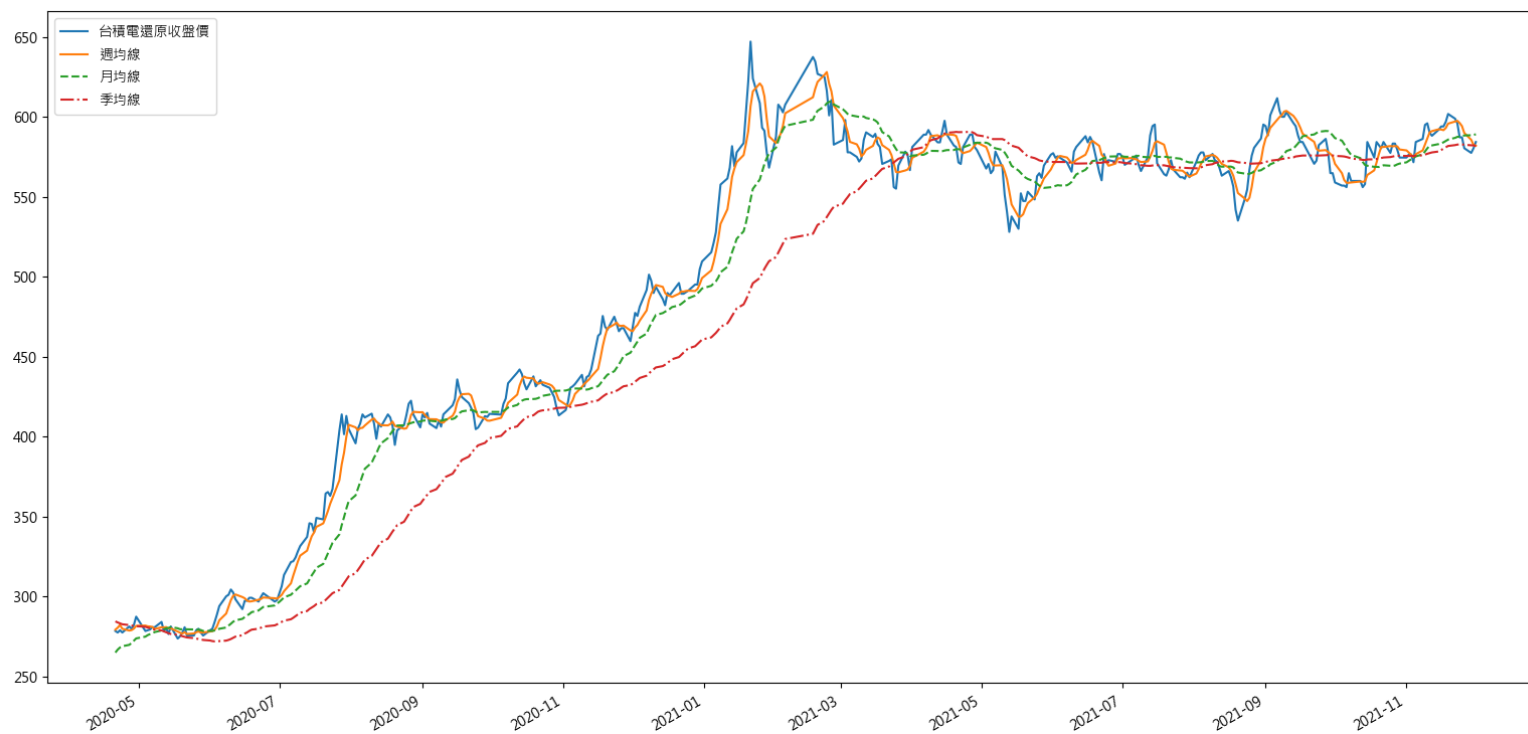




圖10. 股價標準差與收盤價共同繪製圖

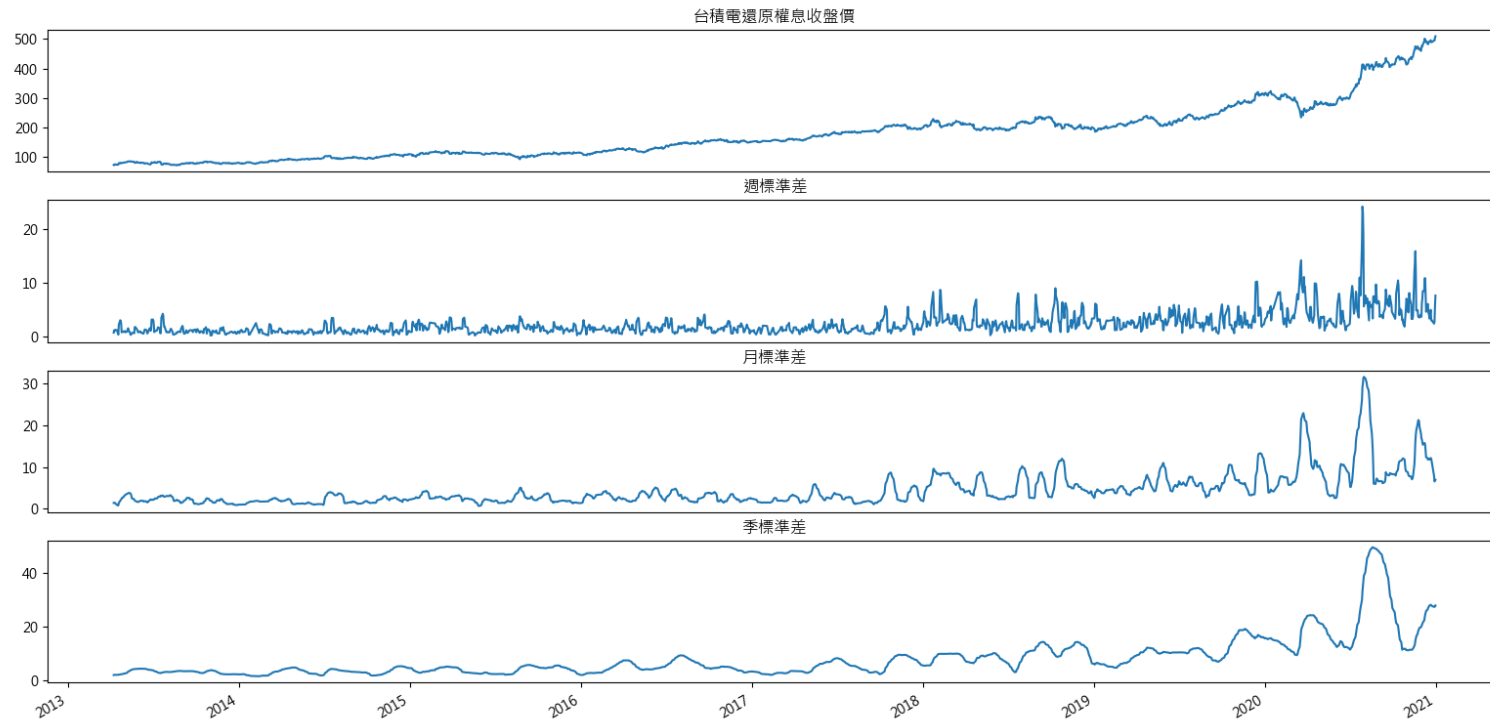


圖11. 股價當前位階與收盤價共同繪製圖

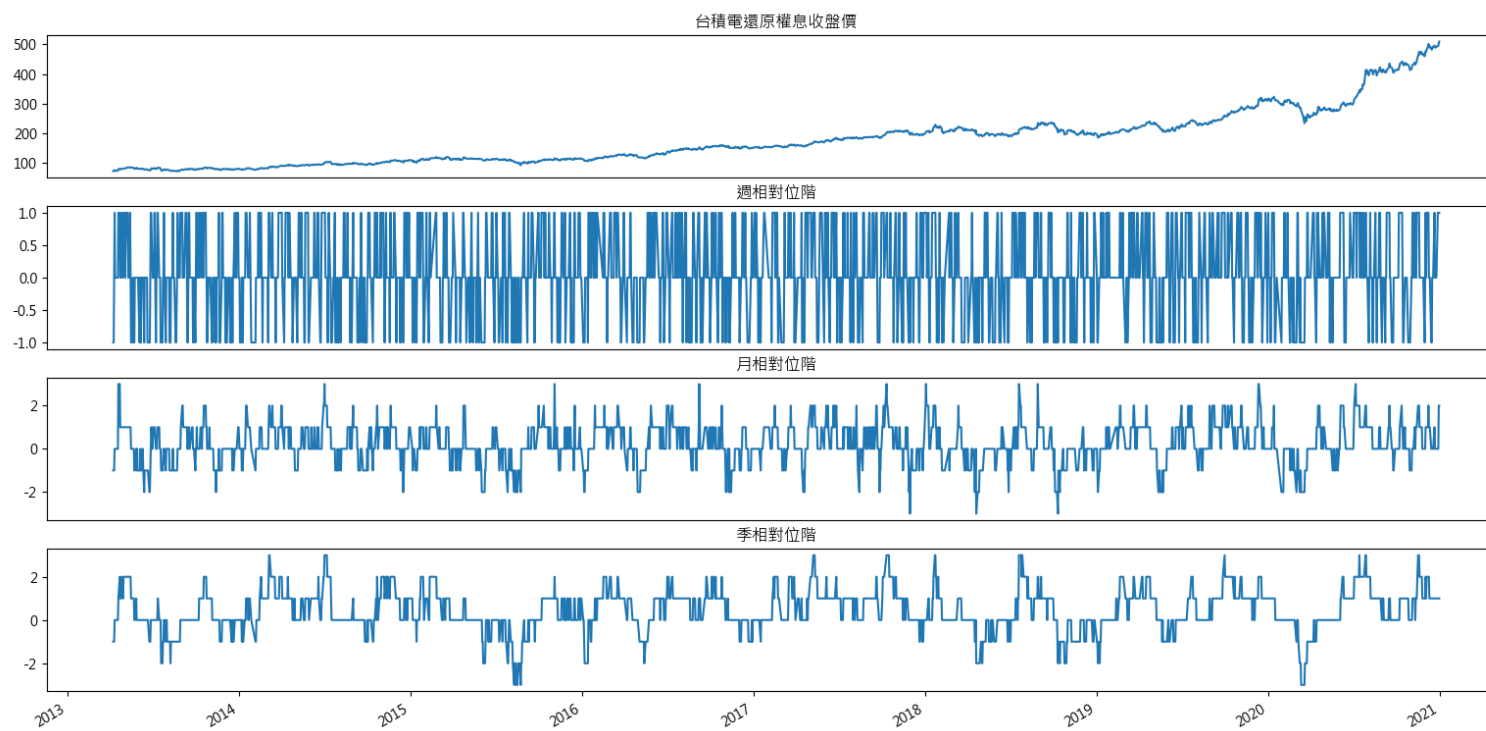


圖12. 價格突破策略訊號與收盤價共同繪製圖

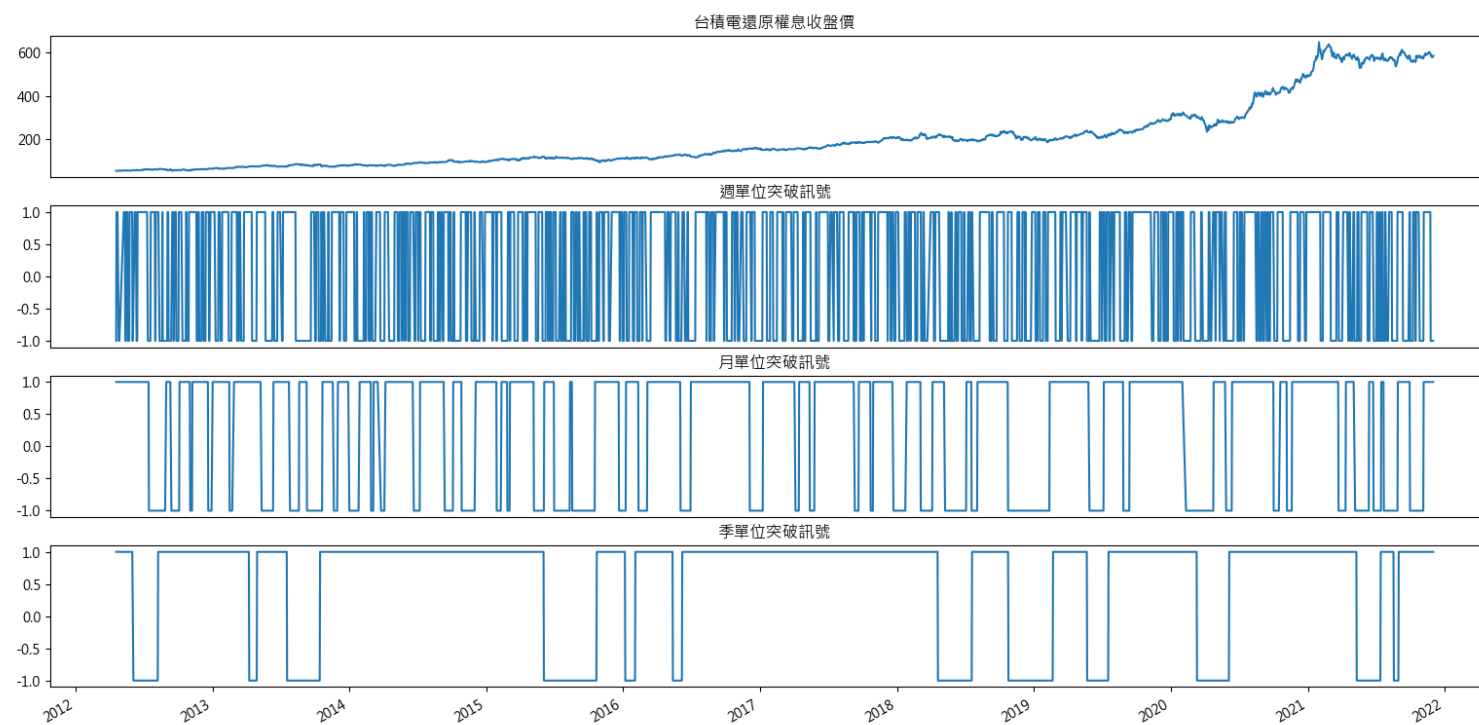


圖13. 模型預測與實際的結果繪製

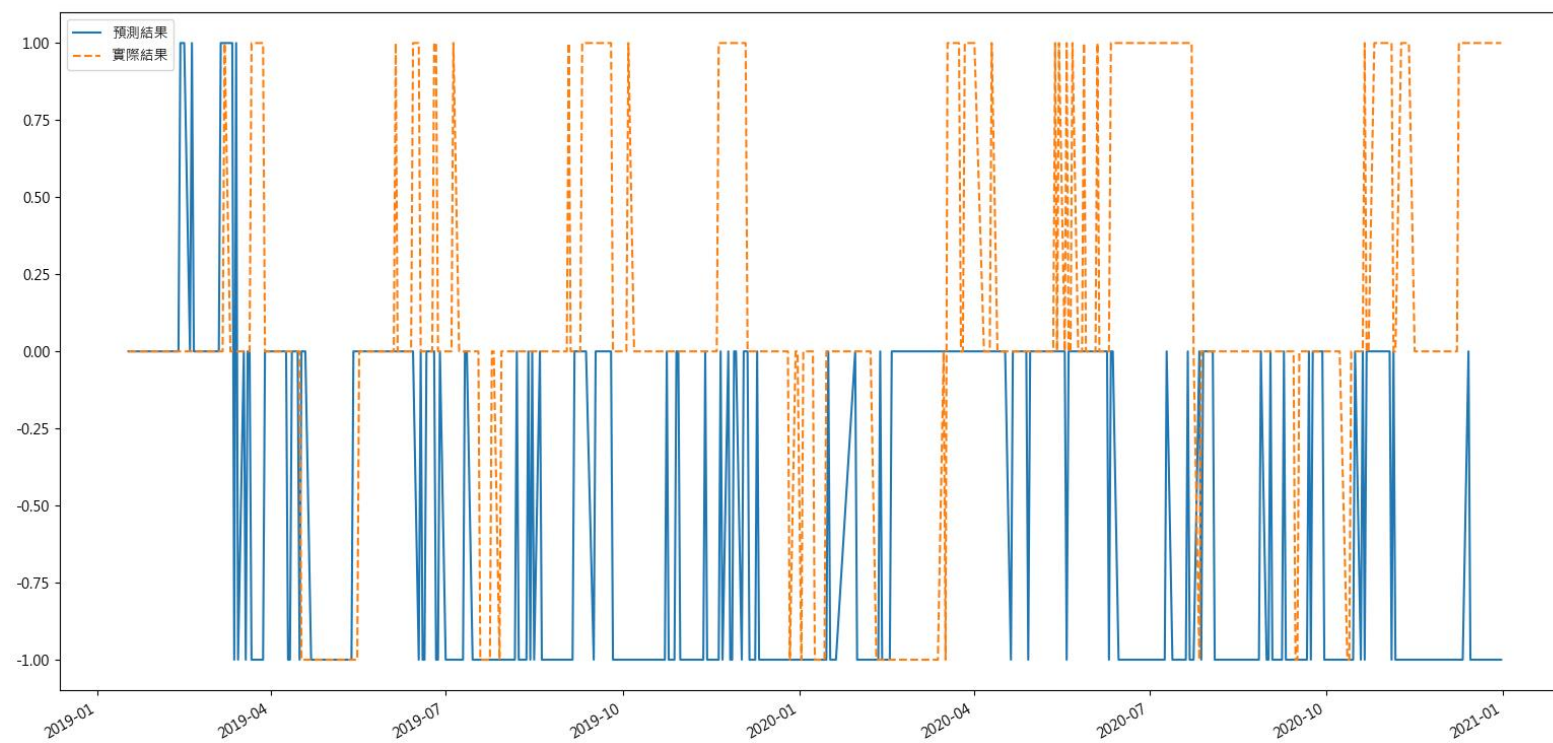


圖14. 台積電特徵相關性

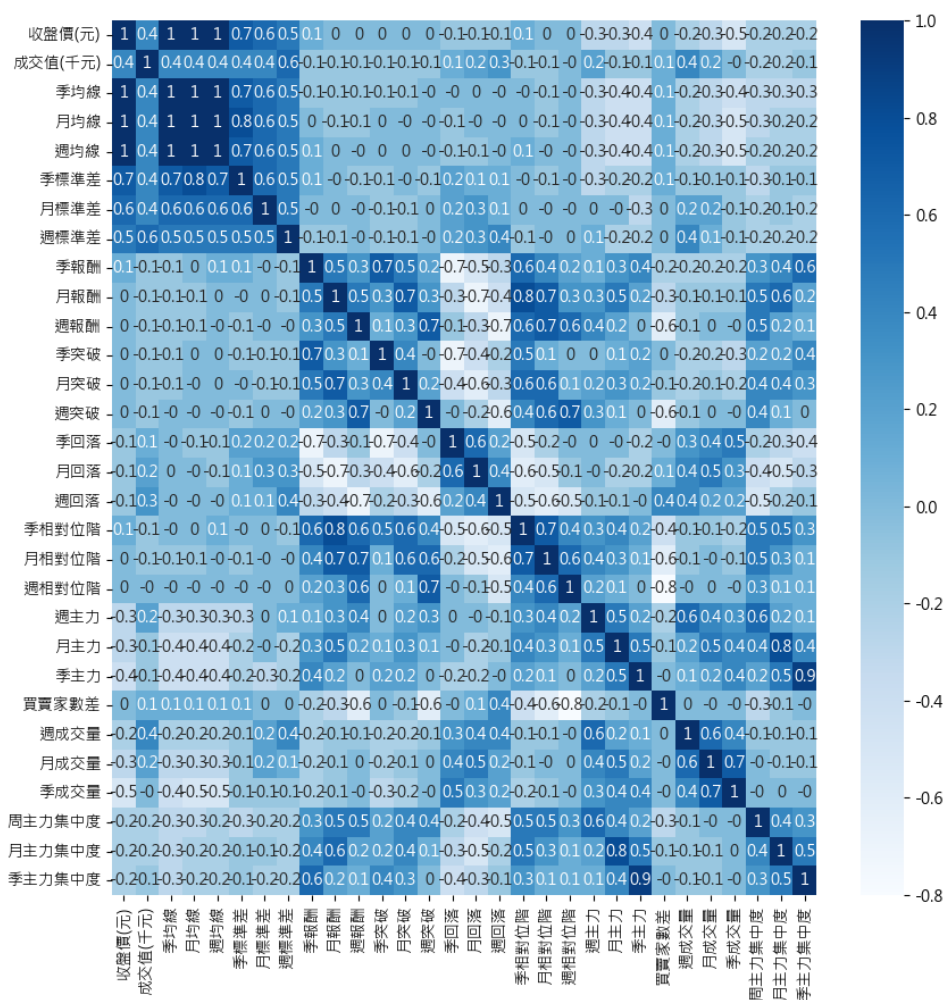


圖15. 台積電特徵重要性

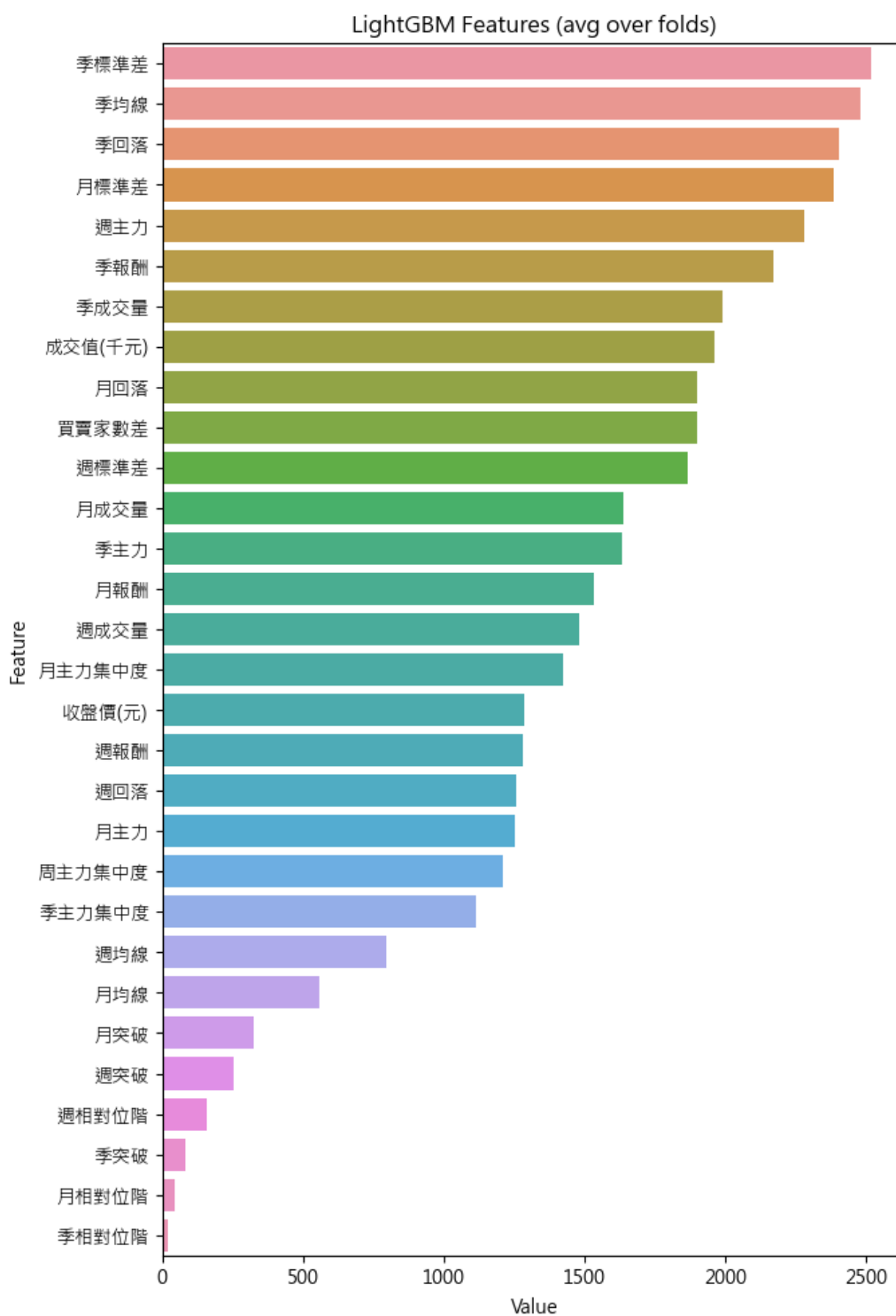


圖16. 重要特徵前 10 名出現排行

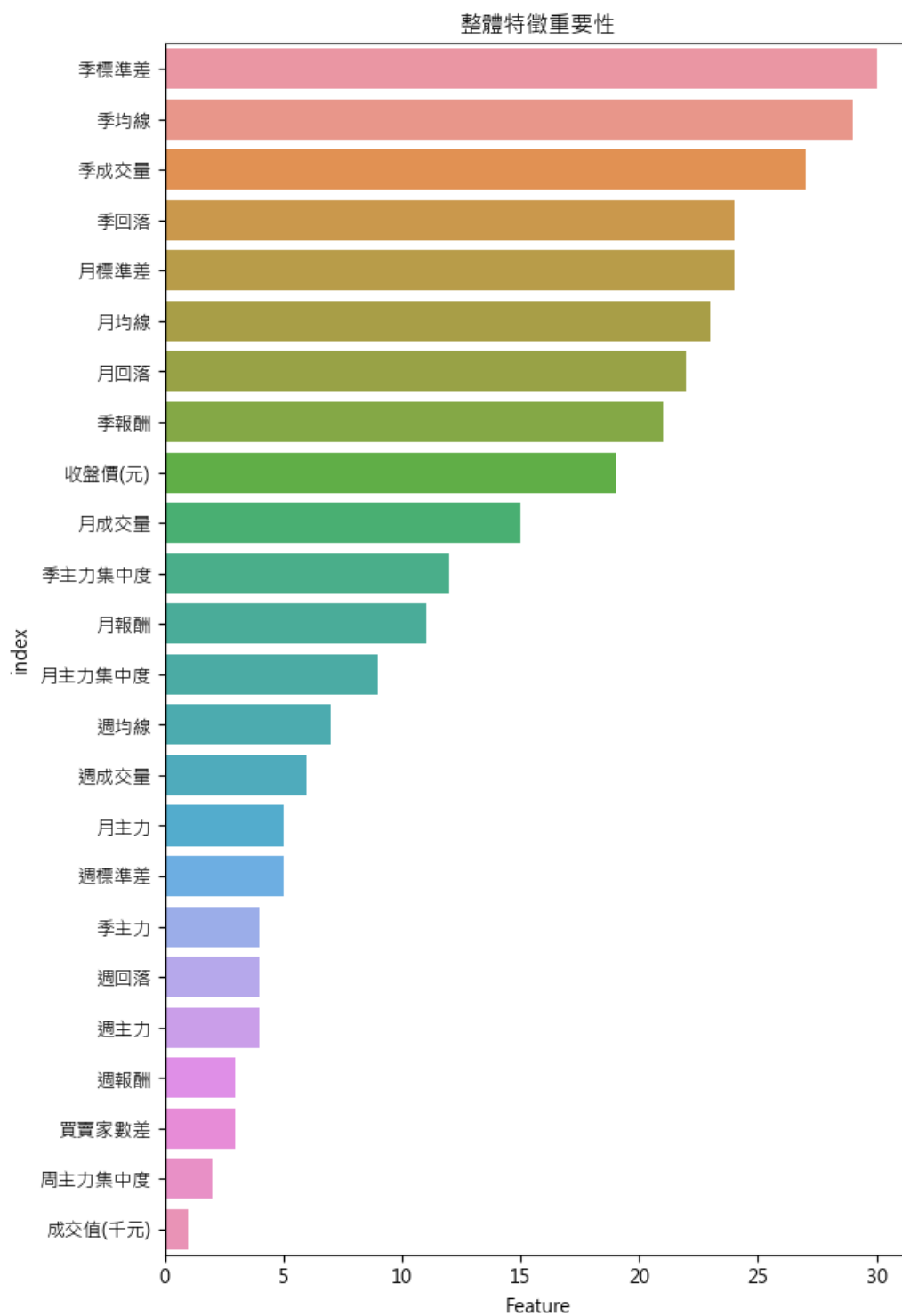


圖17. 台積電混沌矩陣圖

