



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 Thesis

籌碼面分析及機器學習於台股預測之應用

The Application of Chip Analysis and Machine Learning on
Predicting Taiwan Stock Market

研究生：劉淵博

Yuan-Po Liu

指導教授：王昭文 博士

Dr. Chou-Wen Wang

中華民國 110 年 6 月

June 2021

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班

研究生劉淵博（學號：M084030049）所提論文

籌碼面分析及機器學習於台股預測之應用
The Application of Chip Analysis and Machine Learning on predicting
Taiwan stock market

於中華民國 110 年 6 月 21 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 陳勤明 陳勤明

委員 王昭文 王昭文

委員 吳錦文 吳錦文

委員 ☒ 黃振聰 黃振聰

委員 蘇玄啟 蘇玄啟

委員 _____

指導教授(王昭文) 王昭文 (簽名)

摘要

本研究欲探討籌碼面資料以及機器學習模型對於台股股價之預測能力，利用台灣經濟新報(TEJ)籌碼面資料及上市普通股之股價資料，並使用 XGBoost、LightGBM 和 CatBoost 等三種機器學習模型進行預測。預測方法為將訓練期資料以機器學習模型搭配所有特徵、外資特徵、投信和自營商等四種特徵組合並以 Rolling-window 方式訓練後預測測試期每檔股票 20 日後、10 日後和 5 日後之每日股價報酬率和漲跌，以 RMSE 評估不同特徵組合下之預測誤差以及 F1-score 評估預測漲跌之結果。預測每日股價和漲跌後，再選取預測報酬率較高之標的並定期更換標的的建立交易策略，比較不同特徵搭配模型下各種組合之績效，進一步找出三種模型搭配四種特徵組合中表現最佳之組合。結果顯示，每月進場策略和每週進場策略表現最好之組合皆為投信特徵組合，於各組合中各檔數整體之報酬率皆為最高，而此現象在每週進場策略中更為顯著。每週進場策略和每月進場策略相比，每週進場表現較好，顯示較積極之更換投資標的會帶來較高報酬。預測報酬誤差和漲跌準確度方面，投信特徵亦為各組合表現最佳。比較各模型之表現，LightGBM 模型於預測報酬率誤差和預測準確率皆為三模型中表現最佳，且預測報酬率誤差中 LightGBM 模型和投信特徵組合之預測誤差最低，因此推斷 LightGBM 模型搭配投信特徵組合為本文中表現最佳之組合。

關鍵字：三大法人、股價預測、機器學習、籌碼面分析、交易策略

Abstract

This paper focus on investigating impact of chip analysis and machine learning models on predicting stock prices in Taiwan, using chip data and listed stock data from Taiwan Economic Journal (TEJ), and predicting stock prices using three machine learning models including XGBoost, LightGBM and CatBoost. The predicting method starts with combining four feature combinations including all features, foreign investors, investment trust and dealer with machine learning models and using the rolling-window method, then predicting daily return rate of each stock and rise or fall during the test period of 5,10 and 20 days later. Subsequently, evaluating the prediction error by RMSE under different feature combinations as well as F1-score to evaluate the result of the forecasting rise or fall. Finally, select stocks with a highest predicted rate of return and regularly change the portfolio to establish a trading strategy, and compare the performance under different combinations and find the best one. The results show that the investment trust feature performs the best in both monthly-entry strategy and weekly-entry strategy, by which the overall returns in various number of stocks are the highest, and this phenomenon is more significant in the weekly entry strategy. Compared to monthly entry strategy, the weekly strategy performs better, indicating that more active replacement brings higher returns. In terms of prediction error and accuracy, the investment trust feature is also the best among each combination. Comparing the performance of each model, LightGBM model has the best performance among the three models in terms of forecast return error and forecast accuracy, and the combination of LightGBM model and investment feature has the lowest prediction error. Therefore, it is inferred that LightGBM model combined with investment trust feature combination is the best combination in this article.

Key words: Institutional Investors, Stock Price Prediction, Machine Learning, Chip Analysis, Trading Strategy

目錄

| | |
|---------------------|------|
| 論文審定書 | i |
| 摘要 | ii |
| ABSTRACT | iii |
| 圖次 | vi |
| 表次 | viii |
| 第一章、緒論 | 1 |
| 第一節、研究動機 | 1 |
| 第二節、研究目的 | 1 |
| 第三節、研究架構 | 2 |
| 第二章、文獻回顧 | 3 |
| 第一節、三大法人交易策略 | 3 |
| 第二節、機器學習模型 | 5 |
| 第三節、小結 | 6 |
| 第三章、研究方法 | 7 |
| 第一節、資料簡介 | 7 |
| 第二節、研究流程 | 9 |
| 第三節、模型介紹 | 10 |
| 第四節、特徵選取與預測方法 | 12 |
| 第五節、交易策略 | 14 |
| 第四章、實證結果 | 16 |
| 第一節、敘述統計 | 16 |

| | |
|-------------------------------|----|
| 第二節、各模型和特徵組合之預測誤差和預測漲跌結果..... | 19 |
| 第三節、交易策略績效 | 24 |
| 第五章、結論與建議 | 45 |
| 參考文獻 | 46 |

圖次

| | |
|---|----|
| 圖 1、研究架構圖 | 2 |
| 圖 2、研究流程圖 | 10 |
| 圖 3、移動窗格示意圖 | 13 |
| 圖 4、交易策略流程圖 | 15 |
| 圖 5、XGBoost 模型搭配所有特徵每月更換標的策略累積報酬率圖 | 26 |
| 圖 6、XGBoost 模型搭配外資特徵每月更換標的策略累積報酬率圖 | 26 |
| 圖 7、XGBoost 模型搭配投信特徵每月更換標的策略累積報酬率圖 | 26 |
| 圖 8、XGBoost 模型搭配自營商特徵每月更換標的策略累積報酬率圖 | 26 |
| 圖 9、LightGBM 模型搭配所有特徵每月更換標的策略累積報酬率圖 | 29 |
| 圖 10、LightGBM 模型搭配外資特徵每月更換標的策略累積報酬率圖 | 29 |
| 圖 11、LightGBM 模型搭配投信特徵每月更換標的策略累積報酬率圖 | 29 |
| 圖 12、LightGBM 模型搭配自營商特徵每月更換標的策略累積報酬率圖 | 29 |
| 圖 13、CatBoost 模型搭配所有特徵每月更換標的策略累積報酬率圖 | 32 |
| 圖 14、CatBoost 模型搭配外資特徵每月更換標的策略累積報酬率圖 | 32 |
| 圖 15、CatBoost 模型搭配投信特徵每月更換標的策略累積報酬率圖 | 32 |
| 圖 16、CatBoost 模型搭配自營商特徵每月更換標的策略累積報酬率圖 | 32 |
| 圖 17、XGBoost 模型搭配所有特徵每週更換標的策略累積報酬率圖 | 36 |
| 圖 18、XGBoost 模型搭配外資特徵每週更換標的策略累積報酬率圖 | 36 |
| 圖 19、XGBoost 模型搭配投信特徵每週更換標的策略累積報酬率圖 | 36 |
| 圖 20、XGBoost 模型搭配自營商特徵每週更換標的策略累積報酬率圖 | 36 |
| 圖 21、LightGBM 模型搭配所有特徵每週更換標的策略累積報酬率圖 | 39 |
| 圖 22、LightGBM 模型搭配外資特徵每週更換標的策略累積報酬率圖 | 39 |
| 圖 23、LightGBM 模型搭配投信特徵每週更換標的策略累積報酬率圖 | 39 |
| 圖 24、LightGBM 模型搭配自營商特徵每週更換標的策略累積報酬率圖 | 39 |

| | |
|---|----|
| 圖 25、CatBoost 模型搭配所有特徵每週更換標的策略累積報酬率圖 | 42 |
| 圖 26、CatBoost 模型搭配外資特徵每週更換標的策略累積報酬率圖 | 42 |
| 圖 27、CatBoost 模型搭配投信特徵每週更換標的策略累積報酬率圖 | 42 |
| 圖 28、CatBoost 模型搭配自營商特徵每週更換標的策略累積報酬率圖 | 42 |

表次

| | |
|---|----|
| 表 1、資料欄位說明..... | 8 |
| 表 2、新增變數一覽..... | 9 |
| 表 3、特徵組合一覽..... | 14 |
| 表 4、各年度三大法人買賣超金額..... | 16 |
| 表 5、三大法人各年度買進金額佔成交量比重..... | 17 |
| 表 6、三大法人各年度賣出金額佔成交量比重..... | 18 |
| 表 7、各模型和特徵組合之預測 20 日後報酬率誤差比較..... | 19 |
| 表 8、各模型和特徵組合之預測 20 日後漲跌 F1-score 結果..... | 19 |
| 表 9、各模型和特徵組合之預測 10 日後報酬率誤差比較..... | 20 |
| 表 10、各模型和特徵組合之預測 10 日後漲跌 F1-score 結果..... | 20 |
| 表 11、各模型和特徵組合之預測 5 日後報酬率誤差比較..... | 21 |
| 表 12、各模型和特徵組合之預測 5 日後漲跌 F1-score 結果..... | 21 |
| 表 13、所有特徵組合之重要特徵..... | 22 |
| 表 14、外資特徵組合之重要特徵..... | 23 |
| 表 15、投信特徵組合之重要特徵..... | 23 |
| 表 16、自營商特徵組合之重要特徵..... | 24 |
| 表 17、XGBoost 模型搭配各特徵組合每月更換標的策略績效..... | 25 |
| 表 18、XGBoost 模型每月更換標的策略各組合逐年報酬率..... | 27 |
| 表 19、LightGBM 模型搭配各特徵組合每月更換標的策略績效..... | 28 |
| 表 20、LightGBM 模型每月更換標的策略各組合逐年報酬率..... | 30 |
| 表 21、CatBoost 模型搭配各特徵組合每月更換標的策略績效..... | 31 |
| 表 22、CatBoost 模型每月更換標的策略各組合逐年報酬率..... | 33 |
| 表 23、XGBoost 模型搭配各特徵每週更換標的策略績效..... | 35 |
| 表 24、XGBoost 模型每週更換標的策略各組合逐年報酬率..... | 37 |

| | |
|--|----|
| 表 25、LightGBM 模型搭配各特徵組合每週更換標的策略績效..... | 38 |
| 表 26、LightGBM 模型每週更換標的策略各組合逐年報酬率 | 40 |
| 表 27、CatBoost 模型搭配各特徵組合每週更換標的策略績效..... | 41 |
| 表 28、CatBoost 模型每週更換標的策略各組合逐年報酬率 | 43 |

第一章、緒論

第一節、研究動機

台灣股市中，三大法人長期為市場中主力，成交量佔整體成交量比重有一定比例，為影響股價之重要關鍵。由於三大法人資本及成交金額龐大，其連續買超或連續賣超之行為經常造成股價之大幅上漲或下跌，抑或當三大法人於某一交易日大量買入或賣出某檔股票，足以造成當日股價達到 10% 之漲跌幅限制等劇烈波動。莊家睿(2017)透過 VBA 實證研究證實三大法人之公開資訊對大盤指數有顯著的解釋力，並且以外資交易策略獲利最高。由於證交所每交易日於盤後公布三大法人之買賣超資訊，使得台股市場籌碼面資料有公開透明特性，不少投資人亦透過籌碼面指標作為選股時之判斷依據。

機器學習中的 GBDT(Gradient boosting decision trees)方法模型已被廣泛應用於學術領域(Andreea Anghel, 2019)，其中財務領域之應用包含預測信用違約、股價報酬率等皆有不少相關之研究。相較於其他模型和方法，由於其預測準確率高和特徵處理佳等優勢，使得機器學習模型不論是資料分析競賽或是研究上相當熱門。

第二節、研究目的

綜上所述，本研究使用機器學習模型中近年最熱門且經常被做為比較和資料競賽之模型，即 XGBoost、LightGBM 和 CatBoost 等三種機器學習模型作為訓練之模型，並透過籌碼面指標作為特徵進行訓練和預測，預測股價報酬率以及股票漲跌並選取預測報酬較高者作為投資標的，以比較三大法人各特徵組合搭配各模型之預測報酬率之誤差以及預測漲跌準確度，以及檢視特徵重要度，最後檢視各模型和特徵組合選取標的之投資績效，目的為比較各組合之預測報酬誤差及漲跌準確度和

選取標的之投組績效，並探討預測報酬率誤差最低、預測漲跌準確度最高和選取標的的績效最好之模型和特徵組合，進一步得知三種模型和四種特徵組合搭配訓練之十二種組合中表現最佳之組合。本研究欲探討目的分為以下幾點：

- (一) 比較模型和特徵形成組合之預測報酬率誤差和預測漲跌
- (二) 比較各組合選取標的之交易績效
- (三) 結合以上結果，找出三種模型和四種特徵組合搭配訓練之組合中，表現最佳之組合

第三節、研究架構

本文研究架構共分為五部分，如圖 1 所示。第一章為緒論，主要為研究動機與目的；第二章為文獻回顧，探討三大法人和機器學習相關之文獻；第三章為研究方法，內容包含本研究使用之資料、模型建立方法；第四章為實證結果，為各模型搭配預測報酬率和交易策略之結果呈現；第五章為結論和建議。

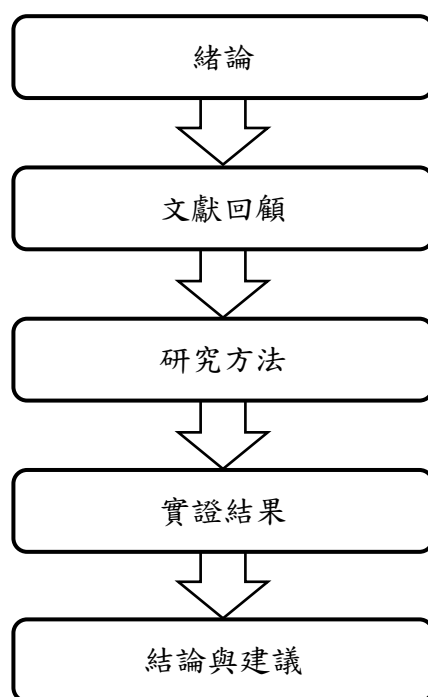


圖 1、研究架構圖

第二章、文獻回顧

第一節、三大法人交易策略

由於證交所每日公布三大法人之買賣超資訊，使得主力進出場之相關資料相對其他市場為公開透明，因此不少研究利用籌碼面指標建立相關之交易策略或進行預測，選擇權未平倉和期貨未平倉量等指標也作為預測台灣加權指數之資料(郭裕涼，2013)。

Hongwei Chuang(2015)利用橫斷面迴歸，探討 2001 至 2014 年期間法人短期和長期持股量變化對於投資組合報酬率之影響，先分別根據過去 3 個月(短期)和過去 12 個月(長期)之外資、投信和自營商持股量變化由高到低排序並分為十群，再探討十個分群之股票於持有數個月後之報酬率。該研究之結果顯示，短期策略中，僅外資法人持股策略分群有顯著正報酬；長期策略中，根據外資、投信和自營商之持股分群皆能產生正報酬。最後再將外資、投信和自營商之投資組合和大盤以及動能策略等五項指數和策略於 2001 至 2014 年期間作為比較，結果顯示外資和投信之投資組合約可產生 4.2 倍之報酬，約為大盤之三倍；動能策略報酬率較大盤為低，而自營商投資組合報酬率最低。

范聖培(2014)探討台灣股市 2001 至 2013 年間三大法人之買賣超行為對於短期股價之影響。該文獻選取三大法人於集中和店頭市場前一週成交量或買賣超金額前五和前十之股票作為投資標的，觀察短期之股價報酬。研究之結果顯示，三大法人之買賣超行為確實造成股價之正(負)報酬，而以投信最顯著；不論是投信買超前五檔或十檔標的皆有 6%左右之正報酬，而投信賣超行為也觀察到-3 至-4%之負報酬；而外資法人之買賣超行為雖然也有顯著正負報酬，不過較投信之報酬略低。

陳冠宇(2020)透過追蹤資料(Panel Data)迴歸模型證實三大法人買賣超對高價股之報酬有顯著正相關，股價最高之 50 檔中，有 42 檔之外資買賣超和報酬正相關，41 檔投信買賣超和報酬正相關，顯示法人買賣超對高價股報酬有解釋力。

莊家睿(2017)利用 VBA 分析三大法人之公開資訊對於台灣股價指數之影響力，主要方法為：以台指期三大法人多空淨額交易口數和未平倉口數建立交易策略。結果顯示，自營商以未平倉多空淨額比例交易策略表現最好，投信則是未平倉淨額和多空口數為交易策略之效果相同，而外資不論任何指標建立之交易策略皆可產生高報酬，因此推論外資為三大法人中表現較好者；而作者也從結果中推論出幾項重要指標：包含多空淨額比例 10%與否、以及多空淨額是否連續兩天突破皆為相當有效的指標。

陳楷恩(2020)以向量自我迴歸模型探討三大法人未平倉和買賣超等相關資訊和台股指數之關聯性，而該研究結果顯示兩者存在因果關係，惟市場出現恐慌時(例如 2008 年之金融海嘯和 2020 年之新冠肺炎)較難利用指標預測市場之走勢，但是於非危機時期時，三大法人指標具有一定程度可預測趨勢之能力。

黃英杰(2016)同樣利用三大法人未平倉量為進出場指標，分別操作三檔 ETF，分別為元大台灣 50、元大 MSCI 金融以及元大高股息等 ETF。投資期間為 2007 至 2016，具體作法為利用三大法人個別以及合計之期貨未平倉口數評估應持有之 ETF 部位。該研究結果顯示，利用外資未平倉量操作之交易策略表現最佳，於操作三檔皆有正報酬，至少有 7-13%之年化報酬率，而三大法人合計未平倉量表現次之，投信未平倉量交易策略僅優於自營商，大約有 0-5%之年化報酬，而自營商未平倉量交易策略表現為各組合之末。

第二節、機器學習模型

機器學習屬於人工智慧領域之分支，主要可分為監督式學習和非監督式學習，其應用相當廣泛。不論是醫學、工業、生物辨識等領域皆有相關研究，而本節主要探討比較本研究所使用之機器學習模型中三種梯度提升 GBDT 演算法之相關文獻，即 XGBoost、LightGBM 和 CatBoost 之相關文獻。

Essam Al Daoud(2019)比較 XGBoost、LightGBM 和 CatBoost 等三種梯度提升方法預測房屋信用貸款資料並以 AUC 評估準確率和檢視模型花費時間。房屋信貸資料共包含 219 個特徵和 356251 筆資料。資料經過刪減缺失值 75%以上或重要性較低等處理後，再經過 5-fold 交叉驗證；結果顯示，LightGBM 在三模型中訓練速度最快以及 AUC 最高，即預測最準確；而 CatBoost 之預測準確率為三者較低，XGBoost 模型則是訓練時間較久。但作者認為如果訓練特徵包含更多類別特徵，則 CatBoost 表現會優於其他兩者。該文獻預測標的為房貸信用違約，屬於分類問題，而本研究之預測股價漲跌同樣為二元分類問題，可作為參考。

Andreea Anghel(2019)探討 XGBoost、LightGBM 和 CatBoost 三種 GBDT 方法之模型效能。該文獻利用 Higgs 和 Epsilon 資料集做為二分類建模資料，以及 Yahoo 和 Microsoft 資料，為包含從網頁取得之特徵並包含相關分數(0-4 分，分別代表完全不相關到完全相關)，即屬於多類別之分類問題。結果顯示，XGBoost 模型在多類別之分類問題表現最好，test score 為三種模型中最高；而 LightGBM 模型和 CatBoost 模型則分別在 Higgs 和 Epsilon 等二分類問題表現最好，test score 分別為 0.8573 和 0.9537。

Sun Xiaolei, Liu Mingxia, Sima Zeqian (2020)利用 LightGBM 模型以預測加密貨幣市場的價格趨勢並和 SVM 和隨機森林做比較，該文獻根據交易量，市場排名、流通量和數據可用性，選擇了 42 種主要加密貨幣並選擇了道瓊指數，美國 S&P 500 指數，恆生指數，美元指數期貨，上證綜合指數等眾多總經指標。結果顯示，各方

法在預測 2 週後加密貨幣價格較 2 日後為準確，而 LightGBM 模型在穩健度之表現顯著優於其他兩方法，顯示 LightGBM 模型在處理大量資料集上較有效。

Mathias Kinnander(2020)利用 XGBoost、LightGBM、CatBoost 等三種機器學習方法以及隨機森林和羅吉斯迴歸等方法，透過顧客消費預未來之獲利能力，該研究屬於二分類問題，主要將顧客之獲利能力分為兩群。該研究結果顯示 CatBoost 模型和隨機森林之分類能力最佳，雖各模型對於有獲利能力之顧客皆有高準確預測能力，但區分無獲利能力之顧客中，XGBoost 和 LightGBM 則相較之下表現不佳，再次顯示 CatBoost 於二分類問題中表現最佳。

Sami Ben Jabeur , Cheima Gharib , Salma Mefteh-Wali , Wissal Ben Arfi(2021)利用 CatBoost 以及其他 8 種方法(包含隨機森林、XGBoost、神經網路等)預測法國企業在 2014 年至 2016 年財務危機下之倒閉風險，分別預測一年、兩年和三年後之倒閉情況，同樣屬於二分類問題。結果顯示，CatBoost 預測能力最佳，平均預測準確率高達 82.90%，而 XGBoost 預測準確為第二高，準確率為 81.62%，顯示 XGBoost 和 CatBoost 兩大 GBDT 方法於二分類問題預測之表現優於其他方法。

第三節、小結

綜合上述文獻，本段整理前兩節之文獻探討，進一步得知本文獻兩大主軸，即籌碼面指標以及機器學習模型相關文獻之結論。首先，第一節之多數文獻之結論皆顯示籌碼指標和股價報酬有高度正相關，以及利用三大法人相關指標建立之策略可創造相對應之正報酬；而第二節之文獻則說明本研究之三種機器學習模型於預測二分類和多分類之表現上皆有顯著優於其他方法之表現，亦為本研究選擇上述三種模型之動機。

第三章、研究方法

第一節、資料簡介

本研究資料使用台灣經濟新報(TEJ)提供之上市普通股每日籌碼面及報酬率資料，資料期間為 2008/1/1 至 2020/12/31。詳細資料特徵及定義如表 1 所示，包含外資、投信、自營商、三大法人合計之每日買賣超張數、連續累計買賣超張數、連續買賣超日數、持股數、持股市值和持股率等；買賣超張數為交易所公告之每檔股票當日買進張數與賣出張數之差值；買賣超日數則為對於該證券或該產業至當日為止，連續買賣超的日數。舉例而言，當外資對於台積電(2330)該股票連續五日為買超，則外資買賣超欄位中該數值為+5；反之當外資連續五日賣超，則數值為-5；連續累計買賣超表示從該次連續買超或賣超之第一天算起之累計買超或賣超之張數。買賣超區間報酬代表法人連續買超或連續賣超期間之累積報酬率，例如：若鴻海(2317)於 2009/10/13「外資買賣超日數」為-4，則將當日往回推算，合計 4 個交易日的日報酬率累乘之。買賣超金額則說明對於該證券至當日為止，連續買超或賣超的累計金額，以千元計。

表 1、資料欄位說明

| 欄位 | 說明 | 分類 |
|-----------------------|--|-------------|
| 外資/投信/自營商/合計買賣超張數 | 交易所公告之外資/投信/自營商/三大法人合計買進張數與賣出張數之差值 | 買賣超張數，單位為張 |
| 信用交易買賣超張數 | 交易所公告之當日之融資融券買超張數和融資融券賣超張數之差值 | |
| 外資/投信/自營商/法人連續累計買賣超張數 | 對於該證券或該產業至當日為止，連續買賣超期間之累計買賣超張數，舉例來說，若 2009/10/15「外資/投信/自營商/法人買賣超日數」為-5，則此處將 2009/10/9~2009/10/15（當日往回推算，合計 5 個交易日）之「外資買賣超」數值加總。 | |
| 資券連續累計買賣超張數 | 對於該證券或該產業至當日為止，連續買賣超期間之使用信用交易方式累計買賣超張數 | |
| 外資/投信/自營/法人買賣超日數 | 「外資/投信/自營/合計買賣超(張)」欄位至當日為止，連續為正值或負值的次數連續買賣超日數 | 買賣超天數，單位為日 |
| 資券買賣超日數 | 採信用交易對於該股連續買超或連續賣超日數 | |
| 外資/投信/自營/法人連續累計買賣超金額 | 此欄位需搭配「外資/投信/自營/法人資券買賣超日數」使用，代表對於該證券或該產業至當日為止連續買超或賣超累計金額，舉例來說，若 2009/10/15「現股買賣超日數」為-5，則此處將 2009/10/9~2009/10/15（當日往回推算，合計 5 個交易日）之金額加總。 | 買賣超金額，單位為千元 |
| 資券連續累計買賣超金額 | 對於該證券或該產業至當日為止，使用信用交易方式進出個股連續買超或連續賣超累計金額 | |
| 外資/投信/自營/法人持股數 | 對於該證券或該產業至當日為止，三大法人持有張數 | 張 |
| 外資/投信/自營/法人持股數市值 | 對於該證券或該產業至當日為止，外資/投信/自營/三大法人合計持有股票之市值 | 百萬 |
| 外資/投信/自營/法人持股率 | 外資/投信/自營/三大法人持股數市值持有股數占流通在外股數之比例 | % |

資料來源：TEJ 台灣經濟新報

除上述特徵外，本研究亦結合技術指標新創變數，另創特徵如表 2 所示：

表 2、新增變數一覽

| 分類 | 特徵 |
|-----|------------------|
| 外資 | 外資買賣超張數/5 日均量 |
| | 外資買賣超張數/20 日均量 |
| | 外資連續累計買賣超金額/成交量 |
| | 外資連續買賣超張數/成交量 |
| 投信 | 投信買賣超張數/5 日均量 |
| | 投信買賣超張數/20 日均量 |
| | 投信連續累計買賣超金額/成交量 |
| | 投信連續買賣超張數/成交量 |
| 自營商 | 自營商買賣超張數/5 日均量 |
| | 自營商買賣超張數/20 日均量 |
| | 自營商連續累計買賣超金額/成交量 |
| | 自營商連續買賣超張數/成交量 |

第二節、研究流程

本文之研究流程分為以下步驟，如圖 2 所示。首先，建立 XGBoost、LightGBM 以及 CatBoost 等三種機器學習模型，再將資料中特徵分為外資特徵組合、投信特徵組合、自營商特徵組合和所有特徵等四種不同特徵組合，再分別搭配上述三種模型形成 12 種組合並預測 5 日後、10 日後和 20 日後之每日報酬率和漲跌，並以 RMSE 比較預測誤差和檢視各模型特徵重要度，最後再分別利用每週和每月之頻率選取預測報酬率最高者建立固定頻率更換投資標的之交易策略。

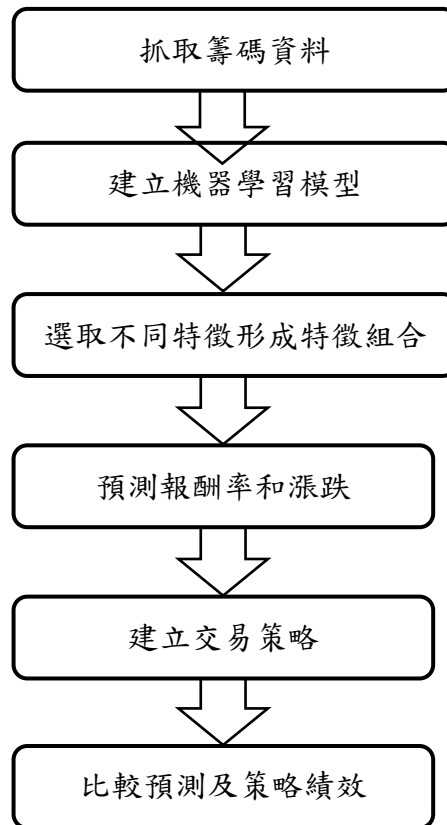


圖 2、研究流程圖

第三節、模型介紹

本研究之模型採用三種機器學習模型：XGBoost、LightGBM 以及 CatBoost 預測報酬率。三種方法皆為 GBDT 之演算法，GBDT (Gradient Boosting Decision Tree) 主要思想是利用弱分類器（決策樹）進行迭代訓練以得到最佳化模型，具有訓練效果好、不容易過度擬合等優點。而上述三種模型也經常作為資料分析競賽中使用之訓練模型，不少文章比較三種模型之表現，甚至稱之為資料分析之王，可見模型之強度；由於模型預測之準確度為眾多模型中表現最好，也具有訓練速度快等特性，因此使得上述三種模型成為近期最熱門之模型。以下分別介紹上述之模型：

3.1 XGBoost

XGBoost 之全名為 eXtreme Gradient Boosting，為 Chen and Guestrin 於 2016 年提出的演算法，為一種極限梯度提升之 Boosting 方法之監督式機器學習模型；Boosting 方法之概念為集合多個弱分類器而形成一個強分類器；而 XGBoost 模型為將多個分類和迴歸樹(CART)集合而成的提升樹模型。XGBoost 演算法的核心理念為：不斷地進行特徵分裂以生成新的樹，每次添加一棵樹即代表模型學習一個新函數，擬合前一次預測之殘差。訓練完成得到 k 棵樹後，根據樣本特徵，在每棵樹中會分配到對應的一個葉子節點，而每個葉子節點即對應一個分數，再將每個節點的分數相加即得到預測值。XGBoost 具有正則化之優點，使模型不會過於複雜能夠避免過度擬合。XGBoost 亦有高度靈活性、內建交叉驗證等幾點優點，缺點則是在訓練時會占用較多記憶體以及相較其他模型耗費較多之訓練時間。

3.2 LightGBM

LightGBM，全名為 Light Gradient Boosting Machine，由微軟公司所開發，是一種實現 GBDT 演算法的輕量模型，能夠支援高效率的並行訓練；LightGBM 和其他模型不同之處為其他模型使用 level-wise 之分裂方法，而 LightGBM 模型係採用加入了單邊梯度採樣(Gradient-based One-Side Sampling ,GOSS)之 leaf-wise 分裂方法，即代表每當 leaf-wise 代表每次之訓練自全部葉子中，模型會找出分裂增益最大者進行再分裂以形成迴圈。Leaf-wise 方法和 level-wise 相比，優點除了在預測上更為精確，降低誤差以外，亦提升訓練效率；缺點則是可能產生過度擬合、over-fitting 等問題，因此 LightGBM 模型設置限制以防止過度擬合，亦能確保預測準確，另外在訓練模型若資料太少容易產生過度擬合之問題。

3.3 CatBoost

CatBoost 全名為 Categorical Boosting，為俄羅斯網路企業 Yandex 於 2017 年所提出的機器學習方法，同樣為極限提升之 Boosting 方法。CatBoost 以對稱決策樹 (oblivious trees) 為基學習器，和其他模型相比其新創之算法能夠解決以往類別變數之處理問題，作法為將計算類別特徵出現的頻率加上超參數以形成新的數值型特徵。CatBoost 和 LightGBM 相同之處為兩模型皆支援類別特徵之輸入，模型則會對類別特徵進行 One-hot encoding 之處理，相較之下 XGBoost 則僅能夠處理數值特徵。CatBoost 最大優點為對於類別特徵和數值特徵皆能有效的處理，另外也透過超參數優化克服了梯度偏差 (Gradient Bias) 和預測偏移 (Prediction Shift)，進而避免過度擬合之問題，以提升模型效能。

第四節、特徵選取與預測方法

本研究將資料分為訓練期和測試期，訓練方式如圖 3 所示，訓練期自 2008 年開始，以移動窗格方式訓練之，以八年訓練期資料長度訓練並預測一年之測試期報酬，測試期間為 2016 至 2020 年共五年；並將每組合取七成樣本訓練，重複訓練五次後再預測每日每檔股票之報酬率，將五次之結果取平均得到最後預測報酬率，漲跌部分則是將五次預測漲跌結果投票；再預測測試期間 20 日後、10 日後和 5 日後之每日報酬率以及每日漲跌。(Ex：1 月 1 日時預測 1 月 21 日之 20 日後報酬率以及 1 月 6 日之 5 日後報酬率，以此類推。) 預測誤差以每日報酬率和當日之預測報酬率之間 RMSE(均方根誤差)衡量，而每日漲跌之預測準確度則以預測期間之預測漲跌和實際漲跌情形比對之 F1-score 結果，F1-score 為結合 Precision 和 Recall，經常作為綜合評估二分類問題精確度之指標，可避免在不平衡資料指標不具參考性等問題；本文將股票上漲視為 1，下跌則視為 0，RMSE 和 F1-score 定義如下：

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (1)$$

其中 \hat{y}_i =每日預測報酬率、 y_i =每日實際報酬率 、 n =測試期資料日數

$$F - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2)$$

$$\text{其中 Precision} = \frac{TP}{TP + FP} ; \text{Recall} = \frac{TP}{TP + FN}$$

TP=預測值和實際值皆為 1，正確預測上漲；

FP=預測值為 1 但實際為 0，實際下跌卻預測為上漲；

TN=預測值和實際值皆為 0，正確預測下跌；

FN=預測值為 0 但實際為 1，實際上漲卻預測為下跌。

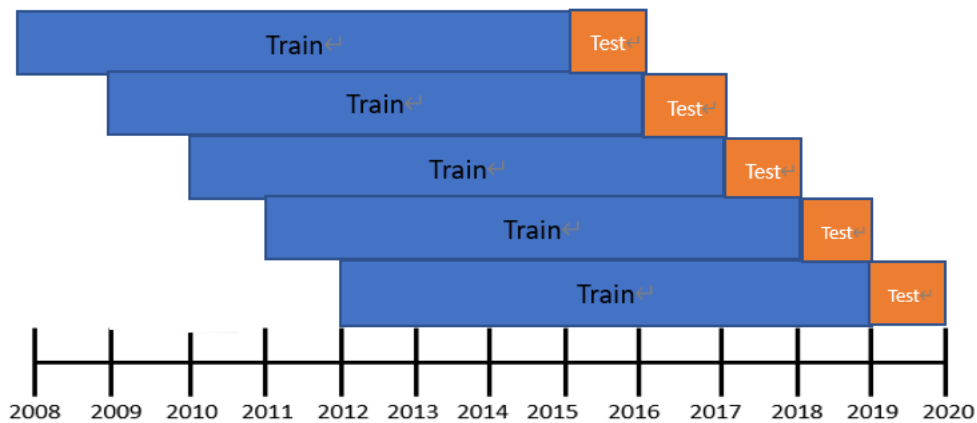


圖 3、移動窗格示意圖

由於不同特徵組合之訓練效果不盡相同，本研究透過建立不同特徵組合比較不同指標之預測誤差，並且透過訓練不同特徵，進一步得知外資、投信、自營商和三大法人中最具有影響力之族群。表 3 為本文模型訓練所建立之特徵組合，包含所有特徵、外資特徵組合、投信特徵組合以及自營商特徵組合等。

表 3、特徵組合一覽

| 特徵組合 | 特徵 |
|---------|---|
| 所有特徵 | 表 1 和表 3 中所有欄位，除三大法人特徵外亦包含融資融券等特徵。 |
| 外資特徵組合 | 外資買賣超張數、外資連續累計買賣超金額、外資買賣超日數、外資連續累計買賣超、外資持股數、外資持股數市值、外資持股率、外資連續累計買賣超金額/成交量、外資買賣超張數/5 日均量、外資買賣超張數/20 日均量、外資連續買賣超張數/成交量 |
| 投信特徵組合 | 投信買賣超張數、投信連續累計買賣超金額、投信買賣超日數、投信連續累計買賣超、投信持股數、投信持股數市值、投信持股率、投信連續累計買賣超金額/成交量、投信買賣超張數/5 日均量、投信買賣超張數/20 日均量、投信連續買賣超張數/成交量 |
| 自營商特徵組合 | 自營商買賣超張數、自營商連續累計買賣超金額、自營商買賣超日數、自營商連續累計買賣超、自營商持股數、自營商持股數市值、自營商持股率、自營商連續累計買賣超金額/成交量、自營商買賣超張數/5 日均量、自營商買賣超張數/20 日均量、自營商連續買賣超張數/成交量 |

第五節、交易策略

比較各特徵和模型組合之預測誤差和準確率後，本研究欲以各模型預測報酬率以選取投資標的建立交易策略，具體作法為將上述之各模型搭配特徵組合之測試期(2016-2020)各檔股票，在 t 期預測 $t+20$ 日之每日報酬率，以月分為單位取平均，並選出每月預測平均報酬率最高之 5、10、20、30、40 和 50 檔股票，再以等權重投資於每月初，每月更換投資標的，投資於 2016 至 2020 年期間，共 60 個月分，並檢視策略績效。本研究亦建立每周更換投資標的之策略，目的為探討較積極之進出場是否有更好之投資績效，作法和每月更換標的之策略相同，預測 $t+5$ 日之報酬率，惟操作頻率為每週更換標的，共 260 週。

圖 4 說明本研究交易策略建構流程。交易策略之績效則於不考慮交易成本下以累積報酬率、年化報酬率、年化標準差、夏普比率和最大回撤等指標評估，定義如式(3)-(7)。

$$\text{累積報酬率} = \prod (1 + \text{每日報酬率}) - 1 \quad (3)$$

$$\text{年化報酬率} = (1 + \text{累積報酬率})^{\frac{1}{\text{年數}}} - 1 \quad (4)$$

$$\text{年化標準差} = \text{每日報酬率標準差} \times \sqrt{252} \quad (5)$$

$$\text{夏普比率} = \text{年化報酬率} / \text{年化標準差} \quad (6)$$

$$\text{最大回撤} = \text{投資期間報酬率最高點至最低點滑落幅度} \quad (7)$$

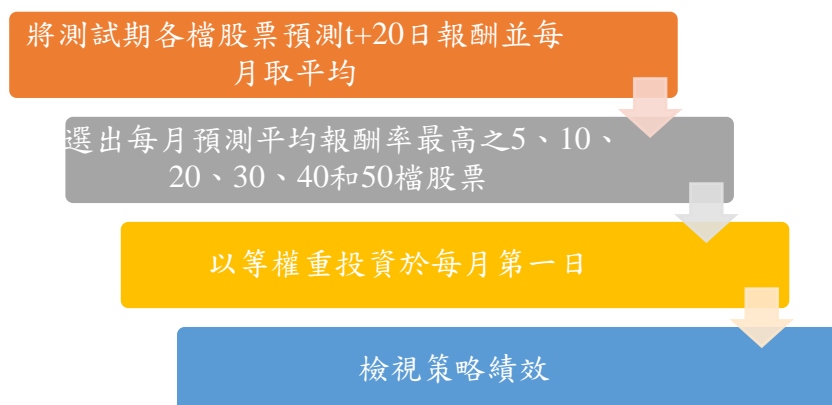


圖 4、交易策略流程圖

第四章、實證結果

第一節、敘述統計

為檢視三大法人金流流向分布，並觀察各年度大盤趨勢和三大法人買賣超金額之關聯，本節分別列出各年度三大法人買賣超金額以及買超和賣超分佔成交量之比例，期間為 2008-2020 年，單位為億元。表 4 顯示 2008 至 2020 年期間各年度中三大法人買賣超金額和指數漲跌百分比，結果顯示三大法人之買賣超金額和指數漲跌有正相關之關係，相關係數為 0.5678；惟 2020 年之指數漲跌和法人買賣超金額為負相關，推測原因可能為 2020 年 3 月股市達到最低點，而隨後反彈甚至創下新高，而跌至最低點前之買賣超金額相當龐大導致該年累計買賣超金額為負。

1.1 各年度買賣超金額

表 4、各年度三大法人買賣超金額

| 年度 | 外資 | 投信 | 自營商 | 三大法人合計 | 指數漲跌 |
|------|--------|--------|--------|--------|---------|
| 2008 | -4,700 | 441.7 | 433 | -3,825 | -46.03% |
| 2009 | 4,801 | -289.3 | 100.2 | 4,612 | 78.34% |
| 2010 | 2,877 | -678.4 | 28.18 | 2,226 | 9.58% |
| 2011 | -2,776 | 94.06 | -456.4 | -3,139 | -21.18% |
| 2012 | 1,442 | -160.6 | -22.41 | 1,259 | 8.87% |
| 2013 | 2,430 | -766.7 | 84.92 | 1,749 | 11.85% |
| 2014 | 3,546 | -381.2 | -3.615 | 3,162 | 8.08% |
| 2015 | 462.3 | -351.1 | -565.9 | -454.7 | -10.41% |
| 2016 | 3,202 | -199 | -866.4 | 2,137 | 10.98% |
| 2017 | 1,552 | -334.9 | 113.8 | 1,331 | 15.01% |
| 2018 | -3,551 | -15 | -994.1 | -4,558 | -8.6% |
| 2019 | 2,442 | 112.8 | 524.8 | 3,079 | 23.33% |
| 2020 | -5,397 | 426.7 | -1,696 | -6,666 | 22.8% |

資料來源：GoodInfo 台灣股市資訊網

1.2 三大法人買進金額占成交量比重

本節為三大法人買進金額占成交量比重之敘述統計量結果，表 5 為 2008 至 2020 年期間各年度三大法人買進金額以及買進金額佔成交量比重，結果顯示三大法人合計買進佔比於 2012 年前為 20-30%左右，2013 年後為 30-40%而買進最低年份為 2009 和 2010 年，大約為 22-24%左右，買進佔比最高為 2015-2016，佔比為 37%-39%；其中外資買進佔成交量比重大約為 20-30%，於 2016 年最高，為 31.30%；外資買進金額佔比最低為 2009 年，僅 17.20%；投信買進金額佔比最低，僅 1-3%；而自營商買進比重則是 3-7%左右，比重最高為 2015 年之 7.65%，而買進金額佔成交比重最低亦為 2009 年，為 2.32%。

表 5、三大法人各年度買進金額佔成交量比重

| 年度 | 成交量 | 外資 | | 投信 | | 自營商 | | 三大法人合計 | |
|------|------------|------------|-----------|------------|-----------|------------|-----------|------------|-----------|
| | (億元) | 買進 (億元) | 佔成交 比重 | 買進 (億元) | 佔成交 比重 | 買進 (億元) | 佔成交 比重 | 買進 (億元) | 佔成交 比重 |
| 2008 | 266,644.92 | 62,470 | 23.40% | 9,214 | 3.46% | 8,418 | 3.16% | 80,101 | 30% |
| 2009 | 301,181.19 | 51,658 | 17.20% | 7,954 | 2.64% | 6,977 | 2.32% | 66,589 | 22.10% |
| 2010 | 288,855.24 | 54,951 | 19% | 6,480 | 2.24% | 8,780 | 3.04% | 70,211 | 24.30% |
| 2011 | 269,957.04 | 57,500 | 21.30% | 6,792 | 2.52% | 9,778 | 3.62% | 74,069 | 27.40% |
| 2012 | 207,885.24 | 47,733 | 23% | 5,209 | 2.51% | 6,868 | 3.30% | 59,810 | 28.80% |
| 2013 | 196,033.28 | 49,573 | 25.30% | 3,755 | 1.92% | 9,025 | 4.60% | 62,353 | 31.80% |
| 2014 | 230,432.84 | 56,602 | 24.60% | 3,457 | 1.50% | 14,438 | 6.27% | 74,497 | 32.30% |
| 2015 | 225,050.96 | 64,103 | 28.50% | 3,103 | 1.38% | 17,207 | 7.65% | 84,413 | 37.50% |
| 2016 | 189,156.19 | 59,201 | 31.30% | 2,474 | 1.31% | 13,304 | 7.03% | 74,979 | 39.60% |
| 2017 | 257,988.82 | 67,692 | 26.20% | 2,755 | 1.07% | 14,829 | 5.75% | 85,276 | 33.10% |
| 2018 | 321,623.76 | 82,840 | 25.80% | 3,834 | 1.19% | 17,245 | 5.36% | 103,523 | 32.20% |
| 2019 | 290,563.19 | 81,602 | 28.10% | 3,825 | 1.32% | 15,553 | 5.35% | 100,963 | 34.70% |
| 2020 | 491,825.26 | 128,580 | 26.10% | 6,387 | 1.30% | 18,700 | 3.80% | 153,641 | 31.20% |

資料來源：GoodInfo 台灣股市資訊網

1.3 三大法人賣出金額占成交量比重

本節為三大法人賣出金額占成交量比重之敘述統計量結果，表 6 為 2008 至 2020 年期間各年度三大法人買進金額以及賣出金額佔成交量比重，結果顯示三大法人合計賣超金額 2012 年前為 30%-40% 左右，而 2013 年後為 20-30%，顯示法人買進比例上升賣出比例下降，買超有上升的趨勢；其中賣出比重最高為 2011 年，佔比為 38.50%，而賣出比重佔比最低為 2017 年之 23.50%；外資賣超金額佔比大約為 20-30%，其中最高為 2011 年之 29.60%，最低為 2018 年之 15.60%。投信賣出金額佔比為 1-3% 左右，而自營商賣出金額佔比為 3-7%。

表 6、三大法人各年度賣出金額佔成交量比重

| 年度 | 成交量 (億元) | 外資 | | 投信 | | 自營商 | | 三大法人合計 | |
|------|-------------|------------|-----------|------------|-----------|------------|-----------|------------|-----------|
| | | 賣出 (億元) | 佔成交 比重 | 賣出 (億元) | 佔成交 比重 | 賣出 (億元) | 佔成交 比重 | 賣出 (億元) | 佔成交 比重 |
| 2008 | 290,563.19 | 79,161 | 27.20% | 3,712 | 1.28% | 15,028 | 5.17% | 97,883 | 33.70% |
| 2009 | 321,623.76 | 86,392 | 26.90% | 3,849 | 1.20% | 18,239 | 5.67% | 108,080 | 33.60% |
| 2010 | 257,988.82 | 66,140 | 25.60% | 3,090 | 1.20% | 14,715 | 5.70% | 83,945 | 32.50% |
| 2011 | 189,156.19 | 55,999 | 29.60% | 2,673 | 1.41% | 14,170 | 7.49% | 72,842 | 38.50% |
| 2012 | 225,050.96 | 63,640 | 28.30% | 3,454 | 1.53% | 17,773 | 7.90% | 84,867 | 37.70% |
| 2013 | 230,432.84 | 53,056 | 23% | 3,838 | 1.67% | 14,441 | 6.27% | 71,335 | 31% |
| 2014 | 196,033.28 | 47,143 | 24% | 4,522 | 2.31% | 8,940 | 4.56% | 60,604 | 30.90% |
| 2015 | 207,885.24 | 46,291 | 22.30% | 5,370 | 2.58% | 6,890 | 3.31% | 58,551 | 28.20% |
| 2016 | 269,957.04 | 60,276 | 22.30% | 6,697 | 2.48% | 10,234 | 3.79% | 77,208 | 28.60% |
| 2017 | 288,855.24 | 52,074 | 18% | 7,158 | 2.48% | 8,752 | 3.03% | 67,985 | 23.50% |
| 2018 | 301,181.19 | 46,857 | 15.60% | 8,243 | 2.74% | 6,877 | 2.28% | 61,977 | 20.60% |
| 2019 | 266,644.92 | 67,170 | 25.20% | 8,772 | 3.29% | 7,985 | 2.99% | 83,927 | 31.50% |
| 2020 | 491,825.26 | 133,976 | 27.20% | 5,960 | 1.21% | 20,396 | 4.15% | 160,306 | 32.60% |

資料來源：GoodInfo 台灣股市資訊網

第二節、各模型和特徵組合之預測誤差和預測漲跌結果

本節為第三章第四節說明之各模型和特徵組合之預測誤差和預測漲跌準確率結果以及特徵重要度，以比較各模型及特徵組合之結果以及檢視具有解釋力之重要特徵。

2.1 預測 20 日後之每日報酬率誤差及漲跌

本節為各模型和特徵組合預測 20 日後報酬和漲跌之結果，如表 7 和表 8 所示；為比較預測誤差，表 7 為各模型搭配不同特徵組合之預測 20 日後報酬率誤差之結果，以 RMSE 評比。結果顯示，三種模型中 LightGBM 之預測誤差最小，其中 RMSE 最小之組合為 LightGBM 模型搭配投信特徵組合之 0.0557；而各法人特徵組合相比之下投信特徵組合之預測誤差最低。

表 7、各模型和特徵組合之預測 20 日後報酬率誤差比較

| 模型 | 所有特徵 | 外資 | 投信 | 自營商 |
|----------|--------|--------|--------|--------|
| XGBoost | 0.2705 | 0.1857 | 0.0948 | 0.2063 |
| LightGBM | 0.0663 | 0.0564 | 0.0557 | 0.0611 |
| CatBoost | 0.1428 | 0.1491 | 0.1088 | 0.1290 |

表 8 為各模型搭配不同特徵組合之預測 20 日後漲跌之結果，準確率以 F1-score 評比。結果顯示預測 20 日後之漲跌中，投信特徵組合各模型中之預測漲跌最為準確，F1-score 皆高於 0.7，其中最高之組合為 CatBoost 模型搭配投信特徵組合，F1-score 為 0.7062。

表 8、各模型和特徵組合之預測 20 日後漲跌 F1-score 結果

| 模型 | 所有特徵 | 外資 | 投信 | 自營商 |
|----------|--------|--------|--------|--------|
| XGBoost | 0.6521 | 0.6790 | 0.7025 | 0.6949 |
| LightGBM | 0.6785 | 0.6815 | 0.7016 | 0.6918 |
| CatBoost | 0.6935 | 0.6950 | 0.7062 | 0.7022 |

2.2 預測 10 日後之每日報酬率誤差及漲跌

本節為各模型和特徵組合預測 10 日後報酬和漲跌之結果，如表 9 和 10 所示；為比較預測誤差，表 9 為各模型搭配不同特徵組合之預測報酬率之結果，以 RMSE 評比。結果顯示，預測 10 日後報酬率結果中，三模型中預測誤差最小仍為 LightGBM 模型，各特徵組合中誤差最小皆為投信特徵，而 RMSE 最小之組合為 LightGBM 模型搭配投信特徵組合，RMSE 為 0.0652。

表 9、各模型和特徵組合之預測 10 日後報酬率誤差比較

| 模型 | 所有特徵 | 外資 | 投信 | 自營商 |
|----------|--------|--------|--------|--------|
| XGBoost | 0.2337 | 0.2535 | 0.1635 | 0.2139 |
| LightGBM | 0.1143 | 0.1227 | 0.0652 | 0.0988 |
| CatBoost | 0.1477 | 0.1759 | 0.0970 | 0.1367 |

表 10 為各模型搭配不同特徵組合之預測 10 日後漲跌之結果，準確率以 F1-score 評比。結果顯示預測 10 日後之漲跌中，投信特徵組合同樣為各組合中預測漲跌表現最佳，F1-score 最高之組合為 XGBoost 模型搭配投信組合之 0.7058。自營商特徵之表現僅次於投信特徵，F1-score 為 0.69 左右。

表 10、各模型和特徵組合之預測 10 日後漲跌 F1-score 結果

| 模型 | 所有特徵 | 外資 | 投信 | 自營商 |
|----------|--------|--------|--------|--------|
| XGBoost | 0.6498 | 0.6677 | 0.7058 | 0.6920 |
| LightGBM | 0.6767 | 0.6796 | 0.7020 | 0.6912 |
| CatBoost | 0.6859 | 0.6887 | 0.7054 | 0.6979 |

2.3 預測 5 日後之每日報酬率誤差及漲跌

本節為各模型和特徵組合預測 5 日後報酬和漲跌之結果，如表 11 和表 12 所示；表 11 為各模型搭配不同特徵組合之預測報酬率之結果，以 RMSE 評比。結果顯示法人特徵組合中，投信特徵組合之預測誤差最低。其中 RMSE 最低之組合為 LightGBM 模型搭配投信特徵之組合，RMSE 為 0.0588；而三種模型相比之下，LightGBM 模型之預測誤差最低。

表 11、各模型和特徵組合之預測 5 日後報酬率誤差比較

| 模型 | 所有特徵 | 外資 | 投信 | 自營商 |
|----------|--------|--------|--------|--------|
| XGBoost | 0.1853 | 0.2225 | 0.0980 | 0.1635 |
| LightGBM | 0.0662 | 0.0595 | 0.0588 | 0.0613 |
| CatBoost | 0.1315 | 0.1442 | 0.0969 | 0.1191 |

表 12 為各模型搭配不同特徵組合之預測 5 日後漲跌之結果，準確率以 F1-score 評比。結果顯示各特徵組合中投信特徵組合預測漲跌最為準確，其中 F1-score 最高之組合為 XGBoost 模型搭配投信特徵組合為 0.7069。且各組合之準確度排序和預測 5 日和 10 日後漲跌之結論相同。

表 12、各模型和特徵組合之預測 5 日後漲跌 F1-score 結果

| 模型 | 所有特徵 | 外資 | 投信 | 自營商 |
|----------|--------|--------|--------|--------|
| XGBoost | 0.6804 | 0.6862 | 0.7069 | 0.7002 |
| LightGBM | 0.6763 | 0.6795 | 0.7005 | 0.6910 |
| CatBoost | 0.6869 | 0.6896 | 0.7020 | 0.6959 |

2.4 特徵重要度

本研究亦分別檢視三種模型中特徵重要度；特徵重要度以 Gain、Frequency 和 Cover 評估。Gain 即代表該特徵對模型的貢獻度，該數值愈高代表對於生成預測結果越重要；Cover 指標為：覆蓋率，代表和該特徵對應有關的觀測值數量；Frequency 則為出現在模型所有樹的相對次數，代表某特徵出現在模型中樹的相對次數之百分比(%)。由於本研究分為四種特徵組合訓練，以下依序列出三種模型訓練之各特徵中 Gain 前五高之特徵。

表 13 為所有特徵於三種模型訓練中重要度最高之前五項特徵。結果顯示，三種模型之重要特徵有相當多共同特徵，其中各模型中貢獻度最高之特徵為信用交易買賣超和外資持股和成交量之交叉項等特徵，顯示外資持股相關之特徵具有一定參考價值。

表 13、所有特徵組合之重要特徵

| 重要度 排名 | XGBoost | LightGBM | CatBoost |
|-----------|----------------|---------------------|--------------------|
| 1 | 外資持股市值 | 外資買賣超張數/ 20 日均量 | 信用交易買賣超 張數 |
| 2 | 三大法人合計 持股市值 | 外資買賣超張數/ 5 日均量 | 外資買賣超張數/ 20 日均量 |
| 3 | 外資持股數 | 外資買賣超金額/ 成交量 | 自營商持股市值 |
| 4 | 三大法人合計 持股數 | 信用交易買賣超 張數 | 信用交易買賣超 金額 |
| 5 | 外資持股率 | 自營商買賣超張 數/20 日均量 | 外資持股市值 |

表 14 為外資特徵於三種模型訓練中重要度最高之前五項特徵，結果顯示各模型中重要度最高之特徵並無顯著差異，整體而言外資持股率在三模型中排名皆位於前兩名，顯示該特徵為外資特徵中最重要特徵；而持股數和持股率排名僅次於持股市值，顯示外資持股對於模型之訓練相當重要。

表 14、外資特徵組合之重要特徵

| 重要度 排名 | XGBoost | LightGBM | CatBoost |
|-----------|-----------------|-----------------|-----------------|
| 1 | 外資買賣超張數/ 成交量 | 外資持股率 | 外資持股市值 |
| 2 | 外資持股市值 | 外資持股市值 | 外資持股數 |
| 3 | 外資持股數 | 外資買賣超張 數/成交量 | 外資持股率 |
| 4 | 外資買賣超金額/ 成交量 | 外資持股數 | 外資買賣超張數 |
| 5 | 外資買賣超金額 | 外資買賣超金 額/成交量 | 外資買賣超張數/ 成交量 |

表 15 為投信特徵於三種模型訓練中重要度最高之前五項特徵，結果顯示各模型中重要度最高之特徵有些許差異，不過持股市值和持股率等特徵仍為重要度排名最高者，顯示投信持股相關特徵為投信特徵組合中重要特徵。

表 15、投信特徵組合之重要特徵

| 重要度 排名 | XGBoost | LightGBM | CatBoost |
|-----------|-------------------|-----------------|----------|
| 1 | 投信買賣超張數/ 成交量 | 投信持股率 | 投信持股率 |
| 2 | 投信持股市值 | 投信持股市值 | 投信持股數 |
| 3 | 投信持股率 | 投信持股數 | 投信持股市值 |
| 4 | 投信持股數 | 投信買賣超張數/ 成交量 | 投信買賣超張數 |
| 5 | 投信買賣超張數 /5 日均量 | 投信買賣超金額/ 成交量 | 投信買賣超金額 |

表 16 為自營商特徵搭配三種模型訓練中重要度最高之前五項特徵，結果顯示自營商特徵於各模型訓練之重要特徵和外資和投信特徵有所不同，其中重要度最高之特徵為自營商買賣超張數除以成交量之交叉項，整體而言自營商特徵中重要特徵和買賣超張數較為相關。

表 16、自營商特徵組合之重要特徵

| 重要度 排名 | XGBoost | LightGBM | CatBoost |
|-----------|---------------------|------------------|------------------|
| 1 | 自營商買賣超金額/ 成交量 | 自營商買賣超張數/ 成交量 | 自營商買賣超張數/ 成交量 |
| 2 | 自營商買賣超張數/ 成交量 | 自營商持股率 | 自營商買賣超金額/ 成交量 |
| 3 | 自營商買賣超張數 /5 日均量 | 自營商持股市值 | 自營商買賣超金額 |
| 4 | 自營商買賣超金額 | 自營商持股數 | 自營商持股率 |
| 5 | 自營商買賣超張數 /20 日均量 | 自營商買賣超金額/ 成交量 | 自營商持股市值 |

第三節、交易策略績效

本節為第三章說明之交易策略之結果，依序呈現 XGBoost、LightGBM 和 CatBoost 模型搭配特徵組合之交易策略在不同股票檔數(5、10、20、30、40 和 50 檔)之策略績效，以累積報酬率、年化報酬率、年化標準差、夏普比率和最大回撤率等指標和累積報酬率圖呈現，並且列出各組合逐年之報酬率。前段為每月更換標的之結果，後段為每週更換標的之結果。

3.1 每月更換標的策略績效

(1)XGBoost 模型交易策略結果

表 17 和圖 5 至圖 8 為 XGBoost 模型搭配各特徵組合預測 20 日後報酬率並取最高之標的檔數每月更換標的之交易策略績效。結果顯示各組合中投信特徵表現最佳，各檔數之年化報酬率高達 20-30%以上；最大回撤率亦為各特徵組合中最低；自營商特徵表現次之，各檔數亦有年化報酬率 20%以上之表現；所有特徵之年化報酬率為 15-20%左右，而外資特徵報酬最低，年化報酬約 10%左右。

表 17、XGBoost 模型搭配各特徵組合每月更換標的策略績效

| 特徵組合 | 檔數 | 累積 報酬率 | 年化 報酬率 | 年化 標準差 | 夏普 比率 | 最大 回撤 |
|------|----|-----------|-----------|-----------|----------|----------|
| 所有特徵 | 5 | 34.69% | 6.14% | 24.49% | 0.2506 | -43.25% |
| 所有特徵 | 10 | 108.18% | 15.79% | 19.03% | 0.8301 | -40.84% |
| 所有特徵 | 20 | 96.19% | 14.43% | 15.11% | 0.9547 | -35.99% |
| 所有特徵 | 30 | 153.53% | 20.45% | 14.32% | 1.4277 | -32.81% |
| 所有特徵 | 40 | 165.64% | 21.58% | 13.71% | 1.5744 | -31.21% |
| 所有特徵 | 50 | 159.39% | 21.00% | 13.33% | 1.5760 | -30.72% |
| 外資 | 5 | 124.62% | 17.57% | 25.05% | 0.7013 | -43.32% |
| 外資 | 10 | 42.40% | 7.33% | 18.80% | 0.3897 | -44.09% |
| 外資 | 20 | 58.30% | 9.62% | 15.15% | 0.6350 | -44.98% |
| 外資 | 30 | 74.14% | 11.73% | 14.00% | 0.8382 | -43.36% |
| 外資 | 40 | 67.05% | 10.81% | 13.56% | 0.7967 | -42.04% |
| 外資 | 50 | 69.69% | 11.16% | 13.08% | 0.8529 | -39.08% |
| 投信 | 5 | 616.75% | 48.28% | 23.34% | 2.0687 | -29.85% |
| 投信 | 10 | 321.53% | 33.34% | 19.19% | 1.7378 | -29.22% |
| 投信 | 20 | 241.87% | 27.87% | 17.22% | 1.6182 | -31.66% |
| 投信 | 30 | 194.13% | 24.08% | 16.39% | 1.4696 | -33.84% |
| 投信 | 40 | 161.76% | 21.22% | 15.97% | 1.3287 | -35.40% |
| 投信 | 50 | 132.27% | 18.36% | 15.77% | 1.1638 | -35.56% |
| 自營商 | 5 | 287.02% | 31.08% | 22.62% | 1.3744 | -39.36% |
| 自營商 | 10 | 206.57% | 25.11% | 18.73% | 1.3406 | -44.50% |
| 自營商 | 20 | 258.45% | 29.09% | 16.92% | 1.7190 | -35.01% |
| 自營商 | 30 | 236.09% | 27.44% | 16.06% | 1.7080 | -34.06% |
| 自營商 | 40 | 205.94% | 25.06% | 15.43% | 1.6241 | -34.11% |
| 自營商 | 50 | 181.58% | 23.00% | 14.90% | 1.5438 | -33.74% |

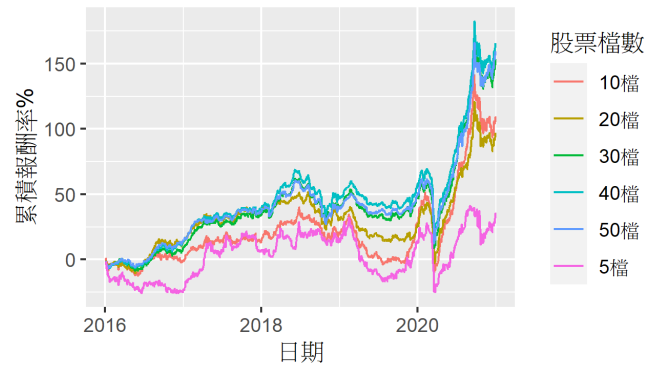


圖 5、XGBoost 模型搭配所有特徵每月更換標的策略累積報酬率圖

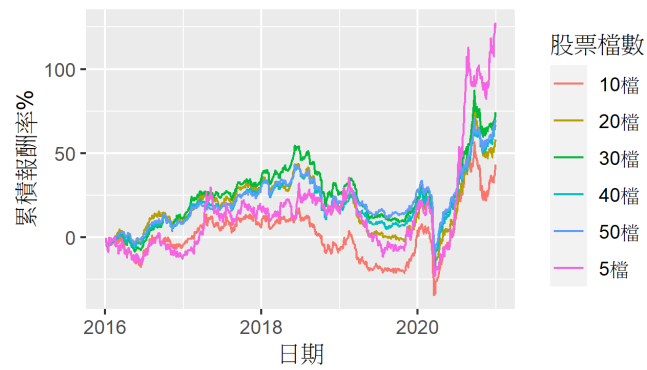


圖 6、XGBoost 模型搭配外資特徵每月更換標的策略累積報酬率圖

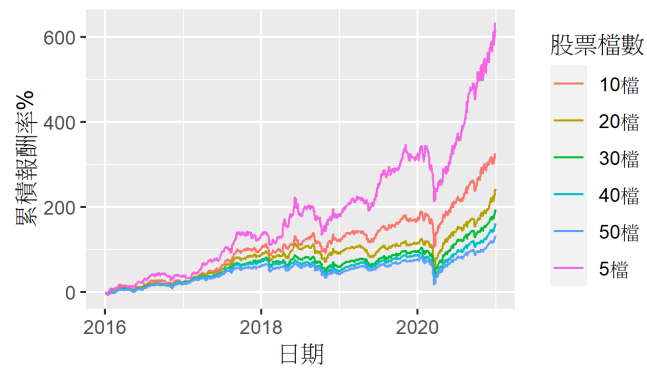


圖 7、XGBoost 模型搭配投信特徵每月更換標的策略累積報酬率圖

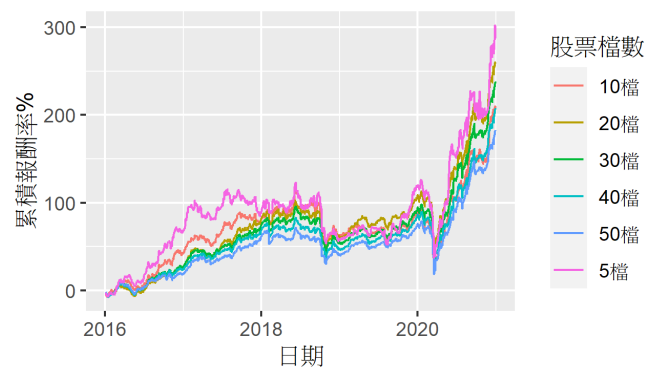


圖 8、XGBoost 模型搭配自營商特徵每月更換標的策略累積報酬率圖

表 18 為 XGBoost 模型搭配各特徵及檔數組合每月更換標的之逐年報酬率結果，結果顯示投信特徵和所有特徵於 2020 有顯著高報酬，顯示兩組合高報酬主要來自於該年度；自營商組合除 2018 年之跌幅較大以外，其餘年份表現良好；而外資特徵組合雖無大幅虧損之年度，不過各年度皆無顯著高報酬導致表現為各組合中不佳。

表 18、XGBoost 模型每月更換標的策略各組合逐年報酬率

| 特徵組合 | 檔數 | 2016 | 2017 | 2018 | 2019 | 2020 |
|------|----|---------|--------|---------|---------|---------|
| 加權指數 | - | 10.98% | 15.01% | -8.60% | 23.33% | 22.80% |
| 所有特徵 | 5 | -21.96% | 28.46% | 8.20% | -2.22% | 21.93% |
| 所有特徵 | 10 | -1.93% | 13.82% | 4.50% | 7.88% | 79.33% |
| 所有特徵 | 20 | 14.38% | 20.76% | -3.87% | -4.33% | 69.22% |
| 所有特徵 | 30 | 9.91% | 24.64% | 6.55% | 8.51% | 104.38% |
| 所有特徵 | 40 | 14.95% | 24.63% | 8.97% | 9.89% | 107.03% |
| 所有特徵 | 50 | 15.75% | 21.78% | 3.10% | 10.31% | 108.18% |
| 外資 | 5 | -10.11% | 27.30% | -0.32% | 2.33% | 105.74% |
| 外資 | 10 | -4.17% | 9.70% | -15.05% | 11.34% | 38.86% |
| 外資 | 20 | 10.52% | 21.66% | -15.71% | -3.93% | 44.50% |
| 外資 | 30 | 13.02% | 22.34% | -9.24% | -2.78% | 50.76% |
| 外資 | 40 | 9.70% | 21.42% | -14.16% | 4.72% | 45.20% |
| 外資 | 50 | 9.60% | 20.79% | -9.24% | 7.01% | 41.29% |
| 投信 | 5 | 39.69% | 95.99% | 46.83% | 140.39% | 289.98% |
| 投信 | 10 | 24.31% | 78.87% | 18.89% | 47.20% | 145.62% |
| 投信 | 20 | 23.66% | 65.31% | 5.11% | 21.82% | 124.51% |
| 投信 | 30 | 20.65% | 55.74% | -14.59% | 34.89% | 96.36% |
| 投信 | 40 | 24.55% | 49.23% | -22.31% | 35.85% | 73.67% |
| 投信 | 50 | 22.17% | 40.53% | -17.98% | 32.83% | 53.95% |
| 自營商 | 5 | 70.77% | 19.59% | -33.03% | 60.62% | 167.25% |
| 自營商 | 10 | 45.38% | 36.54% | -18.58% | 24.56% | 115.65% |
| 自營商 | 20 | 28.78% | 54.82% | -20.87% | 44.62% | 149.52% |
| 自營商 | 30 | 29.56% | 46.39% | -20.10% | 36.98% | 141.10% |
| 自營商 | 40 | 26.40% | 43.73% | -21.53% | 37.02% | 119.16% |
| 自營商 | 50 | 20.97% | 41.72% | -20.40% | 33.37% | 105.22% |

(2)LightGBM 模型交易策略結果

表 19 以及圖 9 至圖 12 為 LightGBM 模型搭配各特徵組合預測 20 日後之報酬並取最高不同檔數下，每月更換標的之交易策略績效，結果顯示整體表現而言，投信特徵表現最好，各檔數年化報酬率高達 20%以上；外資特徵次之則有 20%左右之年化報酬率，而自營商特徵之年化報酬率約為 10-15%。所有特徵年化報酬率則為各組合最低，大約 6-10%。

表 19、LightGBM 模型搭配各特徵組合每月更換標的策略績效

| 特徵組合 | 檔數 | 累積 報酬率 | 年化 報酬率 | 年化 標準差 | 夏普 比率 | 最大 回撤 |
|------|----|-----------|-----------|-----------|----------|----------|
| 所有特徵 | 5 | 24.05% | 4.41% | 23.25% | 0.1894 | -48.37% |
| 所有特徵 | 10 | 71.56% | 11.40% | 18.39% | 0.6198 | -49.56% |
| 所有特徵 | 20 | 40.21% | 6.99% | 15.26% | 0.4584 | -51.32% |
| 所有特徵 | 30 | 52.95% | 8.87% | 13.92% | 0.6375 | -48.40% |
| 所有特徵 | 40 | 49.39% | 8.36% | 13.15% | 0.6358 | -47.17% |
| 所有特徵 | 50 | 44.26% | 7.60% | 12.50% | 0.6085 | -44.06% |
| 外資 | 5 | 241.89% | 27.87% | 33.73% | 0.8264 | -49.01% |
| 外資 | 10 | 113.84% | 16.42% | 21.81% | 0.7529 | -55.29% |
| 外資 | 20 | 158.04% | 20.88% | 16.07% | 1.2988 | -49.53% |
| 外資 | 30 | 157.49% | 20.82% | 14.91% | 1.3962 | -47.60% |
| 外資 | 40 | 131.30% | 18.26% | 13.57% | 1.3454 | -42.45% |
| 外資 | 50 | 124.88% | 17.60% | 12.35% | 1.4248 | -39.83% |
| 投信 | 5 | 185.82% | 23.37% | 23.47% | 0.9958 | -41.59% |
| 投信 | 10 | 221.40% | 26.30% | 20.09% | 1.3093 | -32.02% |
| 投信 | 20 | 262.91% | 29.41% | 18.60% | 1.5814 | -35.93% |
| 投信 | 30 | 265.41% | 29.59% | 17.86% | 1.6564 | -34.64% |
| 投信 | 40 | 255.35% | 28.86% | 17.17% | 1.6812 | -34.27% |
| 投信 | 50 | 235.23% | 27.37% | 16.66% | 1.6429 | -34.43% |
| 自營商 | 5 | 115.70% | 16.62% | 24.86% | 0.6686 | -45.51% |
| 自營商 | 10 | 74.95% | 11.84% | 18.51% | 0.6396 | -41.22% |
| 自營商 | 20 | 64.08% | 10.41% | 15.25% | 0.6829 | -38.80% |
| 自營商 | 30 | 57.47% | 9.51% | 14.08% | 0.6754 | -43.07% |
| 自營商 | 40 | 51.52% | 8.67% | 13.18% | 0.6577 | -40.50% |
| 自營商 | 50 | 48.93% | 8.29% | 12.51% | 0.6626 | -37.43% |

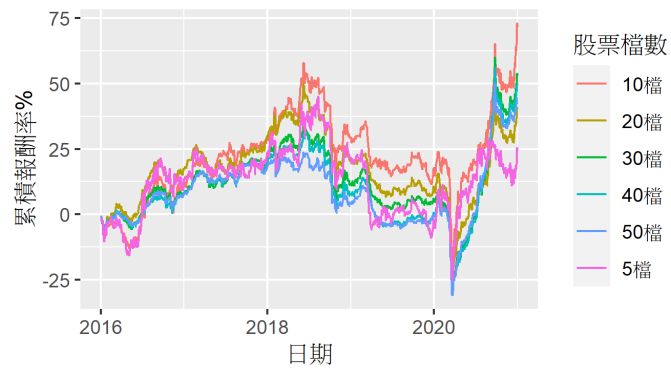


圖 9、LightGBM 模型搭配所有特徵每月更換標的策略累積報酬率圖

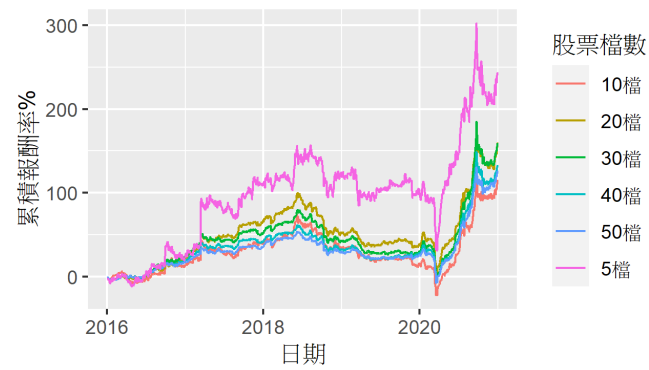


圖 10、LightGBM 模型搭配外資特徵每月更換標的策略累積報酬率圖

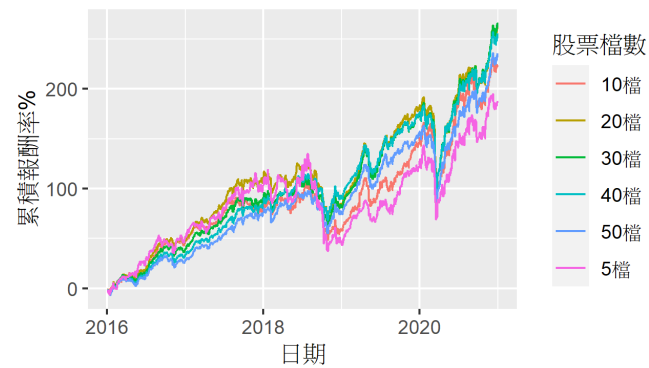


圖 11、LightGBM 模型搭配投信特徵每月更換標的策略累積報酬率圖

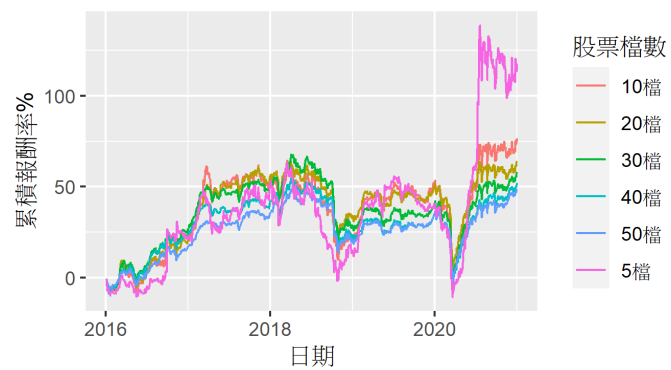


圖 12、LightGBM 模型搭配自營商特徵每月更換標的策略累積報酬率圖

表 20 為 LightGBM 模型搭配各特徵及檔數組合每月更換標的之逐年報酬率結果，結果顯示外資特徵投信特徵主要於 2019 年獲利最多，除 2018 年虧損較多外其餘年份有不錯表現；外資特徵和所有特徵則是於 2019 年表現不佳，而自營商特徵整體而言表現較為穩定。

表 20、LightGBM 模型每月更換標的策略各組合逐年報酬率

| 特徵組合 | 檔數 | 2016 | 2017 | 2018 | 2019 | 2020 |
|------|----|--------|--------|---------|---------|---------|
| 加權指數 | - | 10.98% | 15.01% | -8.60% | 23.33% | 22.80% |
| 所有特徵 | 5 | 14.73% | 10.25% | -1.98% | -23.73% | 26.82% |
| 所有特徵 | 10 | 12.12% | 17.24% | 0.05% | -11.49% | 54.60% |
| 所有特徵 | 20 | 17.85% | 12.10% | -10.07% | -9.83% | 29.97% |
| 所有特徵 | 30 | 8.97% | 13.18% | -11.19% | -5.67% | 47.04% |
| 所有特徵 | 40 | 9.00% | 11.35% | -12.78% | -7.12% | 48.84% |
| 所有特徵 | 50 | 8.81% | 9.67% | -14.36% | -4.37% | 44.17% |
| 外資 | 5 | 25.42% | 83.53% | 12.86% | -21.46% | 141.56% |
| 外資 | 10 | 14.86% | 26.38% | -5.48% | -19.37% | 98.22% |
| 外資 | 20 | 22.99% | 43.23% | -14.40% | -13.34% | 119.69% |
| 外資 | 30 | 24.90% | 26.60% | -9.77% | -12.47% | 127.33% |
| 外資 | 40 | 20.24% | 21.54% | -9.76% | -4.17% | 102.56% |
| 外資 | 50 | 17.87% | 17.68% | -6.70% | -3.34% | 98.50% |
| 投信 | 5 | 52.85% | 50.87% | -56.84% | 77.69% | 57.39% |
| 投信 | 10 | 48.12% | 31.49% | -23.83% | 89.19% | 70.18% |
| 投信 | 20 | 49.55% | 60.83% | -28.16% | 97.21% | 79.06% |
| 投信 | 30 | 40.44% | 49.64% | -7.95% | 91.45% | 88.72% |
| 投信 | 40 | 34.72% | 50.49% | 7.16% | 82.11% | 78.39% |
| 投信 | 50 | 29.54% | 47.23% | 0.09% | 78.47% | 77.77% |
| 自營商 | 5 | 22.10% | 30.18% | -38.53% | 26.61% | 74.74% |
| 自營商 | 10 | 23.15% | 30.81% | -32.46% | 31.77% | 21.57% |
| 自營商 | 20 | 21.40% | 34.83% | -22.52% | 17.42% | 12.37% |
| 自營商 | 30 | 28.05% | 24.13% | -23.49% | 12.60% | 15.39% |
| 自營商 | 40 | 25.15% | 17.00% | -18.82% | 12.70% | 14.84% |
| 自營商 | 50 | 18.00% | 17.85% | -15.43% | 14.62% | 12.73% |

(3)CatBoost 模型交易策略結果

表 21 和圖 13 至圖 16 為 CatBoost 模型搭配各特徵組合預測 20 日後之報酬率並取最高不同檔數下並每月更換標的之交易策略績效，結果顯示投信特徵仍為表現最好，除 50 檔外其餘之年化報酬皆高於 20%以上；所有特徵表現次之，於部分檔數亦有 20%左右之年化報酬率；所有特徵組合之最大回撤為各組合中最低；自營商特徵年化報酬率為 15-20%，而外資特徵於 CatBoost 模型中年化報酬率為各組合最低，大約為 10%左右。

表 21、CatBoost 模型搭配各特徵組合每月更換標的策略績效

| 特徵組合 | 檔數 | 累積 報酬率 | 年化 報酬率 | 年化 標準差 | 夏普 比率 | 最大 回撤 |
|------|----|-----------|-----------|-----------|----------|----------|
| 所有特徵 | 5 | 44.08% | 7.58% | 24.44% | 0.3100 | -54.07% |
| 所有特徵 | 10 | 85.27% | 13.13% | 17.67% | 0.7430 | -39.10% |
| 所有特徵 | 20 | 138.83% | 19.02% | 13.76% | 1.3819 | -30.70% |
| 所有特徵 | 30 | 146.39% | 19.76% | 12.64% | 1.5635 | -25.20% |
| 所有特徵 | 40 | 193.62% | 24.04% | 12.09% | 1.9882 | -24.15% |
| 所有特徵 | 50 | 195.87% | 24.23% | 11.89% | 2.0384 | -24.24% |
| 外資 | 5 | 54.59% | 9.10% | 26.90% | 0.3384 | -54.61% |
| 外資 | 10 | 53.96% | 9.01% | 21.23% | 0.4245 | -51.98% |
| 外資 | 20 | 48.54% | 8.23% | 16.87% | 0.4881 | -45.80% |
| 外資 | 30 | 76.41% | 12.02% | 14.87% | 0.8085 | -44.47% |
| 外資 | 40 | 86.31% | 13.25% | 14.19% | 0.9340 | -37.84% |
| 外資 | 50 | 78.34% | 12.27% | 13.71% | 0.8946 | -41.33% |
| 投信 | 5 | 594.65% | 47.35% | 25.46% | 1.8599 | -35.01% |
| 投信 | 10 | 217.35% | 25.98% | 20.98% | 1.2385 | -42.38% |
| 投信 | 20 | 210.79% | 25.46% | 18.27% | 1.3931 | -38.05% |
| 投信 | 30 | 189.26% | 23.67% | 17.38% | 1.3620 | -36.93% |
| 投信 | 40 | 164.10% | 21.44% | 16.67% | 1.2860 | -35.64% |
| 投信 | 50 | 127.78% | 17.90% | 16.22% | 1.1037 | -35.61% |
| 自營商 | 5 | 102.03% | 15.10% | 22.72% | 0.6647 | -58.95% |
| 自營商 | 10 | 155.18% | 20.61% | 19.32% | 1.0667 | -51.79% |
| 自營商 | 20 | 104.69% | 15.40% | 16.43% | 0.9376 | -42.78% |
| 自營商 | 30 | 74.80% | 11.82% | 15.23% | 0.7760 | -43.17% |
| 自營商 | 40 | 74.61% | 11.79% | 14.52% | 0.8120 | -40.43% |
| 自營商 | 50 | 77.80% | 12.20% | 14.36% | 0.8497 | -41.99% |

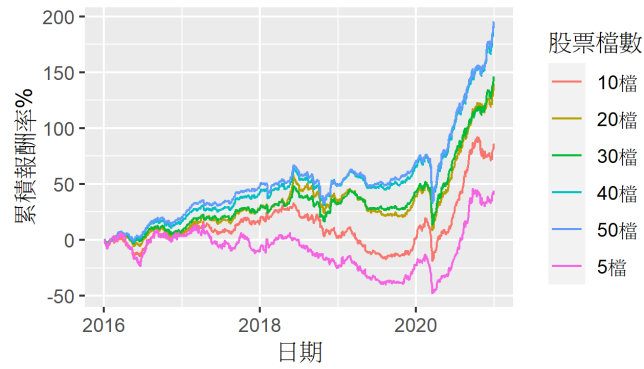


圖 13、CatBoost 模型搭配所有特徵每月更換標的策略累積報酬率圖

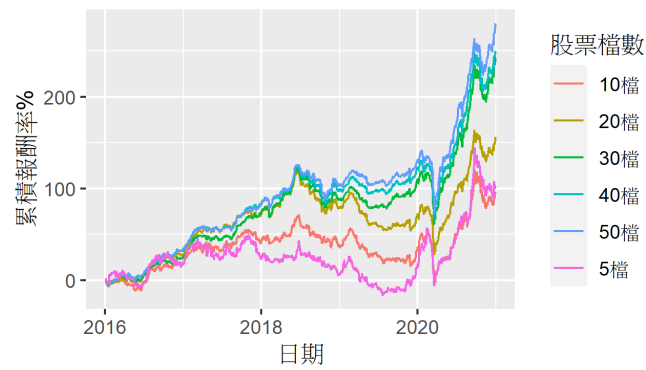


圖 14、CatBoost 模型搭配外資特徵每月更換標的策略累積報酬率圖

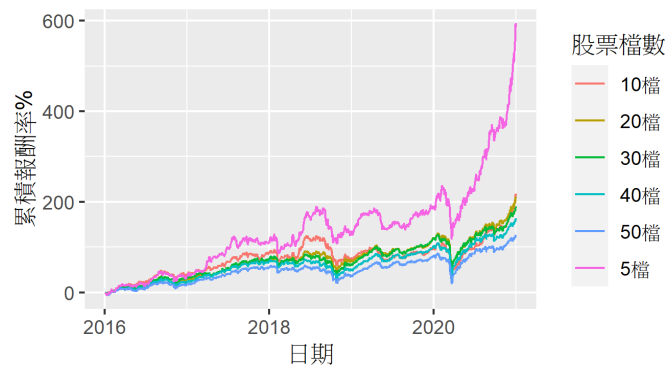


圖 15、CatBoost 模型搭配投信特徵每月更換標的策略累積報酬率圖

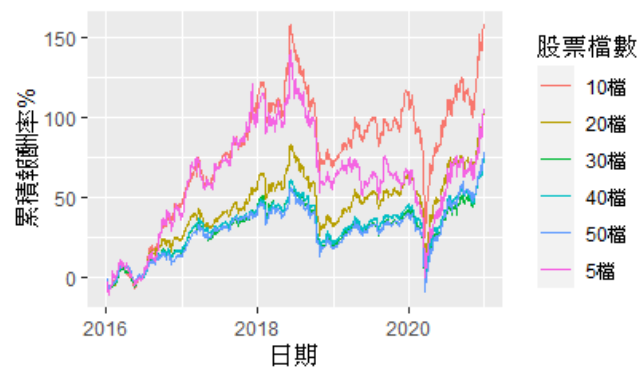


圖 16、CatBoost 模型搭配自營商特徵每月更換標的策略累積報酬率圖

表 22 為 CatBoost 模型搭配各特徵及檔數組合每月更換標的之逐年報酬率結果，結果顯示投信特徵高報酬主要來自多數年份顯著優於大盤之表現；外資特徵和所有特徵於 2019 年表現稍劣於大盤；自營商除 2018 年外則是每年皆有 20-30% 之穩定成長。

表 22、CatBoost 模型每月更換標的策略各組合逐年報酬率

| 特徵組合 | 檔數 | 2016 | 2017 | 2018 | 2019 | 2020 |
|------|----|---------|---------|---------|--------|---------|
| 加權指數 | - | 10.98% | 15.01% | -8.60% | 23.33% | 22.80% |
| 所有特徵 | 5 | 3.27% | -13.67% | -16.67% | 1.85% | 67.07% |
| 所有特徵 | 10 | 4.48% | 10.05% | -15.11% | -0.60% | 83.15% |
| 所有特徵 | 20 | 5.99% | 20.72% | 6.84% | 4.27% | 98.91% |
| 所有特徵 | 30 | 7.34% | 19.35% | 4.30% | 10.84% | 102.65% |
| 所有特徵 | 40 | 16.59% | 25.15% | 6.58% | 17.22% | 126.69% |
| 所有特徵 | 50 | 20.67% | 27.15% | 1.25% | 20.50% | 124.64% |
| 外資 | 5 | -10.59% | -13.30% | -21.37% | 3.50% | 91.35% |
| 外資 | 10 | 3.09% | -7.87% | -25.72% | 1.84% | 79.51% |
| 外資 | 20 | 16.99% | 1.67% | -18.21% | -5.05% | 51.36% |
| 外資 | 30 | 13.49% | 17.89% | -15.12% | 1.83% | 57.39% |
| 外資 | 40 | 10.44% | 15.11% | -6.37% | 7.86% | 59.22% |
| 外資 | 50 | 8.12% | 18.25% | -5.87% | 3.32% | 54.67% |
| 投信 | 5 | 37.48% | 74.03% | 23.50% | 55.80% | 403.82% |
| 投信 | 10 | 41.03% | 48.66% | -11.85% | 21.07% | 117.93% |
| 投信 | 20 | 30.60% | 41.50% | -0.99% | 49.11% | 90.05% |
| 投信 | 30 | 32.64% | 42.82% | -12.59% | 57.39% | 68.10% |
| 投信 | 40 | 24.55% | 44.77% | -18.06% | 51.54% | 60.75% |
| 投信 | 50 | 18.70% | 37.85% | -18.18% | 43.41% | 45.77% |
| 自營商 | 5 | 42.31% | 63.58% | -45.08% | 5.74% | 37.87% |
| 自營商 | 10 | 49.45% | 62.59% | -40.80% | 42.63% | 41.61% |
| 自營商 | 20 | 26.93% | 32.72% | -25.91% | 31.26% | 39.82% |
| 自營商 | 30 | 18.25% | 27.67% | -24.98% | 21.88% | 32.14% |
| 自營商 | 40 | 18.63% | 26.36% | -22.40% | 23.18% | 28.84% |
| 自營商 | 50 | 15.27% | 27.80% | -23.85% | 21.20% | 37.77% |

(4)綜合比較與小結

本段為綜合比較各模型和特徵於各檔數之下表現，首先各模型中 XGBoost 模型之整體報酬率為最高，於 24 檔組合之下有 15 檔年化報酬率高於 20%；相較於 LightGBM 模型之 9 檔和 CatBoost 模型 7 檔高出許多；而法人特徵組合方面，投信特徵組合表現最佳，於三模型之所有檔數年化報酬率皆高於 20%，而觀察各組合逐年之表現，外資特徵普遍於 2020 年有高報酬，但 2019 年之表現則較不理想；而各檔數相比之下，隨著檔數增加不但降低風險，且年化報酬率亦維持一定水準。以報酬率以及夏普比率和最大回撤等指標綜合評估，於每月更換投資標的之結果中，XGBoost 搭配投信特徵以及 CatBoost 搭配投信特徵表現為各組合中績效最佳。

3.2 每週更換標的策略績效

(1) XGBoost 模型交易策略結果

表 23 以及圖 17 至圖 20 為 XGBoost 模型搭配各特徵組合預測 5 日後報酬率並每週更換標的之結果，結果顯示投信特徵和自營商特徵組合為各組合中表現最好，投信特徵組合之各檔數年化報酬率高達 40%以上，最大回撤亦為各特徵組合最低；而自營商特徵之年化報酬則是有 4 種檔數組合高於 40%；其中最高之組合為投信特徵 10 檔，年化報酬率為 48.13%，外資組合亦有 20-30%之年化報酬率。所有特徵之報酬為各組合最低。

表 23、XGBoost 模型搭配各特徵每週更換標的策略績效

| 特徵組合 | 檔數 | 累積 報酬率 | 年化 報酬率 | 年化 標準差 | 夏普 比率 | 最大 回撤 |
|------|----|-----------|-----------|-----------|----------|----------|
| 所有特徵 | 5 | 78.39% | 12.27% | 24.63% | 0.4983 | -43.13% |
| 所有特徵 | 10 | 98.04% | 14.64% | 18.70% | 0.7832 | -41.89% |
| 所有特徵 | 20 | 164.44% | 21.47% | 16.04% | 1.3382 | -33.03% |
| 所有特徵 | 30 | 156.37% | 20.72% | 15.43% | 1.3428 | -33.02% |
| 所有特徵 | 40 | 226.14% | 26.67% | 15.40% | 1.7314 | -30.92% |
| 所有特徵 | 50 | 224.19% | 26.52% | 15.07% | 1.7600 | -31.01% |
| 外資 | 5 | 72.77% | 11.56% | 26.39% | 0.4378 | -49.66% |
| 外資 | 10 | 150.19% | 20.13% | 20.10% | 1.0015 | -49.38% |
| 外資 | 20 | 189.84% | 23.72% | 17.24% | 1.3757 | -51.00% |
| 外資 | 30 | 333.03% | 34.06% | 16.03% | 2.1249 | -35.50% |
| 外資 | 40 | 264.89% | 29.55% | 15.58% | 1.8960 | -33.75% |
| 外資 | 50 | 255.16% | 28.85% | 15.15% | 1.9049 | -33.96% |
| 投信 | 5 | 519.28% | 44.00% | 22.94% | 1.9182 | -30.05% |
| 投信 | 10 | 613.16% | 48.13% | 19.53% | 2.4644 | -28.73% |
| 投信 | 20 | 582.95% | 46.85% | 17.77% | 2.6371 | -29.59% |
| 投信 | 30 | 566.33% | 46.13% | 17.12% | 2.6942 | -27.77% |
| 投信 | 40 | 550.26% | 45.42% | 16.70% | 2.7195 | -28.56% |
| 投信 | 50 | 479.95% | 42.13% | 16.42% | 2.5658 | -30.04% |
| 自營商 | 5 | 415.28% | 38.81% | 23.73% | 1.6352 | -35.46% |
| 自營商 | 10 | 591.09% | 47.20% | 19.87% | 2.3753 | -29.48% |
| 自營商 | 20 | 525.29% | 44.28% | 17.57% | 2.5197 | -30.45% |
| 自營商 | 30 | 500.54% | 43.12% | 17.13% | 2.5178 | -31.96% |
| 自營商 | 40 | 463.56% | 41.32% | 16.66% | 2.4804 | -30.97% |
| 自營商 | 50 | 384.75% | 37.12% | 16.36% | 2.2694 | -31.57% |



圖 17、XGBoost 模型搭配所有特徵每週更換標的策略累積報酬率圖



圖 18、XGBoost 模型搭配外資特徵每週更換標的策略累積報酬率圖

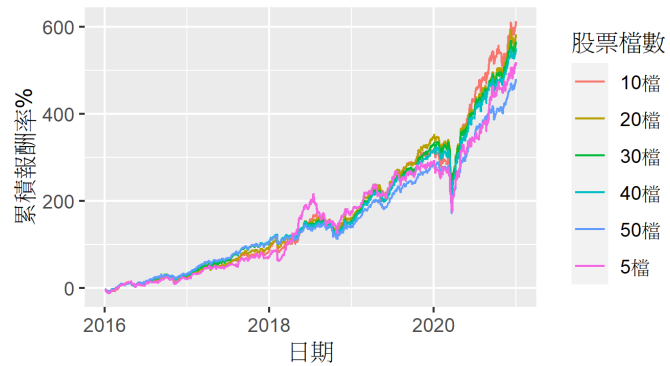


圖 19、XGBoost 模型搭配投信特徵每週更換標的策略累積報酬率圖

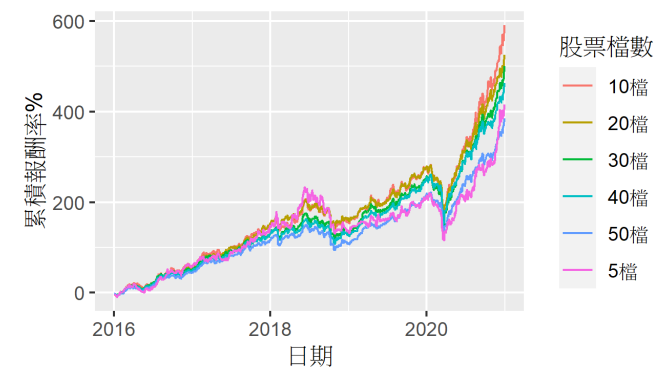


圖 20、XGBoost 模型搭配自營商特徵每週更換標的策略累積報酬率圖

表 24 為 XGBoost 模型搭配各特徵及檔數組合每週更換標的之逐年報酬率結果，結果顯示投信特徵和自營商特徵於 2019 和 2020 兩年度有顯著高報酬；所有特徵於 2018 年部分檔數有較大虧損，而外資特徵除 2016 年外，其餘年份表現亦良好。

表 24、XGBoost 模型每週更換標的策略各組合逐年報酬率

| 特徵組合 | 檔數 | 2016 | 2017 | 2018 | 2019 | 2020 |
|------|----|--------|--------|---------|---------|---------|
| 加權指數 | - | 10.98% | 15.01% | -8.60% | 23.33% | 22.80% |
| 所有特徵 | 5 | 20.38% | 87.38% | -13.51% | 68.76% | 664.06% |
| 所有特徵 | 10 | 37.08% | 63.91% | -40.92% | 26.50% | 158.84% |
| 所有特徵 | 20 | 25.24% | 57.43% | -25.34% | 39.91% | 80.68% |
| 所有特徵 | 30 | 21.02% | 42.15% | -5.14% | 28.46% | 76.74% |
| 所有特徵 | 40 | 22.76% | 35.19% | -5.23% | 27.53% | 68.88% |
| 所有特徵 | 50 | 19.22% | 35.42% | -8.81% | 22.26% | 63.15% |
| 外資 | 5 | -6.41% | 18.40% | -4.81% | -15.40% | 81.15% |
| 外資 | 10 | -2.30% | 53.02% | 0.92% | -14.74% | 111.06% |
| 外資 | 20 | 7.27% | 68.34% | -27.31% | -1.18% | 142.36% |
| 外資 | 30 | 7.88% | 75.81% | 8.93% | 28.39% | 210.41% |
| 外資 | 40 | 11.66% | 60.82% | 2.13% | 32.87% | 156.59% |
| 外資 | 50 | 14.90% | 63.20% | 3.16% | 37.87% | 135.40% |
| 投信 | 5 | 21.61% | 51.29% | 103.08% | 115.65% | 225.87% |
| 投信 | 10 | 28.69% | 58.65% | 66.01% | 166.97% | 291.39% |
| 投信 | 20 | 28.61% | 67.52% | 66.75% | 187.71% | 230.05% |
| 投信 | 30 | 31.40% | 74.89% | 48.49% | 176.51% | 232.02% |
| 投信 | 40 | 32.73% | 76.08% | 42.04% | 166.09% | 229.44% |
| 投信 | 50 | 35.19% | 71.88% | 35.79% | 142.64% | 190.44% |
| 自營商 | 5 | 61.13% | 88.65% | -10.51% | 74.68% | 200.06% |
| 自營商 | 10 | 62.72% | 88.92% | 9.79% | 117.23% | 314.39% |
| 自營商 | 20 | 55.52% | 99.71% | 10.68% | 114.33% | 245.01% |
| 自營商 | 30 | 53.42% | 80.09% | 3.87% | 119.26% | 242.29% |
| 自營商 | 40 | 50.87% | 81.77% | -1.68% | 122.97% | 207.55% |
| 自營商 | 50 | 46.63% | 71.46% | -7.06% | 107.08% | 165.09% |

(2) LightGBM 模型交易策略結果

表 25 及圖 21 至 24 為 LightGBM 模型搭配各特徵組合預測 5 日後報酬率並每週更換標的之結果，結果顯示投信特徵表現最佳，於各檔數年化報酬率高於 40% 以上；外資特徵和所有特徵表現次之，年化報酬率約為 20-30%；而自營商特徵之表現為 LightGBM 各組合中之末，各檔數年化報酬率介於 10-20%。

表 25、LightGBM 模型搭配各特徵組合每週更換標的策略績效

| 特徵組合 | 檔數 | 累積 報酬率 | 年化 報酬率 | 年化 標準差 | 夏普 比率 | 最大 回撤 |
|------|----|-----------|-----------|-----------|----------|----------|
| 所有特徵 | 5 | 387.77% | 37.29% | 25.25% | 1.4771 | -39.14% |
| 所有特徵 | 10 | 344.63% | 34.77% | 19.46% | 1.7866 | -37.47% |
| 所有特徵 | 20 | 175.84% | 22.50% | 16.28% | 1.3817 | -36.68% |
| 所有特徵 | 30 | 148.00% | 19.92% | 14.74% | 1.3511 | -39.82% |
| 所有特徵 | 40 | 127.82% | 17.90% | 14.09% | 1.2704 | -41.68% |
| 所有特徵 | 50 | 142.06% | 19.34% | 13.51% | 1.4316 | -37.90% |
| 外資 | 5 | 769.48% | 54.12% | 31.41% | 1.7229 | -36.37% |
| 外資 | 10 | 399.93% | 37.97% | 21.52% | 1.7640 | -38.86% |
| 外資 | 20 | 273.10% | 30.13% | 15.69% | 1.9200 | -41.36% |
| 外資 | 30 | 155.82% | 20.67% | 13.52% | 1.5292 | -36.54% |
| 外資 | 40 | 158.16% | 20.89% | 13.02% | 1.6045 | -35.67% |
| 外資 | 50 | 134.25% | 18.56% | 12.51% | 1.4831 | -36.17% |
| 投信 | 5 | 1064.36% | 63.39% | 24.23% | 2.6163 | -34.72% |
| 投信 | 10 | 748.58% | 53.37% | 20.67% | 2.5825 | -30.34% |
| 投信 | 20 | 669.48% | 50.40% | 18.40% | 2.7397 | -29.13% |
| 投信 | 30 | 606.30% | 47.84% | 17.47% | 2.7386 | -28.75% |
| 投信 | 40 | 522.53% | 44.16% | 16.87% | 2.6176 | -28.69% |
| 投信 | 50 | 449.00% | 40.58% | 16.57% | 2.4494 | -28.59% |
| 自營商 | 5 | 200.35% | 24.60% | 23.37% | 1.0526 | -38.90% |
| 自營商 | 10 | 186.09% | 23.40% | 17.65% | 1.3255 | -36.85% |
| 自營商 | 20 | 125.74% | 17.68% | 14.25% | 1.2412 | -32.19% |
| 自營商 | 30 | 93.87% | 14.16% | 12.95% | 1.0930 | -30.59% |
| 自營商 | 40 | 67.74% | 10.90% | 12.18% | 0.8949 | -30.58% |
| 自營商 | 50 | 59.58% | 9.80% | 11.76% | 0.8329 | -30.88% |



圖 21、LightGBM 模型搭配所有特徵每週更換標的策略累積報酬率圖

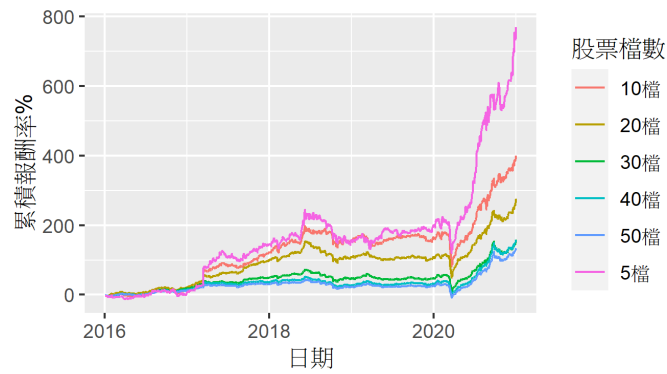


圖 22、LightGBM 模型搭配外資特徵每週更換標的策略累積報酬率圖

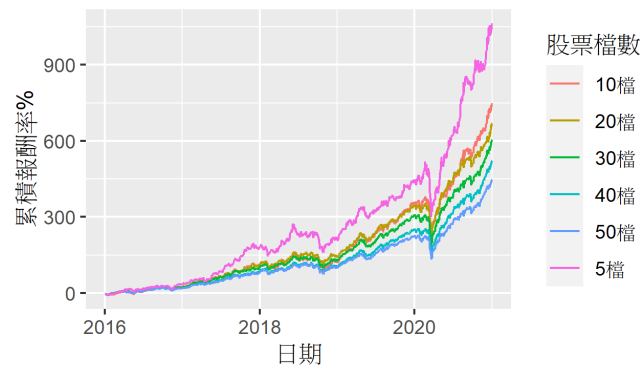


圖 23、LightGBM 模型搭配投信特徵每週更換標的策略累積報酬率圖

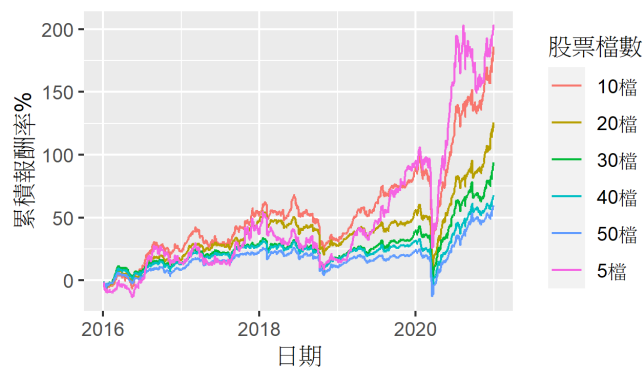


圖 24、LightGBM 模型搭配自營商特徵每週更換標的策略累積報酬率圖

表 26 為 LightGBM 模型搭配各特徵及檔數組合預測 5 日後報酬並每週更換標的之逐年報酬率結果，結果顯示投信特徵於 2019 年和 2020 年之高報酬使得該組合表現最好；而外資特徵組合和所有特徵主要於 2017 年和 2020 年表現較好，僅 2019 年表現較不佳；而自營商特徵於各年之漲跌程度相較其他組合較小。

表 26、LightGBM 模型每週更換標的策略各組合逐年報酬率

| 特徵組合 | 檔數 | 2016 | 2017 | 2018 | 2019 | 2020 |
|------|----|--------|---------|---------|---------|---------|
| 加權指數 | - | 10.98% | 15.01% | -8.60% | 23.33% | 22.80% |
| 所有特徵 | 5 | 12.94% | 79.26% | -5.50% | 72.73% | 221.70% |
| 所有特徵 | 10 | 19.85% | 75.35% | 8.22% | 74.92% | 157.40% |
| 所有特徵 | 20 | 15.60% | 50.06% | -21.22% | 28.47% | 96.19% |
| 所有特徵 | 30 | 21.12% | 40.97% | -21.14% | 6.38% | 95.79% |
| 所有特徵 | 40 | 17.76% | 33.87% | -16.16% | -2.27% | 90.39% |
| 所有特徵 | 50 | 20.94% | 27.34% | -12.08% | 2.00% | 101.27% |
| 外資 | 5 | 9.33% | 135.93% | 11.72% | 43.99% | 570.40% |
| 外資 | 10 | 18.10% | 102.96% | 34.46% | 8.42% | 236.11% |
| 外資 | 20 | 22.42% | 76.84% | 7.89% | -0.71% | 163.95% |
| 外資 | 30 | 19.57% | 29.70% | -2.26% | 0.65% | 107.62% |
| 外資 | 40 | 15.69% | 20.94% | -9.23% | 4.76% | 125.65% |
| 外資 | 50 | 13.06% | 17.46% | -8.14% | 3.58% | 107.58% |
| 投信 | 5 | 37.05% | 143.52% | 35.60% | 223.69% | 618.49% |
| 投信 | 10 | 24.86% | 83.73% | 15.08% | 233.34% | 390.73% |
| 投信 | 20 | 26.15% | 87.18% | 31.59% | 200.49% | 321.52% |
| 投信 | 30 | 23.97% | 78.18% | 28.00% | 176.26% | 297.78% |
| 投信 | 40 | 21.06% | 66.56% | 18.60% | 144.98% | 270.15% |
| 投信 | 50 | 20.44% | 62.15% | 19.69% | 123.09% | 222.96% |
| 自營商 | 5 | 15.14% | 22.75% | -20.97% | 76.49% | 104.91% |
| 自營商 | 10 | 24.24% | 28.26% | -18.71% | 54.69% | 96.15% |
| 自營商 | 20 | 17.23% | 30.33% | -19.43% | 27.33% | 69.33% |
| 自營商 | 30 | 14.17% | 15.34% | -11.94% | 21.41% | 53.96% |
| 自營商 | 40 | 14.06% | 14.56% | -11.32% | 14.35% | 35.67% |
| 自營商 | 50 | 9.60% | 13.66% | -11.80% | 12.77% | 35.12% |

(2)CatBoost 模型交易策略結果

表 27 及圖 25 至 28 為 CatBoost 模型搭配各特徵組合預測 5 日後報酬率並每週更換標的之結果，結果顯示投信特徵組合仍為各特徵組合表現最佳，於各檔數之年化報酬率皆高達 50%以上；而自營商表現次之，年化報酬率為 30%-40%左右；而外資和所有特徵雖表現不如其他組合，但亦有 20-30%之年化報酬率。

表 27、CatBoost 模型搭配各特徵組合每週更換標的策略績效

| 特徵組合 | 檔數 | 累積 報酬率 | 年化 報酬率 | 年化 標準差 | 夏普 比率 | 最大 回撤 |
|------|----|-----------|-----------|-----------|----------|----------|
| 所有特徵 | 5 | 180.09% | 22.87% | 25.63% | 0.8924 | -48.21% |
| 所有特徵 | 10 | 330.53% | 33.91% | 19.91% | 1.7029 | -36.33% |
| 所有特徵 | 20 | 335.43% | 34.21% | 16.84% | 2.0319 | -33.61% |
| 所有特徵 | 30 | 332.40% | 34.02% | 15.77% | 2.1568 | -33.06% |
| 所有特徵 | 40 | 303.08% | 32.15% | 15.18% | 2.1176 | -35.02% |
| 所有特徵 | 50 | 292.18% | 31.43% | 14.71% | 2.1361 | -35.08% |
| 外資 | 5 | 256.26% | 28.93% | 28.14% | 1.0281 | -48.17% |
| 外資 | 10 | 318.55% | 33.15% | 21.14% | 1.5683 | -45.59% |
| 外資 | 20 | 201.09% | 24.66% | 17.58% | 1.4026 | -44.56% |
| 外資 | 30 | 231.33% | 27.07% | 16.06% | 1.6854 | -38.04% |
| 外資 | 40 | 219.67% | 26.17% | 15.14% | 1.7281 | -40.35% |
| 外資 | 50 | 191.46% | 23.86% | 14.65% | 1.6285 | -41.73% |
| 投信 | 5 | 692.59% | 51.29% | 24.24% | 2.1160 | -35.91% |
| 投信 | 10 | 706.30% | 51.81% | 21.11% | 2.4541 | -34.57% |
| 投信 | 20 | 1031.32% | 62.45% | 19.25% | 3.2440 | -33.16% |
| 投信 | 30 | 933.81% | 59.55% | 18.35% | 3.2443 | -31.96% |
| 投信 | 40 | 856.37% | 57.08% | 17.81% | 3.2052 | -31.96% |
| 投信 | 50 | 752.95% | 53.53% | 17.36% | 3.0839 | -31.40% |
| 自營商 | 5 | 626.82% | 48.69% | 24.08% | 2.0222 | -36.54% |
| 自營商 | 10 | 710.81% | 51.98% | 20.61% | 2.5221 | -32.26% |
| 自營商 | 20 | 793.25% | 54.95% | 18.76% | 2.9297 | -34.71% |
| 自營商 | 30 | 564.88% | 46.07% | 17.74% | 2.5965 | -35.82% |
| 自營商 | 40 | 506.82% | 43.42% | 17.03% | 2.5491 | -35.89% |
| 自營商 | 50 | 407.89% | 38.41% | 16.53% | 2.3232 | -35.95% |

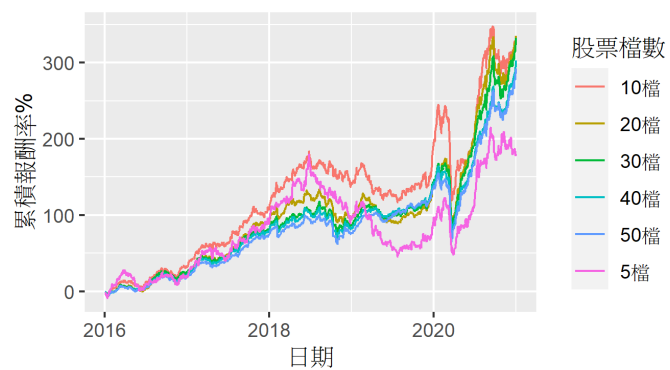


圖 25、CatBoost 模型搭配所有特徵每週更換標的策略累積報酬率圖

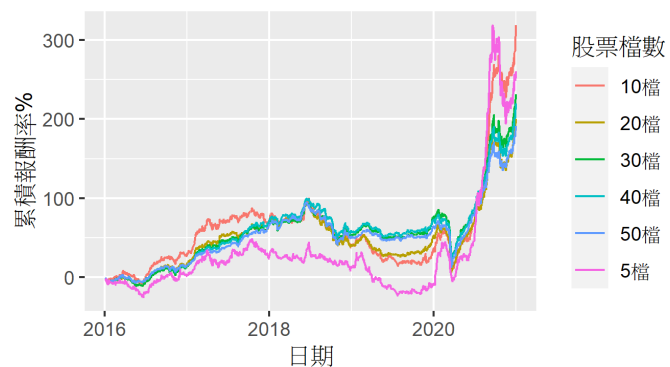


圖 26、CatBoost 模型搭配外資特徵每週更換標的策略累積報酬率圖

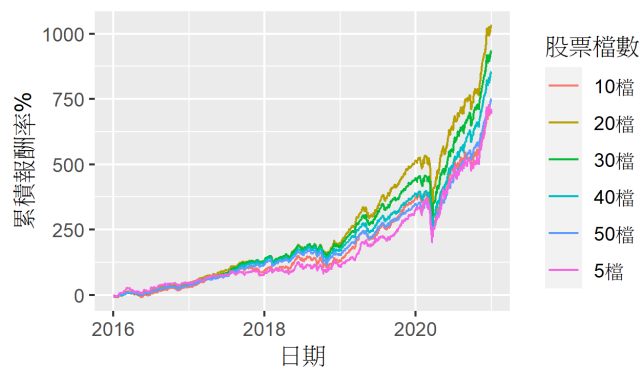


圖 27、CatBoost 模型搭配投信特徵每週更換標的策略累積報酬率圖

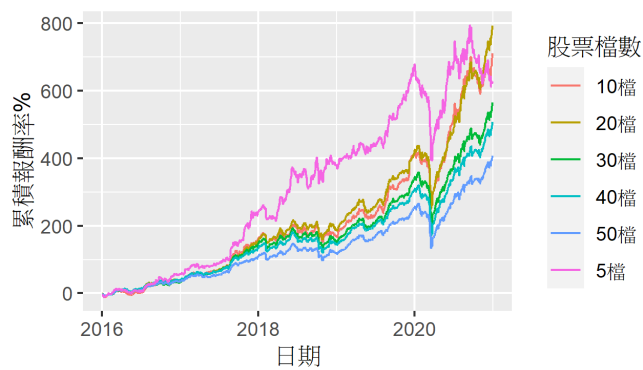


圖 28、CatBoost 模型搭配自營商特徵每週更換標的策略累積報酬率圖

表 28 為 CatBoost 模型搭配各特徵及檔數組合預測 5 日後報酬並每週更換標的之逐年報酬率結果，結果顯示各組合於 2020 年皆有顯著報酬，而投信特徵和自營商特徵於 2019 年和 2017 年有優於其他組合之顯著高報酬；外資特徵則是於 2018 年和 2019 年表現較不佳。

表 28、CatBoost 模型每週更換標的策略各組合逐年報酬率

| 特徵組合 | 檔數 | 2016 | 2017 | 2018 | 2019 | 2020 |
|------|----|--------|---------|---------|---------|---------|
| 加權指數 | - | 10.98% | 15.01% | -8.60% | 23.33% | 22.80% |
| 所有特徵 | 5 | 20.91% | 69.73% | 1.07% | 1.73% | 86.10% |
| 所有特徵 | 10 | 33.45% | 74.25% | 30.39% | 55.51% | 130.88% |
| 所有特徵 | 20 | 23.27% | 67.49% | 3.36% | 42.19% | 194.61% |
| 所有特徵 | 30 | 25.96% | 55.53% | 2.96% | 57.73% | 185.43% |
| 所有特徵 | 40 | 22.46% | 57.06% | 1.51% | 59.04% | 159.11% |
| 所有特徵 | 50 | 19.52% | 54.61% | 1.20% | 60.93% | 153.12% |
| 外資 | 5 | 5.19% | 22.08% | -16.33% | -19.79% | 264.93% |
| 外資 | 10 | 30.66% | 40.13% | -33.22% | -3.15% | 283.14% |
| 外資 | 20 | 18.28% | 48.57% | -28.57% | 12.94% | 148.37% |
| 外資 | 30 | 15.07% | 51.51% | -14.25% | 23.53% | 155.09% |
| 外資 | 40 | 15.21% | 55.72% | -14.41% | 14.98% | 147.22% |
| 外資 | 50 | 15.45% | 49.56% | -15.27% | 13.12% | 126.98% |
| 投信 | 5 | 47.50% | 34.64% | 31.36% | 214.17% | 363.47% |
| 投信 | 10 | 32.88% | 61.89% | 37.76% | 239.82% | 332.14% |
| 投信 | 20 | 47.91% | 80.20% | 70.15% | 314.31% | 516.39% |
| 投信 | 30 | 40.22% | 92.84% | 55.50% | 258.49% | 485.38% |
| 投信 | 40 | 40.39% | 85.57% | 46.84% | 215.14% | 466.44% |
| 投信 | 50 | 37.57% | 79.11% | 38.93% | 191.86% | 403.24% |
| 自營商 | 5 | 59.85% | 182.67% | 138.28% | 296.50% | -51.65% |
| 自營商 | 10 | 40.98% | 122.79% | 5.83% | 242.93% | 286.63% |
| 自營商 | 20 | 40.12% | 114.91% | 31.35% | 224.69% | 368.13% |
| 自營商 | 30 | 34.89% | 111.98% | 5.88% | 182.86% | 219.54% |
| 自營商 | 40 | 41.42% | 92.34% | 11.03% | 158.47% | 196.64% |
| 自營商 | 50 | 37.02% | 72.09% | 9.05% | 133.22% | 150.62% |

(4)綜合比較與小結

比較三種模型各檔數下之結果，和每月更換標的之結果相同，CatBoost 模型中各組合之整體表現為三模型中最佳，而三大法人特徵組合中，投信特徵表現最佳，於三種模型中投信特徵之報酬率皆高於其他特徵組合；以報酬率以及夏普比率和最大回撤等指標綜合評估，於每週更換投資標的之結果中，CatBoost 模型搭配投信特徵組合之績效為各組合中最佳。

而綜合比較每週更換標的和每月更換標的之結果，每週更換標的之報酬率顯著較高，顯示每週更換標的之績效表現較每月更換標的為好，顯示較積極之進出場動作會顯著提高報酬率。

第五章、結論與建議

本研究之結論分為三大部分：第一部分為三大法人交易策略之績效以及每月和每週進場策略比較，探討三大法人中何者之表現較佳，有利於投資人於投資時判斷利用主力資訊做為參考；第二部分為預測誤差及預測準確度探討，綜合比較各組合下之表現以研判最佳之組合；第三部分為後續研究建議。

各交易策略結果中，每月進場策略和每週進場策略報酬率最高之組合皆為投信特徵組合，而投信特徵報酬率優於其他組合之情況，以每週進場策略更為顯著；而每週和每月進場之策略相比，每週進場之策略表現較好，顯示較積極之更換投資標的會帶來較高報酬。而預測誤差和預測漲跌準確率之結果中，LightGBM 模型搭配投信特徵組合於預測 20 日後、10 日後和 5 日後之預測報酬率誤差皆為最小，而漲跌準確度方面，準確度最高之組合分別出現在其他兩模型(CatBoost 和 XGBoost) 搭配投信特徵組合，選取之標的績效表現最好亦為 LightGBM 模型和投信特徵組合之結果，因此整體而言仍推斷 LightGBM 模型和投信特徵為最佳之組合。

後續研究之建議部分，由於本研究建立之特徵組合以法人作為特徵分類之依據為主（即外資相關特徵為同一類、投信相關特徵為同一類、自營商相關特徵為同一類等），後續研究可研究以同性質之特徵作為分類特徵（舉例來說，金額相關特徵為同一類，法人買賣超張數為同一類），或是利用法人指標建立進出場訊號以建立交易策略。

參考文獻

中文文獻

何建緯(2017)，「外資台指期貨未平倉量對於市場報酬率的影響」，國立中正大學財務金融學系碩士在職專班碩士論文。

呂岳暘(2016)，「探討三大法人買賣超對股票報酬之影響」，國立高雄大學國際高階經營管理碩士在職專班(IEMBA)碩士論文。

洪瑞延(2013)，「台灣期貨三大法人和散戶交易行為與報酬關係」，國立中央大學財務金融學系碩士論文。

范聖培(2014)，「三大法人之買賣超行為對股價短期報酬之研究」，國立中央大學財務金融學系在職專班碩士論文。

郭裕涼(2013)，「三大法人籌碼面預測臺灣加權股價指數之研究」，國立高雄科技大學金融資訊研究所碩士論文。

陳彥豪(2002)，「外資與投信法人持股比率變化對股價報酬率影響之研究-以上市電子股為例」，國立中山大學財務管理研究所碩士論文。

陳楷恩(2020)，「三大法人未平倉量、買賣超與台指選擇權波動率指數對台灣指數期貨之影響與關聯性分析」，國立中正大學財務金融學系研究所碩士學論文。

陳冠宇(2020)，「三大法人買賣超對高價股報酬之影響」，國立中央大學產業經濟研究所在職專班碩士論文。

許立豫(2014)，「外資投資行為對股價報酬影響」，國立中正大學財務金融所碩士論文。

莊家睿(2017)，「三大法人資訊對於大盤指數的解釋力及期貨市場的報酬率-以台灣加權指數為例」，國立交通大學財務金融碩士班碩士論文。

黃英杰(2016),「運用三大法人台股期貨未平倉資料於台灣 ETF 交易」,國立中央大學財務金融學系碩士論文。

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研究所碩士論文。

簡笠庭(2019),「台灣股市漲跌對三大法人買賣超與報酬關係之影響」,國立中正大學財務金融研究所碩士論文。

謝鴻安(2014),「三大法人籌碼面預測臺灣加權指數與台股期貨—倒傳遞類神經網路與計量方法之應用」,國立高雄應用科技大學金融資訊研究所碩士論文。

英文文獻

Aggarwal, R., Klapper, L., & Wysocki, P. D. (2005). Portfolio preferences of foreign institutional investors. *Journal of Banking and Finance*, 29(12), 2919-2946.

Andreea Anghel, Nikolaos Papandreou, Thomas Parnell, Alessandro de Palma, Haralampos Pozidis (2018). “Benchmarking and Optimization of Gradient Boosting Decision Tree Algorithms”. *Workshop on Systems for ML and Open-Source Software at NeurIPS 2018*.

Asquith, Paul & Pathak, Parag A. & Ritter, Jay R., (2005). “Short interest, Institutional ownership, and stock returns,” *Journal of Financial Economics, Elsevier*, vol.78(2), 243-276, November.

Chen, T., & Guestrin, C. (2016). “XGBoost: A scalable tree boosting system.” *In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794, ACM.

Chuang Hongwei (2015). “Institutional Ownership and Stock Returns”, *DSSR Discussion Papers No.47*.

Chuang Hongwei (2020). "The impacts of institutional ownership on stock returns," *Empirical Economics, Springer*, vol. 58(2), 507-533, February.

Dreassi A., Miani S., Paltrinieri A. (2015). "Investment Strategies of Institutional Investors: An International Comparison of Sovereign Pension and Social Security Reserve Funds". *Lending, Investments and the Financial Crisis*.

Edelen, Roger M. & Ince, Ozgur S. & Kadlec, Gregory B. (2016). "Institutional investors and stock return anomalies," *Journal of Financial Economics, Elsevier*, vol. 119(3), 472-488.

Essam Al Daoud," Comparison between XGBoost, LightGBM and CatBoost Using a Home Credit Dataset", *World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering* Vol.13, No.1, 2019.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu (2017). "LightGBM: A Highly Efficient Gradient Boosting Decision Tree", *Advances in Neural Information Processing Systems 30* (NIPS 2017).

Kinnander, M. (2020). Predicting profitability of new customers using gradient boosting tree models: Evaluating the predictive capabilities of the XGBoost, LightGBM and CatBoost algorithms. *University of Skövde, School of Informatics*.

Leo de Haan, Jan Kakes. "Momentum or contrarian investment strategies: Evidence from Dutch institutional investors". *Journal of Banking & Finance Volume 35, Issue 9*, September 2011, 2245-2251.

Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, Andrey Gulin (2018). "CatBoost: unbiased boosting with categorical features". *Advances in Neural Information Processing Systems 31* (NeurIPS 2018).

Sami Ben Jabeur, Cheima Gharib , Salma Mefteh-Wali , Wissal Ben Arfi(2021).
“CatBoost model and artificial intelligence techniques for corporate failure prediction”.
Technological Forecasting & Social Change 166.

Sun Xiaolei, Liu Mingxi, Sima Zeqian (2018). “A novel cryptocurrency price trend
forecasting model based on LightGBM”. *Finance Research Letters* Volume 32.