

輔仁大學金融與國際企業學系

# 交易的聖杯-AI 股票預測系統

指導教授：高銘淞

組員姓名：謝鎮陽

何岳桓

李念祖

劉厚謙

蕭亦雯

蘇婉瑄

中華民國一百零九年五月

## 摘要

本專題以 Python 語言建構一個理財機器人系統，用來協助投資人預測股票之漲跌，核心的演算法是利用四種不同機器學習方法彼此競爭，將最佳的預測結果提供給投資人作為投資參考。四種分析方法分別是隨機森林、極限梯度提升、LightGBM、長短期記憶模型，乃是近期金融科技比賽中預測績效較好的方法。在系統的使用上，使用者只需輸入標的股票代號，每個分析法即會自動搜尋該股票對應之 250 種技術指標與歷史股價資料進行預測模型訓練，透過模型的準確率、精準度、召回率等績效指標重複測試找出最佳預測參數，並將模型預測結果傳回給使用者，並提供策略選股、財報分析、歷史資料分析，輔助使用者作出選股決策。在資料處理的方面，本系統包含整體上市櫃公司自 2007 年的歷史資料，並結合網路爬蟲技術，讓資料可以不斷更新。在結果呈現方面，除了用電腦版呈現外，還可使用 Telegram 以聊天機器人的方式在手機上展現成果，讓投資人使用上更加便利。

關鍵詞：股票預測、機器學習、投資機器人

## **Abstract**

Our topic focuses on constructing a stock forecasting system. This system can provide the user with the basic information of stock, stock price prediction, K-line chart, and the individual stock news. In the prediction aspect, our system uses four machine learning models, including Random Forest, XGBoost, LightGBM, and LSTM. To train our machine models, we use 250 different technical indicators as the variables and inputs. In order to make our model become better and more precise, we also used Shap and Skater to observe the reasoning process of the machine learning and improved our models by analyzing observation and changing parameters. We evaluate our models by examining the precision and recall rate and found that LightGBM has the best precision and recall rate.

In addition, in the field of data processing, this study uses the network crawler to extract the share price of listed cabinet companies and combined with MySql access to individual stock data. Finally, we created a Telegram Bot to present our results to users. Users of our bot can enter several commands to our bot and our bot will read the information and return the result to users.

Keywords: stock forecast, machine learning, Robo-Advisor

## 誌謝辭

首先，我們想向專題指導教授 高銘淞 教授致上最高敬意。

在我們仍對主題訂定感到迷茫之初，教授帶領我們腦力激盪，開放足夠思考空間，讓我們自由地發揮想法，引導我們將腦中抽象化為可行具體，階段性給予專業、寶貴建議，細心指導至研究結果，讓我們學習無限，完成的不僅是一份研究專題、一個預測系統，更從中收穫無比成就感。

此外，特別感謝黃弘開學長與楊雅薇老師為學生舉辦「機器學習營」，在這為期一個禮拜的時間中，以淺顯易懂的教學方式，帶領我們踏進機器學習的領域。

在此由衷感謝每位組員，在本次專題研究中，組員們各取所長，包括組織能力、邏輯思考、專業知識、資料整合能力、創意及合作精神，在探究過程中，不僅培養出積極的學習態度，也培養出卓越的合作互助默契。透過此次專題研究的學習，組員們不單是將未來趨勢掌握在手裡，更能整合專業技術及問題處理技巧，相信無論對於未來升學、甚至進入職場，皆有莫大助益。

輔仁大學金融與國際企業學系

謝鎮陽 何岳桓 李念祖 劉厚謙 蕭亦雯 蘇婉瑄

謹誌

中華民國一百零九年五月

## 目錄

中文摘要	i
Abstract	ii
誌謝辭	iii
目錄	iv
表目錄	vi
圖目錄	vii
第一章 緒論	1
第一節 研究背景	1
第二節 研究動機與目的	2
第三節 研究架構與流程	2
第二章 文獻回顧	3
第一節 機器學習(隨機森林、XGBoost、LightGBM)	3
第二節 深度學習(長短期記憶模型)	6
第三節 網路爬蟲	7
第三章 市場現況分析	10
第一節 產業背景	10
第二節 產業分類	11
第三節 與市面上軟體比較	12
第四章 系統建構	15
第一節 資料蒐集與特徵建立	16
第二節 資料預處理	17
第三節 建立預測模型	17
第四節 績效表現	19
第五節 評估優化	19
第六節 結果呈現	26
第七節 額外功能	27
第五章 實際操作方法	28

第一節	Jupyter Lab 實際操作流程-----	28
第二節	Telegram Bot 操作流程-----	42
附錄	-----	47
參考文獻	-----	55

表目錄

表 3-1 與市面軟體比較表-----13

表 4-1 參數優化前-----19

表 4-2 AUC ROC Score 解釋-----20

表 4-3 參數優化後-----22

表 4-4 取出權重排名前 20 高的特徵-----23

## 圖目錄

圖 2-1 決策樹例圖	4
圖 2-2 一般情況示意圖	8
圖 2-3 爬蟲情況示意圖	8
圖 3-1 評分圖	13
圖 4-1 系統建立流程圖	15
圖 4-2 MySQL 示意圖	16
圖 4-3 MinMaxScaler 函數	17
圖 4-4 一次性學習	18
圖 4-5 漸進式學習	18
圖 4-6 loss 和 val_loss	20
圖 4-7 驗證結果	21
圖 4-8 混淆矩陣	21
圖 4-9 特徵重要性	22
圖 4-10 kater 漲跌分析	23
圖 4-11 漲跌強弱河流圖	24
圖 4-12 特徵總體分析	25
圖 4-13 單變量對多變量交互作用	25
圖 4-14 Telegram Bot 流程圖	26
圖 4-15 大立光策略回測	27
圖 5-1 輸入股票代號	28
圖 5-2 選擇預測的天數	28
圖 5-3 即時股價查詢	29
圖 5-4 股價預測	29
圖 5-5 K 線圖	30
圖 5-6 歷史資料概要	31
圖 5-7 大立光歷史虧損幅度前 5 名	32
圖 5-8 大立光策略歷史虧損幅度前 5 名	32
圖 5-9 累積報酬	33



圖 5-10 累積報酬（對數）	33
圖 5-11 與基準比較的累積收益波動率	34
圖 5-12 策略的每日收益	34
圖 5-13 策略滾動 Beta 係數	34
圖 5-14 歷史虧損幅度前 5 名期間	35
圖 5-15 虧損圖表	35
圖 5-16 每月報酬	36
圖 5-17 年報酬	36
圖 5-18 每月報酬分佈	37
圖 5-19 報酬分點	37
圖 5-20 財報評分	38
圖 5-21 策略監控	39
圖 5-22 策略回測	40
圖 5-23 換手報酬	41
圖 5-24 持有清單	42
圖 5-25 自訂回測條件	43
圖 5-26 凱利公式	43
圖 5-27 機器人啟動畫面	44
圖 5-28 機器人提供資訊結果	45
圖 5-29 機器人提供個股頭條	46

# 第一章 緒論

## 第一節 研究背景

在效率市場假說(Fama(1970))中，市場價格將被已揭露的資訊影響，回到應有的水準。投資人無法因預測未來股價而從中取得超額報酬。以弱勢效果(Weak form efficiency)解釋，市場能充分反映出歷史的價格訊息，包含成交價、收盤價、買賣空金額等。台灣股市以散戶為大宗，截至 2019 年底，散戶比例佔投資者中的 60.7%，但其中因投資股票而獲利的比率不到 20%，造成散戶投資失利的原因大致有：散戶往往追求過高的投資報酬率、對標的不熟悉、心理建設不足、容易追高殺低等。綜上所述，大多數人無法從已揭露的公開資訊為自己累積財富，可見台灣股市並不完全符合效率市場假說。

股市分析通常由三個方面組成：基本面、技術面與籌碼面分析。基本面分析是一種透過評估企業價值來預測股價走勢的方式，藉此了解標的公司內部營運、未來潛力與成長性。技術面分析則是透過研究過去的金融市場來預測價格的趨勢，籌碼面則是根據大戶的動向，來決定自己的交易策略。

技術分析的優勢在於可以幫助投資者迅速的判斷買入或賣出的時機、獲得的收益周期短，不必長期持股且能夠用圖像化的方式更好的理解衡量市場風險。根據何培基(2003)、李良俊(2003)、陳惠郁(2018)皆成功證明技術指標能成功挑選出報酬率較高的標的，證明技術標仍具有一定的參考價值。

近年來，機器學習(Machine Learning)逐漸成為趨勢，透過與大數據(Big Data)的結合，不僅能快速分析大量的資料，也可以用於分析和預測，應用在不同領域中。以金融業為例，透過機器學習快速分析大量的資料，如股市歷史資料、客戶資料等，以提供客製化的服務，還能夠透過文字探勘、情緒分析、價格波動等制定不同的投資策略。透過大數據分析的投資比賽如雨後春筍般冒出，可以發現的是，機器學習技術與我們的社會密不可分，如何運用如此龐大的資訊將成為這個時代的新課題。

## 第二節 研究動機與目的

目前市面上的選股軟體，大多為要求使用者輸入自身設定的條件，來選出符合使用者要求的個股，但對於部分不具備金融知識的用戶而言，仍需要花費大量的時間成本學習相關知識，故本研究的目的為透過不同的機器學習演算法分析過去股價歷史資料，找出最適合的模型以判斷股價未來漲跌，並設計出一套讓投資人能夠輕易上手的自動化選股機器人。

## 第三節 研究範圍

於本研究中，我們以台灣上市櫃股票作為分析對象，從 TEJ 擷取上市公司共 943 家，上櫃 777 家自 2007 年 4 月 23 日至 2020 年 5 月 12 日之股價日資料，共 30,940,050 筆，並以 250 種技術指標作為特徵選取，以預測股價為目標撰寫程式。

## 第二章 文獻回顧

機器學習之主要架構是設計出一套演算法，讓電腦自動學習。這類演算法涉及大量統計學理論，能夠將大量資料進行分析與分群，並且利用其中的規律或關係來進行預測。機器學習一般分為監督式、半監督式與非監督式學習。其中，最常見的為監督式學習，即每筆資料上皆會標記標準答案、有預期輸出，監督式學習多用於價格預測。非監督式學習則是給出完全沒有標記過的資料，讓演算法自行尋找出其中的關係，且沒有預期輸出，信用卡盜刷檢測多半都是利用這種方式。半監督式學習則是介於這兩者之間，即資料含有標記與無標記，常用於醫療影像分析。

與傳統回歸相比，傳統回歸較注重於過程，除了預測之外還要能解釋特徵間的關係與顯著性等。機器學習則是重視結果，解釋力稍微薄弱，而準確度、精確度、召回率與 f1-score 都是評判此模型是否良好的標準。

“隨機森林”、“極限梯度提升”、“LightGBM”與“長短期記憶模型”皆為目前主流的監督式學習演算法，在本次研究中也將採用這些演算法進行股價預測和建構預測系統，以下為這四種演算法之介紹：

### 第一節 機器學習

#### 一、隨機森林

決策樹(Decision Tree)是一種過程直覺單純、執行效率也相當高的監督式機器學習模型，它是一個用來處理分類問題的樹狀結構，每個內部節點表示一個評估欄位，每個分支則代表一個可能的欄位輸出，如圖 2-1。

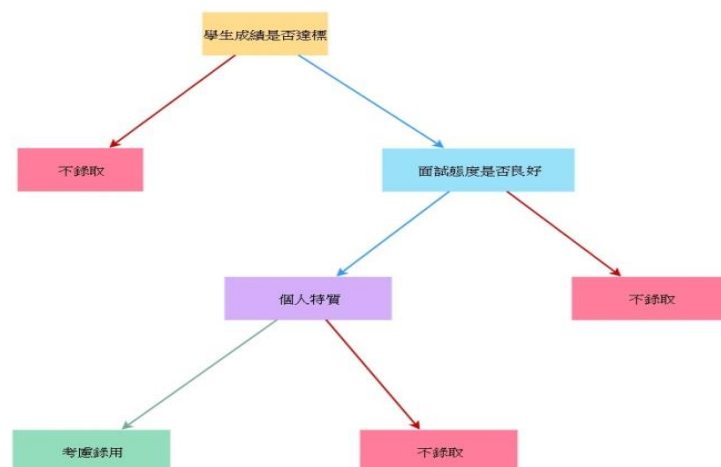


圖 2-1 決策樹例圖

隨機森林則是由許多隨機的決策樹所組成的一種分類演算法，它通過自助法（bootstrap）重取樣技術，從原始訓練樣本集  $N$  中有放回地重複隨機抽取  $k$  個樣本生成新的訓練樣本集合，然後根據自助樣本集生成  $k$  個分類樹組成隨機森林，新資料的分類結果按分類樹投票多少形成的分數決定。在本研究中，我們採用 CART(Classification And Regression Tree)演算為二元樹的技術，此種分類法不受限於應變數與自變數的類型，並以吉尼係數(Gini index)作為屬性的依據。

隨機森林於 2001 年由 Breiman 提出，屬於集成式演算法，Balling & Michel et al.(2015)使用了羅吉斯迴歸(Logistic Regression)、隨機森林、AdaBoost、類神經網路(Neural Network)、支援向量機(Support Vector Machine, SVM)與最佳鄰近法(K-nearest neighbors)等傳統回歸模型和機器學習演算法，並使用 5700 家歐洲公司於 2014 年的股價進行訓練，得到的結果為隨機森林具有最佳的預測準確度。

Patel, J., Shah, S., Thakkar, P., K Kotecha (2014)則比較隨機森林，類神經網路(Artificial Neural Network, ANN)、支援向量機、單純貝氏分類器(Naive Bayes, NB)對印度指數進行預測，以十種技術指標作為特徵選取，再以 2003~2012 年間每日的開盤價、收盤價、最高價、最低價作為訓練資料，並利用「Trend Deterministic Data Preparation」來進行優化，經過非常多參數的比對，發現隨機森林對日資料的預測比其他演算法的績效來的更加優秀。

Khaidem, Saha& Dey(2016)使用隨機森林，以相對強弱指標(Relative Strength Index, RSI)、隨機指標(Stochastic Oscillator, KD)、威廉指標(Williams %R, W%R)、指數平滑異同移動平均線(Moving Average Convergence Divergence, MACD)、價格變動率(Price Rate Of Change, PROC)及能量指標(On Balance

Volume,OBV)作為選取特徵(Feature)，對 APPLE、GE、SAMSUNG 等幾間個股之股價進行訓練並預測三個月後股價，其準確度高達 89%。

## 二、極限梯度提升(XGBoost)

XGBoost 是目前最熱門的機器學習演算法，風靡於各大國內外數據競賽。XGBoost 是 GBTD(Gradient Boosting Decision Tree)演算法的延伸。GBTD 是一種透過組合大量的弱學習者  $f(x)$ ，形成強學習者  $F(x)$ 。弱學習者指得是比亂猜好一點的模型，這種模型的優勢在於訓練成本低，減少計算，不容易過度擬和。當我們使用大量的弱學習者作為基底組合成一個強學習者，可以更好的提升預測的精準度，卻又有原本弱學習者的特徵。在 GradientBoosting 算法中，關鍵就是利用損失函數的負梯度方向在當前模型的值作為殘差的近似值，進而擬合一棵 CART 回歸樹。GBDT 會累加所有樹的結果，而這種累加是無法通過分類完成的，因此 GBDT 的樹都是 CART 回歸樹，而不是分類樹。

XGboost 的性能又比 GBTD 有更進一步的提升，它可以自動運用 CPU 的線程進行運算。

GBDT 以 CART 樹作為基礎學習器，XGBoost 還支持線性分類器，這個時候 XGBoost 相當於 L1 和 L2 正規化的羅吉斯迴歸，且傳統的 GBDT 在優化的時候只用到一階導數信息，XGBoost 則對代價函數進行了二階泰勒展開得到一階和二階導數，降低了模型的方差，使學習出來的模型更加簡單。

Shubharthi, D., Snehanthu, S., Yash, K., and Suryoday, B. (2016)透過 XGBoost 對 Yahoo 與 APPLE 的股價進行預測，針對不同天期：28 天、60 天、90 天期進行比較，以 RSI、KD、威廉指標、PROC 作為特徵選取，從實驗結果發現不論是 28 天、60 天或 90 天期，訓練區間越長，其均方根誤差(Root Mean Square Error, RMSE)越小，且預測天期越短，RMSE 也會越小、準確度越高。並將該篇研究結果與其他學者的論文在同樣環境下進行比較，發現 XGBoost 的績效遠大於類神經網路、隨機森林、支援向量機，與羅吉斯迴歸，有將近 99%的準確度。

Suryoday Basaka, Saibal Kar(2019)則是透過隨機森林與 XGBoost，以 RSI、KD、MACD、PROC、OBV 等技術指標作為特徵，並對 APPLE、FB、AMAZON、NIKE 同樣進行不同時長的預測，實驗結果發現選取的時間範圍越大，不論是隨機森林或 XGBoost 的準確度都比短期預測來的精準，尤其以 90 日來的最佳，F-score 皆高達 0.85。

## 三、LightGBM

LightGBM 是 GBDT 的輕量化版，GBDT 與 XGBoost 皆是增加樹的層級(Level-wise)作為演算法，而 LightGBM 則是選擇具有最大誤差的樹葉進行生長

(Leaf-wise)，以這種方法可以大幅度降低電腦所需的效能。並利用基於直方圖均衡化(histogram)的演算法，通過將連續特徵（屬性）值分段為間斷數值(discrete bins)來加快訓練的速度並減少記憶體的使用，且速度更快，更容易處理大規模資訊。而在準確度方面，當輸入資料大於 10000 筆時的預測準確度與 XGBoost 不分軒輊。再大規模數據分析上，LightGBM 不僅幾乎擁有 XGBoost 所有的功能，還擁有更快的速度與更好的優化。缺點則是容易產生過擬合(Overfitting)的情況。

### (一) histogram

#### 1. 介紹

直方圖算法的基本思想是先把連續的浮點特徵化成 k 個整數，同時建構一個寬度為 k 的直方圖，在遍歷(traversal)數據時，以離散化的值為索引，在直方圖中累積統計量，當遍歷一次數據後，直方圖累積了需要的統計量，然後根據直方圖的離散值，遍歷尋找最優的分割點。

#### 2. 優勢

(1). 在二叉樹中可以通過利用葉節點的母節點和相鄰節點的直方圖的相減來獲得該葉節點的直方圖，所以僅僅需要為一個葉節點建立直方圖就可以通過直方圖的相減來獲得相鄰節點的直方圖，而這花費的代價極小。

(2). 可以將連續的值替換為間斷數值。可以利用較小的資料型別來儲存訓練資料，如 uint8\_t。無需為預先分類的特徵值儲存額外的資訊。

## 第二節 深度學習

### 一、長短期記憶(Long Short-Term Memory, LSTM)

遞歸神經網路(Recurrent Neural Networks, RNN)之長期記憶不足，因此在長期記憶的表現上較差，而長短期記憶就是設計用來改善遞歸神經網路之長期記憶表現不佳問題(Sepp Hochreiter and Jurgen Schmidhuber (1997))。Xiong et al.(2016)使用長短期記憶和其他傳統模型，預測 S&P500 之波動率走勢，結果顯示長短期記憶在誤差表現上明顯優於傳統模型；而 Deng Yue et al.(2017)討論了遞歸神經網路、卷積神經網路 (Convolutional Neural Networks, CNN)以及長短期記憶等深度學習模型的績效表現，其中以長短期記憶的績效最佳。

Murtaza Roondiwala, Harshal Patel and Shraddha Varma (2015)使用不同的實驗對照組，包括只使用開盤價、收盤價和最高價、最低價、收盤價以及最高價、最低價、開盤價和收盤價等三組，並且使用長短期記憶模型預測股價，再將每一組

區分成 250 和 500 epochs 兩種，一共六組進行預測，並以均方根誤差比較各組預測績效，其結果顯示每一組的均方根誤差都非常小，其中誤差最小的為使用最高價、最低價、開盤價和收盤價並且使用 500 epochs 的實驗組，訓練集和測試集誤差僅有 0.00983 和 0.00859，因此本研究欲建構之預測系統將取最高價、最低價、開盤價和收盤價以建構技術指標參數。

David M. Q. Nelson, Adriano C. M. Pereira, Renato A. de Oliveira (2017) 主要使用長短期記憶和多層感知器 (Multilayer perceptron, MLP)、隨機森林以及偽隨機數產生器 (Pseudo Random Number Generator, PRNG) 進行比較，針對 BOVA11、BBDC4、CIEL3、ITUB4 及 PETR4 等五檔個股進行預測，其結果發現長短期記憶預測之準確度明顯優於其餘三者。然而 Fischer Thomas and Krauss Christopher (2017) 比較長短期記憶、深度神經網路 (Deep Neural Network, DNN)、隨機森林和羅吉斯迴歸等模型之預測績效，結果顯示雖然在時間區間較短時，長短期記憶之績效較其他三者佳，但是隨著時間區間拉長，長短期記憶和隨機森林之回報率將逐漸相等，並且兩者標準差雖然皆較其餘兩者更小，但隨機森林仍稍微優於長短期記憶，可知長短期記憶並非一定優於隨機森林等機器學習模型。

本研究欲建構一預測系統，而就目前的文獻而言，每種演算法皆有其適合的配置，因此不容易抉擇出哪一項演算法能最符合本研究之需求，於是在本研究中將會分別使用隨機森林、XGBoost、LightGBM 以及長短期記憶等四種不同的演算法建構預測系統和進行預測。

### 第三節 網路爬蟲 (Web Crawler)

網路爬蟲又名網路蜘蛛，是一種透過程式主動獲取網路資訊的技術，通過網頁的連結地址來尋找網頁，讀取該網頁的資料並存取，從中再找出其他連接地址，再透過新的連結地址找下一個的網頁，然後一直迴圈下去直到依某種策略將網路上能獲取的網頁抓取完。一般而言，我們瀏覽網站是對伺服器發出指令 (request)，伺服器接受指令後回傳網站原始碼，瀏覽器再將這些原始碼組成網站形式，如圖 2-2。而爬蟲則是利用程式語言對伺服器發出 request，再從回傳的原始碼中抽取需要的資訊存入資料庫中以備後續使用，如圖 2-3。



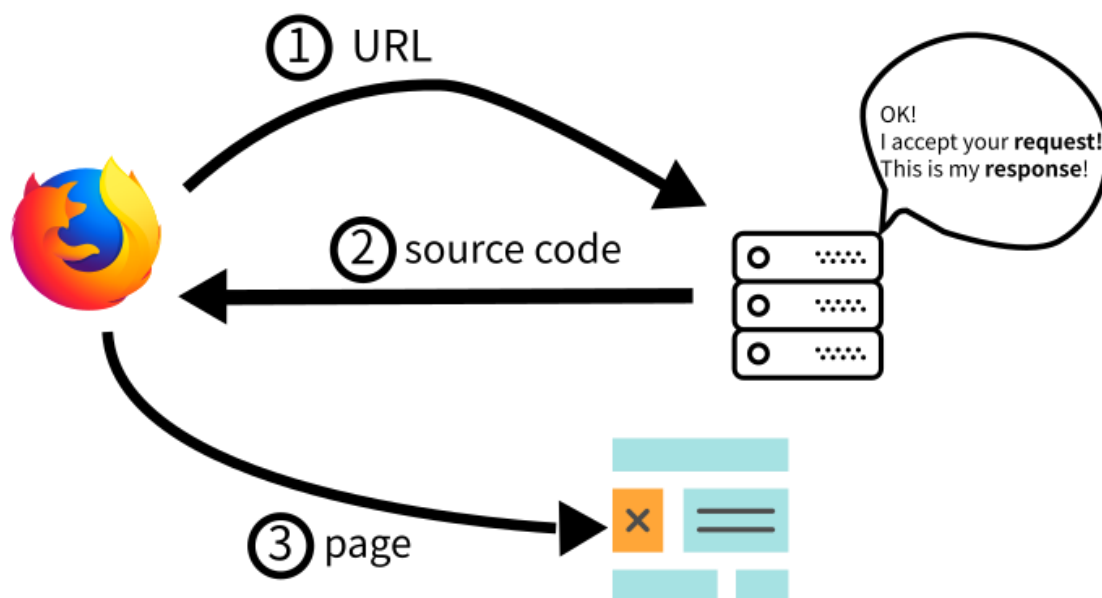


圖 2-2 一般情況示意圖

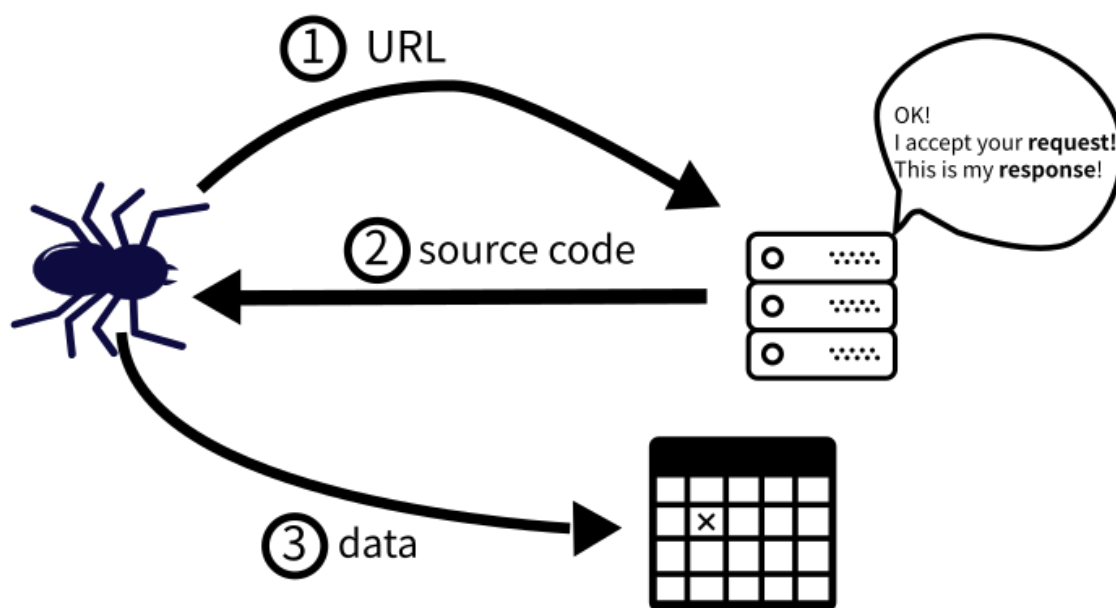


圖 2-3 爬蟲情況示意圖

網路爬蟲的流程是透過一個擷取一個原始 URL，又稱為種子，再根據種子中的超連結將其他 URL 放入佇列，並不斷重複形成迴圈，直到滿足停止條件為止。在傳回的響應內容 (html) 中，會帶有 css、js、圖片等 url 地址，以及 ajax 代碼，瀏覽器按照響應內容中的順序依次發送其他的請求，並獲取相應的響

應內容，所有被爬蟲抓取的網頁將會被系統存貯，進行一定的分析、過濾，並建立索引，以便之後的查詢和檢索。

## 第三章 市場現況分析

### 第一節 產業背景

隨著人工智慧不斷地進步，機器學習技術也不斷發展，機器學習開始被廣泛的運用在各個領域，像是在金融界中最普遍的便是預測金融商品價格的走勢，例如：預測股價漲跌、衍生性金融商品價格趨勢等等，又或是應用在 P2P 借貸上，可以建構信用評估模型來評估和預測借款人違約風險等等。機器學習同樣能夠使用在其他不同領域上，例如：醫學界中，分析回診病患的檢驗資料，可以預測檢驗者罹患疾病的機率，或是透過分析病患病歷資料，來判斷治療有效與否。更有將機器學習和其他技術結合的研究，像是使用機器學習建構智能投資，再與區塊鏈技術結合形成一個區塊鏈平台，最後透過遊戲化概念包裝以吸引更多參與者使用的新興應用方式。

在機器學習不斷發展的同時，機器學習也衍生出一個新的分支—深度學習，深度學習是近年來成果最佳、最令人期待的技術，深度學習有著以下應用。首先是影像辨識，目前深度學習在影像辨識上已經有十分成熟的技術，可以使用在醫學、媒體、零售、交通…等方面上，例如：在醫學界中，已經有著和醫生同等甚至超越醫生的能力，透過分析檢驗對象的資料可以更快且更準確的判斷檢驗者是否罹患疾病；或是最近很熱門的換臉技術、人臉支付，同樣是深度學習在影像辨識上的應用之一。接著則是語音辨識，和影像辨識相同，深度學習在語音辨識上也已經非常成熟，像是應用在醫療診斷上，可以透過分析聽診器聲音的特性或是臨床病患呼吸聲音，來幫助醫生進行診斷。最後則是自然語言處理，藉由深度學習讓 AI 學習和處理與人類語言有關的技能，使 AI 能夠理解人類寫的文字和說的話語，其中最具代表性的應用便是聊天機器人。

#### 一、金融業對機器學習之應用

金融業對於機器學習的應用主要在於預測金融市場。金融市場有許多不確定因素，機器學習的出現使得人們能以更低的成本從繁雜的資料中分析出重要有意義的結果，並協助投資者建立更好的投資組合。

目前常見的標的有道瓊股市、標普指數、NASDAQ、台股、台指期等。這些標的皆行之有年，資料更新快速且容易取得，也是本國投資人主要進攻的市場。

演算法的部分則以類神經網路，基因演算法，隨機森林與 XGBoost 為大宗。類神經網路與基因演算法屬於深度學習的範圍，隨機森林與 XGBoost 則是傳統機器學習的範疇，而深度學習具有以下優點：

1. 深度學習的網路神經層數多，寬度廣，理論上能反映到任意函數，所以能解決複雜問題。
2. 深度學習高度依賴數據，數據量越大，表現越好，在圖像識別等任務甚至能超越人類的表現。

深度學習看似能夠更好的對大數據進行分析，但實際上，並無任何證據顯示其在金融市場預測的績效優於傳統機器學習。兩派皆有不少支持者。

隨著資訊科技的發展，網路上所乘載的資料量越來越大，除了上述方法，我們也可以透過網路搜尋量，新聞文字，社群討論度等資料來進行預測。文字探勘則是針對這些資訊所產生的技術。

文字探勘是一種跨領域的應用，結合資料探勘與自然語言處理，資訊檢索技術等。透過特徵萃取，分類與詞頻等，並區別出文本的類別與屬性，找出詞彙間的群集關係。透過這個技術，投資者不僅能夠利用傳統的基本面、技術面等因素，且能透過分析財經新聞、論壇上的資訊與情緒指標，提升趨勢預測的準確度。

## 第二節 產業分類

European Supervisory Authorities (ESAs) 將「自動化財務建議」(automation in financial advice) 定義為：「消費者使用自動化工具（一般為網頁）以接收財務建議（例如推薦買進或賣出特定財務商品），全無（或極少）人力之參與介入（human intervention）」，其中自動化財務建議即為機器人投顧 (Robo-Advisor)。EFAs 認為 Robo-Advisor 應有三大主要特徵：

1. 消費者可自行使用機器人投顧，全然無須他人之參與或僅需有限之參與
2. 演算法可依據投資人所輸入資訊而運算結果
3. 機器人投顧系統之產出為一定財務建議

洪志麟、賴冠妤、陳國瑞、呂馥伊(2016)提到若是以提供的主要服務分類，則可將機器人投顧概分為「諮詢建議型」、「資產管理型」兩種：

1. 諮詢建議型：

即透過自動化服務提供投資建議和投資組合選擇，取代傳統的證券投顧服務。

2. 資產管理型：

除提供諮詢建議服務外，進一步協助客戶就投資組合提供交易執行及風險管理服務，服務項目包括投資組合再平衡、獲利再投資、稅務虧賣等等。

我國由於法規限制，目前還未像國外發展出完全擺脫人力介入的機器人投顧服務，主要是由投顧公司透過自動化工具，提供投資人投資顧問的服務，然而真正的執行者仍是投資者自己，也就是說目前的機器人投顧還只是半自動化。近年來也開始有類似於歐美地區的機器人投顧公司，像是大拇哥、阿爾發等等，但是同樣受限於法規的限制，目前未能如國外完全的自動化，例如若是要全權委託全自動再平衡，門檻高達 500 萬台幣，低於此門檻之投資人需自行執行。

除了典型的機器人投顧外，台灣市場上目前還有幾種類似於機器人投顧的服務，大略可分為以下兩種：

#### 1. 線上理財篩選試算工具

提供每日的投資理財或金融訊息和投資篩選工具、試算程式供使用者運用，使用者輸入基本資料後，網站便提供推薦的股票或基金投資清單，相關基本服務不需收取費用，倘若要使用進階功能則需加入會員或收費。

#### 2. 看盤軟體工具

透過即時更新及分析資訊，提供使用者買賣股票之具體建議，除彙整每日股市資訊外，亦會透過電腦程式進行技術分析，提供用戶進出場時機、張數以及股價漲跌走勢等分析意見，並可以線上立即下單，但是多數均需要額外付費購買。

### 第三節 與市面上軟體比較

本研究所建構之預測系統屬於看盤軟體工具一類，挑選出市場上同屬於看盤軟體工具類型，並將本系統與這些軟體工具進行比較，再進行評分、繪製出本預測系統與其他軟體工具之產品特色差異。

根據市場上的看盤軟體挑選出五套軟體，分別為 XQ 操盤高手、HUNTA 智慧型股票、轟天雷選股、法拉利贏家、Money A+，並針對這五套軟體與本預測系統，選擇出主要的特色與功能，分別為即時股價、預測系統、策略回測、基本面評分、新聞資訊、API 手機端、自選策略與標的多樣性，再針對這些功能進行比較分析，整理出下列表 3-1 並進行評分，◎代表程式具備該項功能，而✕則表示不具備該項功能。

表 3-1 與市面軟體比較表

	本預測系統	XQ 操盤高手	HUNTA 智慧型股票	轟天雷選股	法拉利贏家	Money A+
即時股價	◎	◎	◎	◎	◎	◎
預測系統	◎	×	◎	◎	×	◎
策略回測	◎	◎	×	×	×	×
基本面評分	◎	×	×	×	×	×
新聞資訊	◎	◎	◎	◎	◎	◎
API 手機端	◎	◎	×	×	◎	×
自選策略	×	◎	◎	◎	◎	◎
標的多樣性	×	◎	◎	◎	◎	◎

評分的部分，我們以表 3-1 作為評分標準，並且將其分為兩大類，以預測/回測功能為主，其他功能為輔。預測/回測功能以預測的準確度、直觀性為做綜合評分，其他功能則以◎和×的數量進行評分，所得結果如圖 3-1。

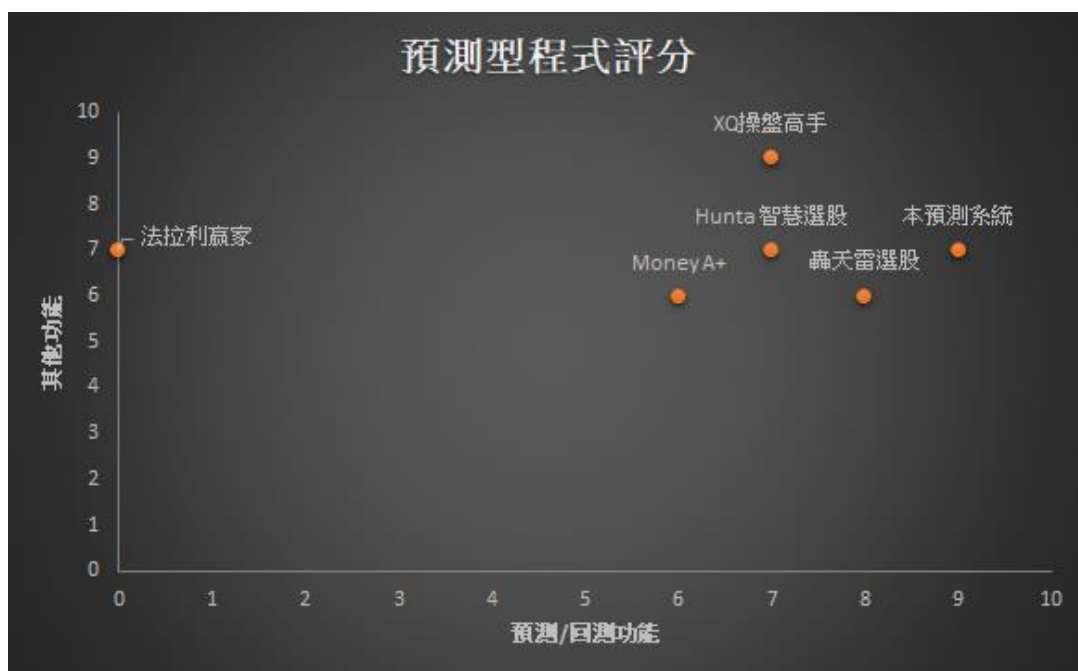


圖 3-1 評分圖

可以發現本研究所建構之預測系統在預測/回測的功能上是較優於其他市面上之預測程式，但是在其他功能上則是 XQ 操盤高手較為高分，然在綜合評分上，本研究之預測系統和 XQ 操盤高手分數相近。

## 第四章 系統建構

本系統之建構共分成七個步驟：第一步為資料蒐集與特徵建立，標的為國內上市上櫃的所有股票資料，並建構 250 種技術指標，第二步為資料預處理，針對遺失值填補及特徵標準化，第三步則是建立預測模型，透過機器學習和深度學習，處理大量資料和變數及建立適當模型，第四步績效表現，從模型中得知預測的準確率，第五步評估優化，第六步呈現結果，第七步附加功能。以下分別就每個步驟進行說明：



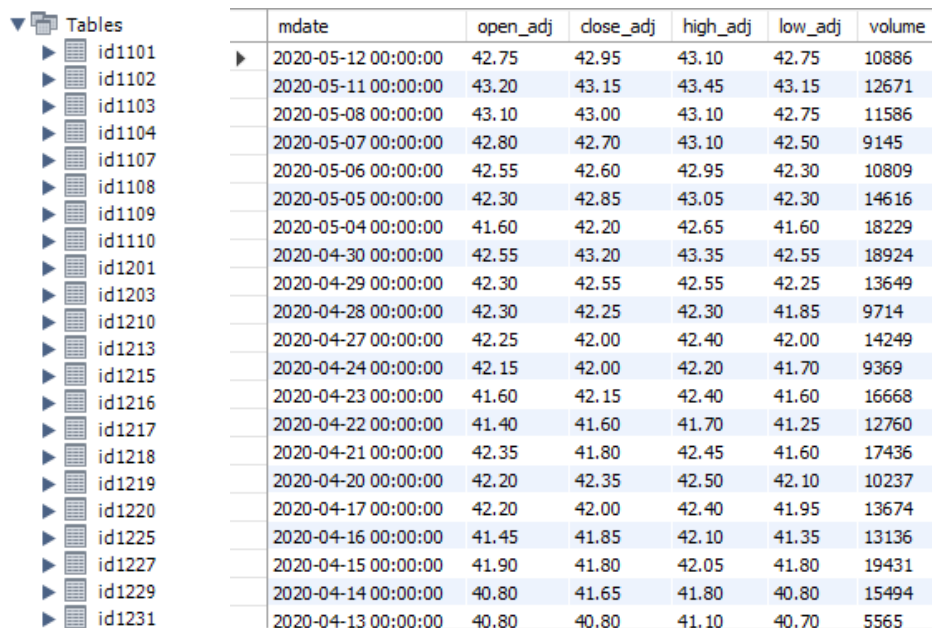
圖 4-1 系統建立流程圖



## 第一節 資料蒐集與特徵建立

本系統研究資料為國內上市上櫃的所有股票資料，資料來源為台灣證券交易所。從金融海嘯時期可看出整體經濟循環的過程，故資料期間為西元 2007 年 4 月 23 日至 2020 年 5 月 12 日，資料頻率為日資料。利用 Python 爬蟲爬取資料並加以整理，取出所有股票的開盤價、收盤價、最高價、最低價、成交股數，並依照爬取下來的財報資料，將個股的資料依照個股的除權息資料，轉換成還原股價，並利用 Python-TA-Lib 建構出 250 個技術指標作為我們預測的特徵。

資料存取的部分使用 MySQL 儲存所有上市櫃公司的開盤價、收盤價、最高價、最低價以及成交股數，Schemas 為每隻股票個別創立一個 Table，如圖 4-2 並將資料按照時間先後順序進行儲存。取用資料則使用 Python-Sqlalchemy 套件將 MySQL 上的个股資料一一節錄並轉換成資料集供接下來的程式計算使用。



	mdate	open_adj	close_adj	high_adj	low_adj	volume
id1101	2020-05-12 00:00:00	42.75	42.95	43.10	42.75	10886
id1102	2020-05-11 00:00:00	43.20	43.15	43.45	43.15	12671
id1103	2020-05-08 00:00:00	43.10	43.00	43.10	42.75	11586
id1104	2020-05-07 00:00:00	42.80	42.70	43.10	42.50	9145
id1107	2020-05-06 00:00:00	42.55	42.60	42.95	42.30	10809
id1108	2020-05-05 00:00:00	42.30	42.85	43.05	42.30	14616
id1109	2020-05-04 00:00:00	41.60	42.20	42.65	41.60	18229
id1110	2020-04-30 00:00:00	42.55	43.20	43.35	42.55	18924
id1201	2020-04-29 00:00:00	42.30	42.55	42.55	42.25	13649
id1203	2020-04-28 00:00:00	42.30	42.25	42.30	41.85	9714
id1210	2020-04-27 00:00:00	42.25	42.00	42.40	42.00	14249
id1213	2020-04-24 00:00:00	42.15	42.00	42.20	41.70	9369
id1215	2020-04-23 00:00:00	41.60	42.15	42.40	41.60	16668
id1216	2020-04-22 00:00:00	41.40	41.60	41.70	41.25	12760
id1217	2020-04-21 00:00:00	42.35	41.80	42.45	41.60	17436
id1218	2020-04-20 00:00:00	42.20	42.35	42.50	42.10	10237
id1219	2020-04-17 00:00:00	42.20	42.00	42.40	41.95	13674
id1220	2020-04-16 00:00:00	41.45	41.85	42.10	41.35	13136
id1225	2020-04-15 00:00:00	41.90	41.80	42.05	41.80	19431
id1227	2020-04-14 00:00:00	40.80	41.65	41.80	40.80	15494
id1229	2020-04-13 00:00:00	40.80	40.80	41.10	40.70	5565

圖 4-2 MySQL 示意圖

## 第二節 資料預處理

利用圖 4-3 Python-SKlearn-MinMaxScaler 將建構的 250 種技術指標，進行特徵縮放，將所有的技術指標收斂在 0 到 1 之間，並剔除已下市的股票，以及我們將有空值的資料以近一年之收盤價的平均值進行補值。

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

圖 4-3 MinMaxScaler 函數

## 第三節 建立預測模型

本研究主要以機器學習和深度學習進行建模，機器學習包含隨機森林 (Random Forest)、極限梯度提升 (Extreme Gradient Boosting)、LightGBM，上述方法皆為機器學習的分類，而深度學習則是長短期記憶 (Long Short-Term Memory)。

預測目標的建立：將明日還原後的開盤價/今日的還原後開盤價，並將漲設定為下一個交易日後的調整開盤價比今日開盤價高，跌設定為下一個交易日後的調整開盤價比今日開盤價低，並將兩種答案轉換成布林值 True 與 False，讓機器清楚辨識，並將技術指標與預測目標整合為新的資料集。

選擇預測模型：我們採用了四種預測模型，1. Random Forest (隨機森林)、2. XGBoost (Extreme Gradient Boost 極限梯度提升)、3. LightGBM、4. Long Short-Term Memory (長短期記憶)。特徵使用：本研究建構的 250 種技術指標，詳細指標參考附錄，機器學習採用一次性學習，一次性學習介紹如圖 4-4，切分資料：以 2007 年 4 月 23 日至 2017 年 12 月 31 日作為我們的訓練集，2018 年 1 月 1 日至 2020 年 5 月 12 日作為我們的測試集，資料頻率為日資料，深度學習採用漸進式學習，漸進式學習介紹如圖 4-5，切分資料：以 2007 年 4 月 23 日至 2015 年 12 月 31 日作為我們的訓練集及驗證集，2016 年 1

月 1 日至 2020 年 5 月 12 日做為我們的測試集資料頻率為三日一維。最後我們以大立光(3008)作為我們的測試標的。

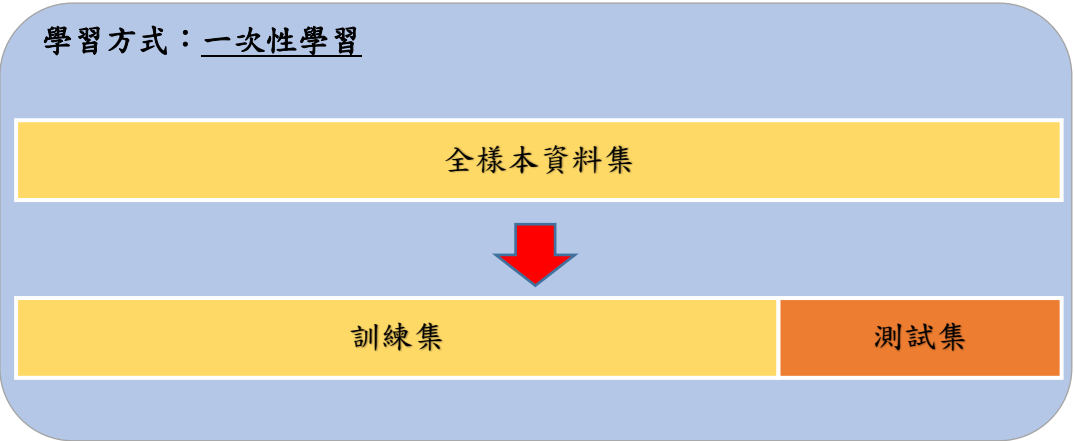


圖 4-4 一次性學習

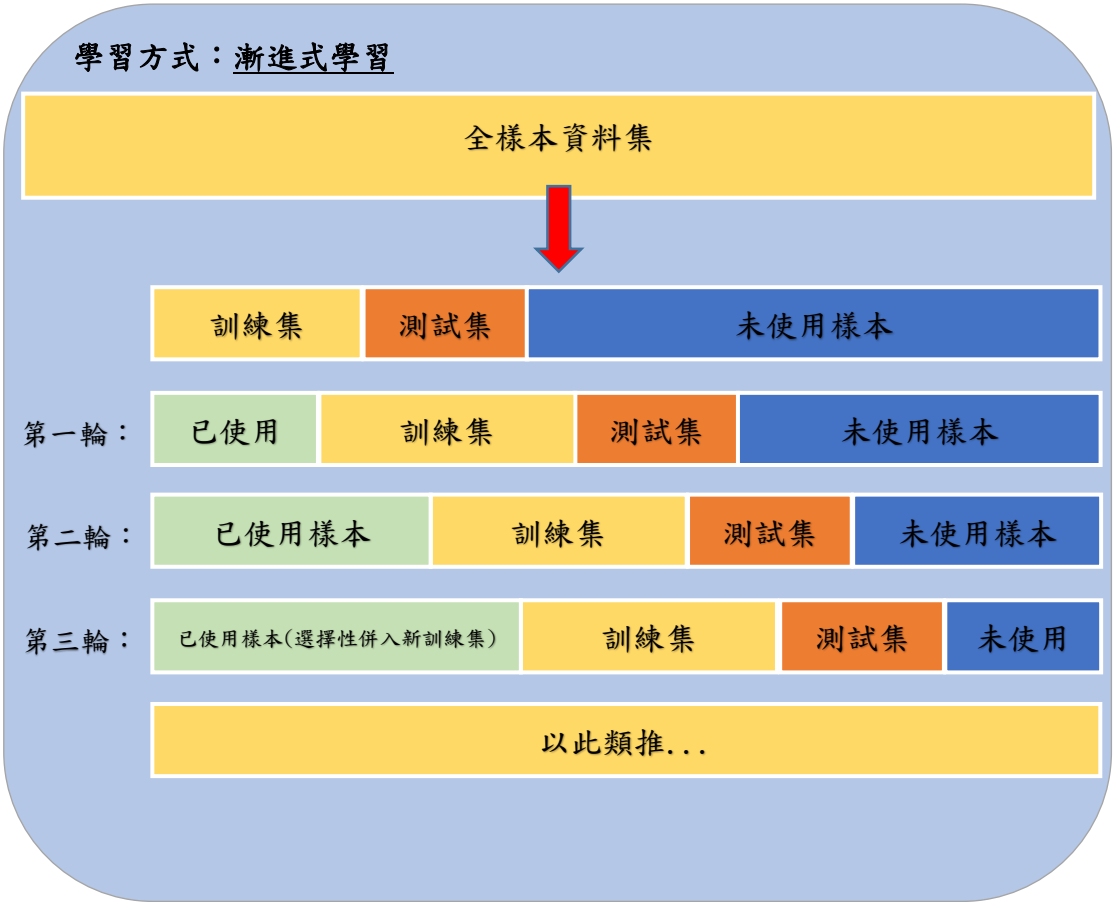


圖 4-5 漸進式學習

## 第四節 績效表現

表 4-1 為未經過參數優化及變數調整前，預測模型的績效表現。

表 4-1 參數優化前

Classification Report			
特徵 250 個	Random Forest	XGBoost	LightGBM
Accuracy	0.77	0.81	0.80
Precision (漲)	0.81	0.85	0.84
Precision (跌)	0.73	0.77	0.76
Recall	0.77	0.81	0.80

## 第五節 評估優化

一次性學習，利用 Python-SKlearn-GridsearchCV 進行參數優化，並利用 AUC ROC Score(Area under the Curve of ROC)表 4-2 作為參數優化的衡量標準，來避免過擬合。

漸進式學習，LSTM 的 Activation Functions 使用 Sigmoid Function、Relu，以均方差(Mean-Square Error, MSE)作為衡量準確性的依據，並調整 Batch\_Size、Epoch、dropout 來最適模型的優化程度和速度，圖 4-6 可以看到 loss 跟 val\_loss 都是持續下降，圖 4-7 驗證結果按照預測結果進行持有，證明模型是有學習到，但 acc 上升而 val\_acc 卻維持在 7 成左右，代表可能有些許過擬合，以至於後面的結果不是太好。

表 4-2 AUC ROC Score 解釋

AUC 判斷預測模型優劣的標準	
AUC = 1	完美分類預測模型，採用這個預測模型時，存在至少一個閾值能得出完美預測，但絕大多數預測的場合，不存在完美分類器。
$0.5 < \text{AUC} < 1$	優於隨機猜測。這個分類器（模型）妥善設定閾值的話，能有預測價值。
AUC = 0.5	跟隨機猜測一樣（例：丟銅板），模型沒有預測價值。
AUC < 0.5	比隨機猜測還差；但只要總是反預測而行，就優於隨機猜測。

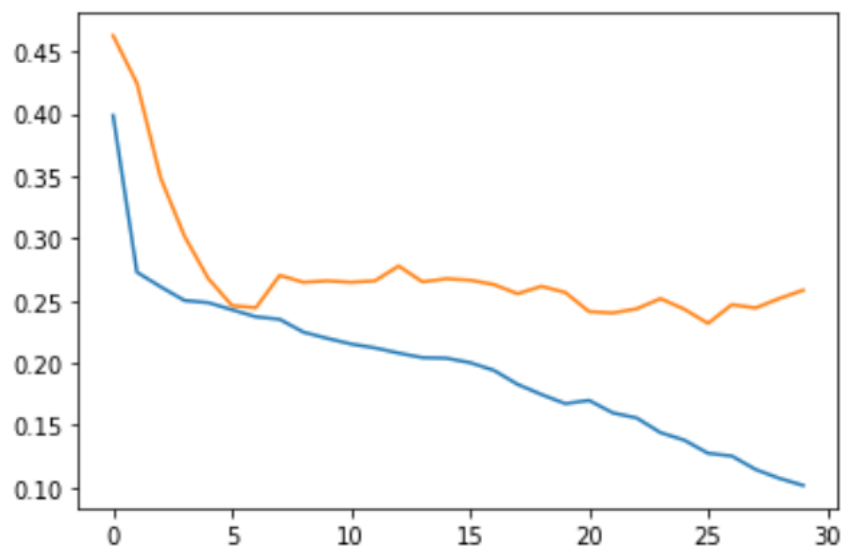


圖 4-6 loss 和 val\_loss

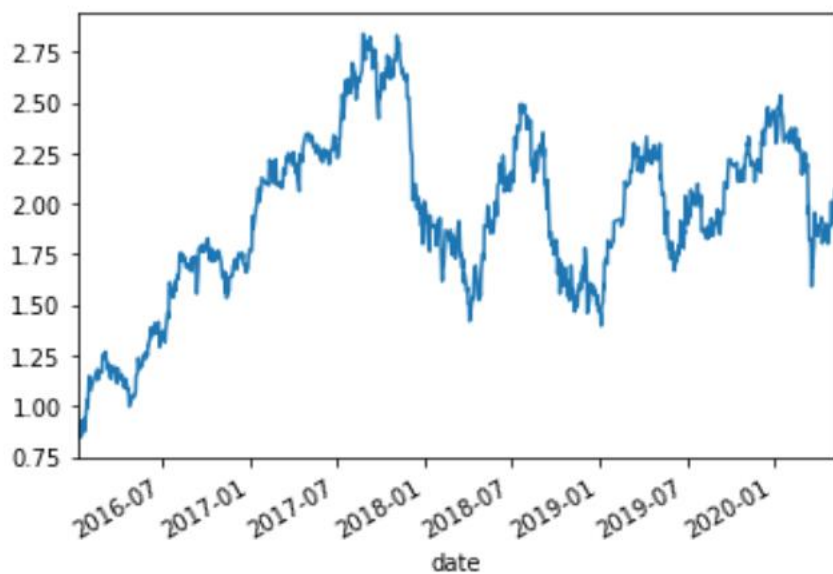


圖 4-7 驗證結果

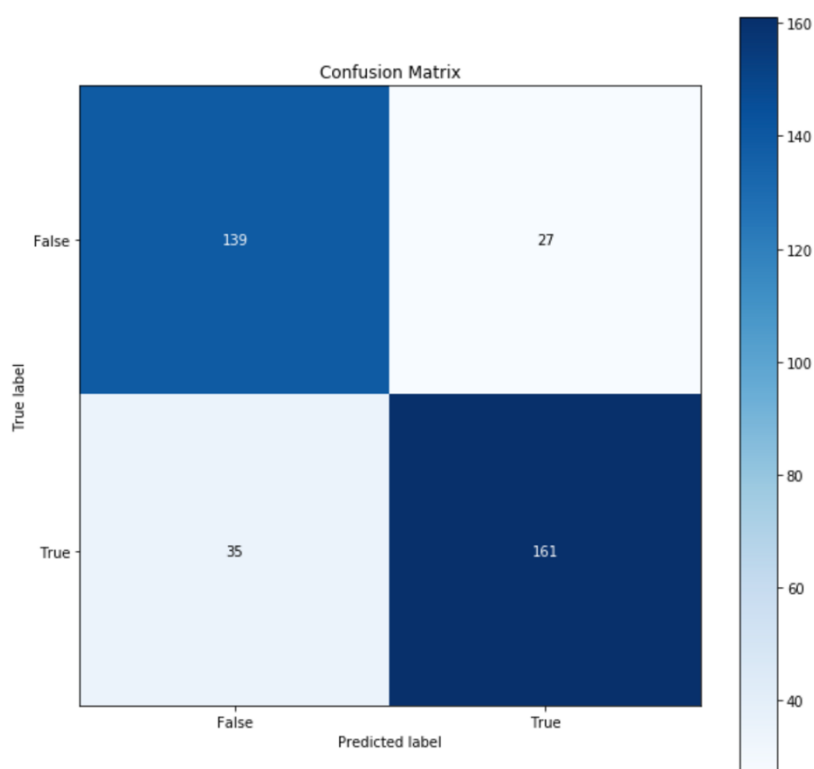


圖 4-8 混淆矩陣

表 4-3 為我們觀察圖 4-8 混淆矩陣及利用表 4-2 AUC ROC Score 來優化模型參數之後的結果，我們對於每隻個股，進行客製化的參數優化，使得每一隻的

股票都有一個對應優化過的參數，並透過圖 4-9 特徵重要性從 250 種技術指標選出權重前 20 高的技術指標，作為模型的新資料集，並比較新舊資料集精準度的差異，表 4-4 的結果發現精準度差異不大，但實務上應用使用 20 個特徵會更適合操作。

表 4-3 參數優化後

Classification Report			
特徵 250 個	Random Forest	XGBoost	LightGBM
Accuracy	0.76	0.82	0.80
Precision (漲)	0.79	0.84	0.83
Precision (跌)	0.72	0.80	0.77
Recall	0.76	0.82	0.80

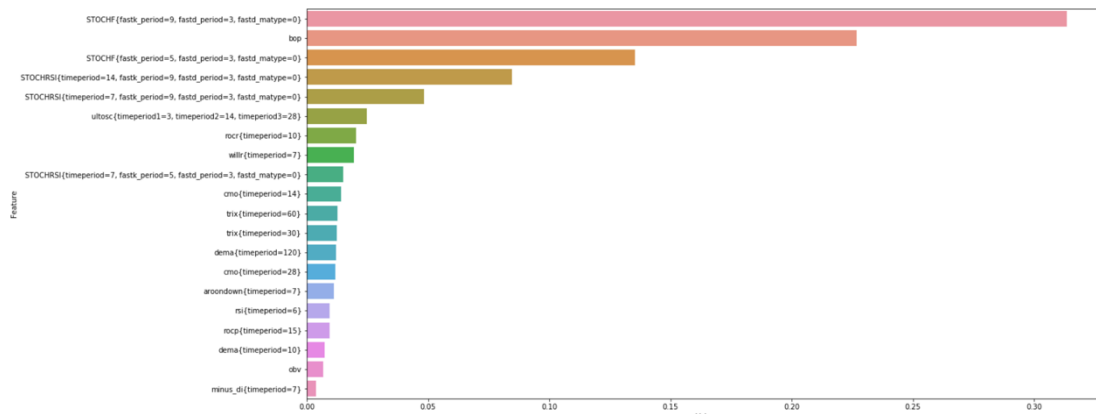


圖 4-9 特徵重要性

表 4-4 取出權重排名前 20 高的特徵

Classification Report			
特徵 20 個	Random Forest	XGBoost	LightGBM
Accuracy	0.79	0.81	0.83
Precision (漲)	0.82	0.83	0.85
Precision (跌)	0.77	0.80	0.81
Recall	0.79	0.81	0.83

進一步地利用 Python-Skater、Shap 可視化模型來視覺化觀察機器學習及深度學習如何做決定，以大立光（3008）這檔股票作為範例，將漲設定為 1、跌設定為 0。下圖 4-10 為大立光在 2020 年 4 月 29 日開盤價公佈後，本研究之預測模型預測隔日的開盤價為漲，預測模型所參考的指標，以及指標的權重。

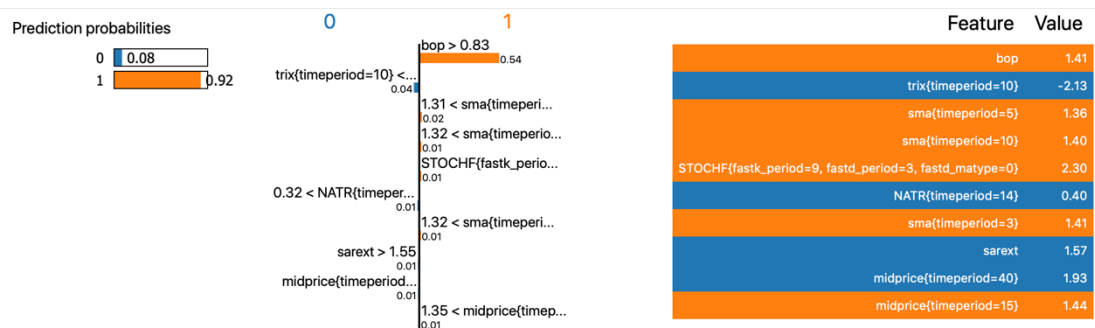


圖 4-10 skater 漲跌分析

以往我們會認為機器學習就是一個黑盒子，機器學習可以做出很好的預測，但我們無法得知預測背後的邏輯。於是我們利用 SHAP Value（SHapley Additive exPlanation）來進行分析，SHAP Value 是由 Shapley value 啟發的解釋模型，能對於每個預測樣本，模型都產生一個預測值，來測量我們預測模型中每個特徵對每個預測的正面或負面貢獻的程度。

下圖 4-11 為漲跌力道強弱的對峙圖，是大立光（3008）在 2018 年 1 月 1 日至 2020 年 5 月 12 日所有測試集的資料，可一次性的觀察每一日的預測，且預測



模型是如何做出漲跌的決定，紅色代表正向的影響，藍色代表負向的影響也就是跌。

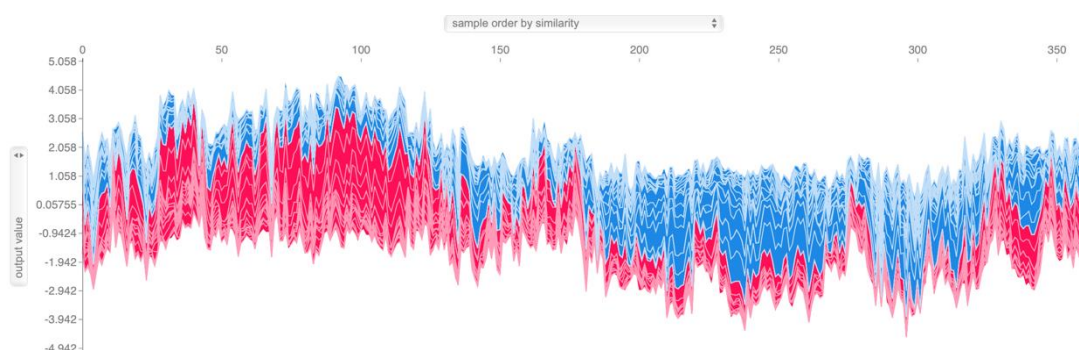


圖 4-11 漲跌強弱河流圖

下圖 4-12 為特徵總體分析，很明顯可以看出與上方我們所算出的特徵重要性（Feature Importance）相比，SHAP Value 的優勢在於可以將模型預測的結果轉換成是平均值及各個特徵影響程度的總和，也就是能反映出每一個樣本中特徵的影響力，且表現出所造成的影響是正向亦或是負向，因此也讓複雜的模型得以局部的被解釋。

以圖 4-12 特徵總體分析 cci 指標為例，縱座標為 Feature Value，當 cci 指標被轉為 SHAP Value 時，SHAP Value 的值會以顏色來呈現：高（紅色）、中（紫色）、低（藍色），橫坐標為正向及負向影響的程度，也就是當 cci 指標（圖 4-9 中紅色框處）的值過高（紅色）時，會被認為對於預測目標是正向因子（漲）；值過低（藍色）時，會被認為對於預測目標是負向因子（跌）。

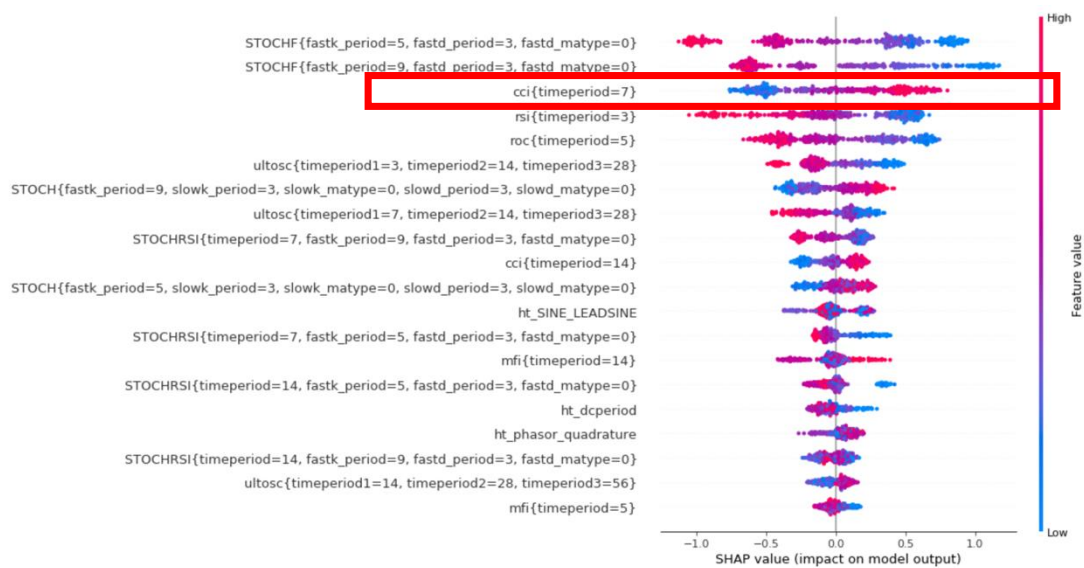


圖 4-12 特徵總體分析

下圖 4-13 為單變量對多變量交互作用，可以看出特徵與特徵之間互相的影響程度，可藉由觀察特徵與特徵影響的結果，作為傳統策略研發的依據。

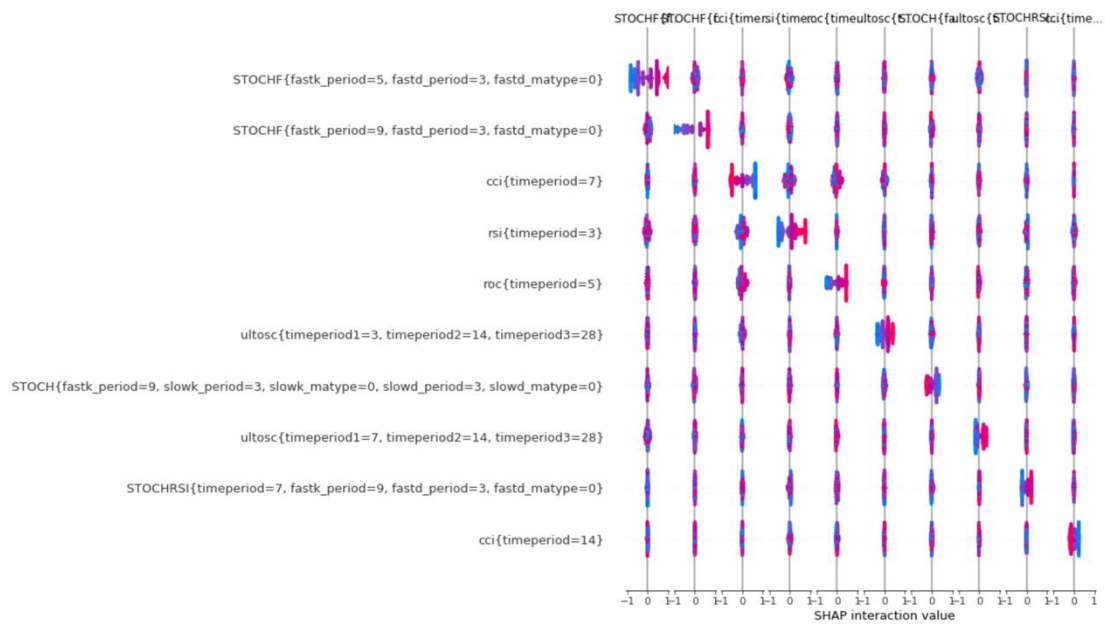


圖 4-13 單變量對多變量交互作用

## 第六節 結果呈現

本專題利用 Telegram Bot 將結果推播給使用者，利用 Telegram Bot 結合 Python 與 MySql 能夠將我們建立的系統做最有效率的呈現。如圖 4-14 Telegram Bot 的製作流程圖，利用了 Telegram API 擷取使用者對機器人傳送的訊息，透過判斷關鍵字執行與回傳不同的結果，分別提供三種指令\start、help、@股票代號，功能分別為回傳問候語，使用手冊，以及主程式呼叫。經由呼叫 Python 的主程式，能夠及時從 MySql 與網路爬蟲擷取主要資料，並計算擷取下的資料轉換成結果呈現給使用者。

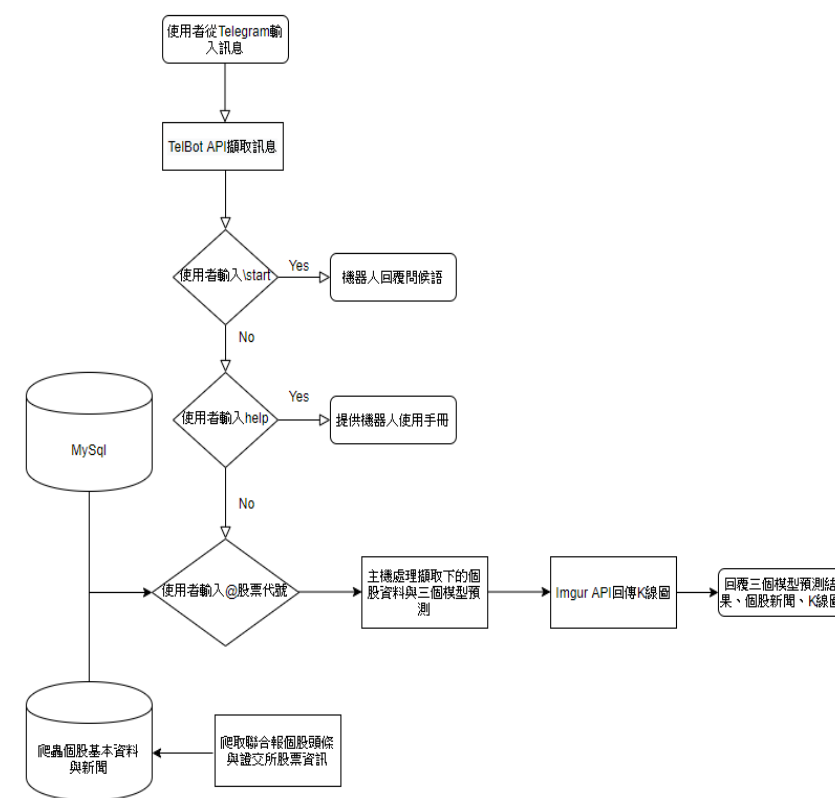


圖 4-14 Telegram Bot 流程圖

## 第七節 額外功能

藉由觀察上圖 4-12 特徵總體分析，實作一個簡易的傳統策略，參數上使用 BOP(fastperiod=7), STOCHF(fastk\_period = 5, fastd\_period = 3, fastd\_matype = 0) 來計算 K、D 值，傳統上認為 K 值大於 D 值為黃金交叉宜買入、K 值小於 D 值為死亡交叉宜賣出，然而上圖 4-9 特徵總體分析給出了不一樣的見解，針對大立光 (3008)，預測模型計算出來的結果與傳統認知截然不同，K 值大於 D 值為不利因素為賣出訊號、K 值小於 D 值為有利因素為買入訊號，並考量滑價及手續費，能獲得接近 3 倍的獲利。

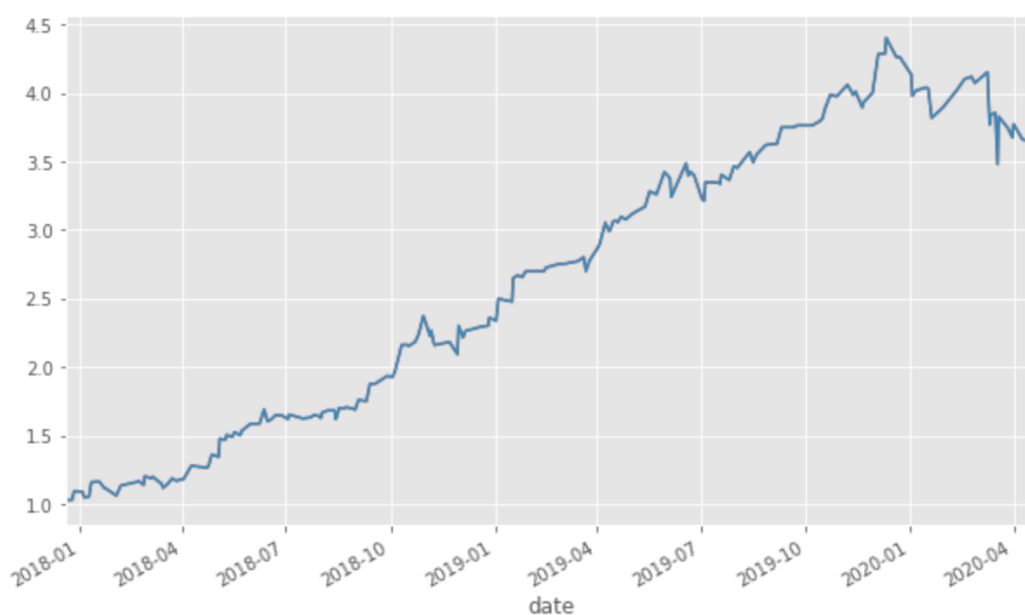


圖 4-15 大立光策略回測

## 第五章 實際操作方法

### 第一節 Jupyter Lab 實際操作流程

本系統使用 Jupyter Lab、Telegram Bot 作為介面呈現。以大立光（3008）、預測 1 日後的漲跌作為範例，在 Jupyter Lab 上提供以下功能供使用者使用：即時股價查詢、未來漲跌預測、歷史資料分析、財報評分、策略回測、策略選股。如圖 5-1 輸入股票代號在空格處輸入欲預測之股票（可自行選擇）。



歡迎來到AI股票預測分析系統

-----

請輸入股票代號：

圖 5-1 輸入股票代號

如圖 5-2 選擇預測的天數在空格處輸入預測幾日後的漲跌（可自行選擇）。



請輸入股票代號： 3008

請輸入預測幾日後的漲跌：

圖 5-2 選擇預測的天數

#### （一）即時股價查詢

圖 5-3 即時股價查詢，根據圖 5-1 使用者所選擇的股票代號，列出今日的即時資料，包含現價、開盤價、最高價、最低價、成交股數、以及近五日的收盤價。

<b>3008-即時股價查詢</b>			
-----			
<b>3008 (大立光) 2020-05-14 13:30:00</b>			
<b>現價: 4110.0000 / 開盤: 4165.0000</b>			
<b>最高: 4195.0000 / 最低: 4100.0000</b>			
<b>量: 434</b>			
-----			
<b>最近五日價格:</b>			
<b>[2020-05-14]</b>	<b>4110.0</b>		
<b>[2020-05-13]</b>	<b>4195.0</b>		
<b>[2020-05-12]</b>	<b>4110.0</b>		
<b>[2020-05-11]</b>	<b>4200.0</b>		
<b>[2020-05-08]</b>	<b>4225.0</b>		

圖 5-3 即時股價查詢

## (二) 漲跌預測

圖 5-4 股價漲跌預測，根據圖 5-1、圖 5-2 使用者所選擇的股票代號、預測天數所計算出的結果，True 代表為漲，False 代表為跌，並給出模型歷史的精準度供使用者參考。

	Stockid	Precision	漲跌預測
lightgbm	3008	0.792920	True
XGBoost	3008	0.814159	True
RandomForest	3008	0.814159	True

圖 5-4 股價預測

## (三) K 線圖繪製

圖 5-5 K 線圖繪製，繪製出從 2020 年 1 月 1 號至今日的 K 線圖，且包含五日、二十日均線以及成交量的繪製，供使用者參考。



圖 5-5 K 線圖

#### (四) 歷史資料分析

本系統歷史資料分析功能，對於使用者提供針對單一個股所寫的策略與歷史資料進行分析，本系統提供以下圖表供使用者參考。以下圖表為使用我們所撰寫出的圖 4-12 大立光策略進行分析圖 5-6 歷史資料概要，圖中的資訊提供分析的開始日、結束日、累積報酬、Sharpe ratio。

圖表解釋：由於我們策略是根據機器學習的分析結果所制訂出來，我們策略的起始時間為 2018/1/1，觀察圖 5-6 紅框部分可以得知以下資訊：

1. Annual Return (年化收益率)：為 72.9%是指把目前的收益率，換算成年收益率來計算，是一種理論收益，並不是真正已經取得的收益率。
2. Cumulative Returns (累積報酬)：策略開始到結束，總資產的變化率為 245.9%。
3. Sharpe Ratio：策略在單位總風險下所能獲得的超額受益為 3.58%。
4. Max Drawdown (最大虧損)：描述策略的最大虧損，越小越好，此策略的最

大虧損為 21%。

<b>Start date</b>	2007-04-23		
<b>End date</b>	2020-05-11		
<b>In-sample months</b>	126		
<b>Out-of-sample months</b>	27		
	All	In-sample	Out-of-sample
<b>Annual return</b>	25.9%	17.6%	72.9%
<b>Cumulative returns</b>	1804.3%	450.5%	245.9%
<b>Annual volatility</b>	47.0%	46.1%	50.6%
<b>Sharpe ratio</b>	1.72	1.29	3.58
<b>Calmar ratio</b>	0.71	0.49	3.48
<b>Stability</b>	0.86	0.84	0.94
<b>Max drawdown</b>	-36.3%	-36.3%	-21.0%
<b>Omega ratio</b>	1.35	1.25	1.94
<b>Sortino ratio</b>	2.66	1.93	6.47
<b>Skew</b>	NaN	NaN	NaN
<b>Kurtosis</b>	NaN	NaN	NaN
<b>Tail ratio</b>	1.26	1.10	1.67
<b>Daily value at risk</b>	-5.6%	-5.6%	-5.7%
<b>Alpha</b>	1.25	1.03	2.93
<b>Beta</b>	0.00	0.10	-0.44

圖 5-6 歷史資料概要

圖 5-7 歷史虧損幅度前 5 名，圖中的資訊提供所分析的個股從開始日以來的最大虧損幅度（Net drawdown in %）以及花了多長的時間股價重新回到起跌前的價位，可對比圖 5-8 大立光策略歷史虧損幅度前 5 名，使用策略後跟未使用策略的差異。

圖表解釋：以第一列作為範例，在 2007-06-22 至 2008-12-05 為最大虧損共 62.20%，在 2009-07-27 共花了 547 天才重新回到 2007-06-22 當日的股價。



Worst drawdown periods	Net drawdown in %	Peak date	Valley date	Recovery date	Duration
0	62.20	2007-06-22	2008-12-05	2009-07-27	547
1	53.22	2011-08-02	2012-04-30	2013-05-21	471
2	50.81	2017-08-25	2019-01-04	NaT	NaN
3	49.83	2015-07-03	2016-01-07	2016-08-08	287
4	25.07	2014-07-16	2014-10-27	2015-01-22	137

圖 5-7 大立光歷史虧損幅度前 5 名

Worst drawdown periods	Net drawdown in %	Peak date	Valley date	Recovery date	Duration
0	36.25	2007-06-23	2008-12-01	2010-04-02	725
1	34.00	2011-09-02	2011-11-22	2014-03-18	663
2	22.21	2015-10-12	2016-01-05	2016-01-20	73
3	20.96	2019-12-19	2020-03-18	NaT	NaN
4	13.77	2015-06-25	2015-07-24	2015-10-08	76

圖 5-8 大立光策略歷史虧損幅度前 5 名

圖 5-9 累積報酬將自製的策略進行回測，綠色部分為訓練期間（2018 年前），紅色部分為驗證期間（2018 年以後）的累計報酬，灰色部分則是使用 0050 作為基準與策略進行比較。圖 5-10 則是將 Y 軸單位轉化為對數表示。

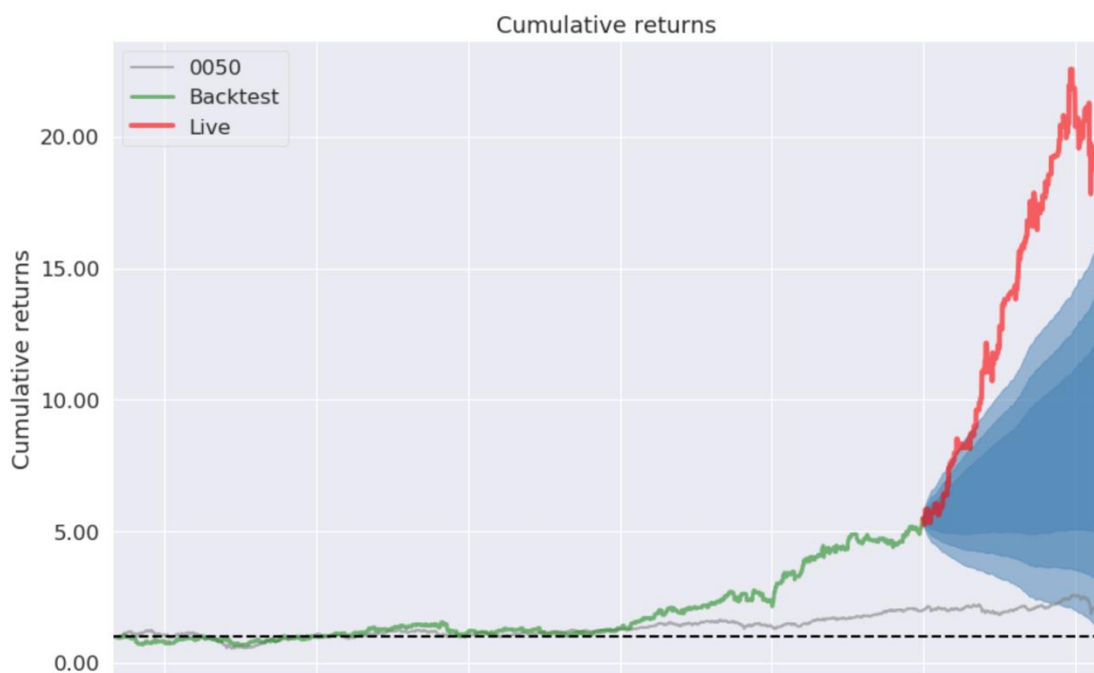


圖 5-9 累積報酬



圖 5-10 累積報酬（對數）

圖 5-11 策略與 0050 比較的累積收益波動率，綠色部分為訓練期間（2018 年前），紅色部分為驗證期間（2018 年以後）的累計收益波動率，灰色部分則是使用 0050 作為基準與策略進行比較



圖 5-11 與 0050 比較的累積收益波動率

圖 5-12 策略的每日收益，綠色部分為訓練期間（2018 年前），紅色部分為驗證期間（2018 年以後）的每日收益。



圖 5-12 策略的每日收益

圖 5-13 策略 Beta 值與 0050 之比較。

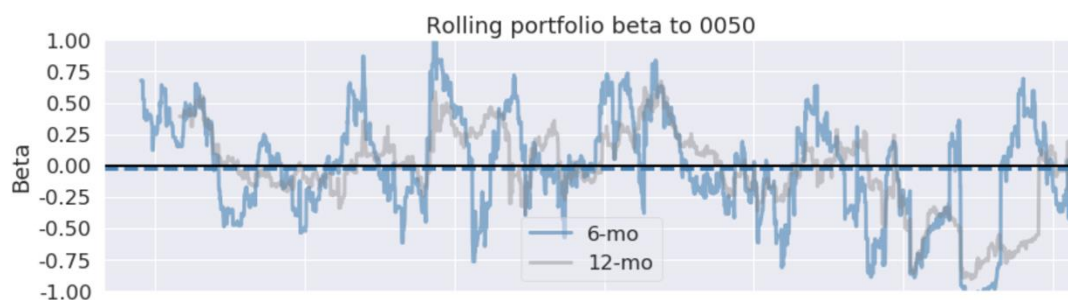


圖 5-13 策略滾動 Beta 係數

圖 5-14 觀察策略在歷史虧損幅度前五大期間的表現。

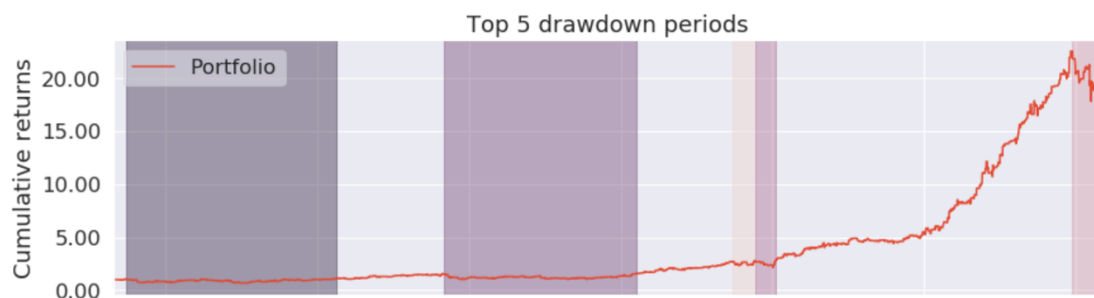


圖 5-14 歷史虧損幅度前 5 名期間

圖 5-15 虧損圖表，可搭配圖 5-14 一併觀察策略在歷史虧損幅度前五大期間的表現。



圖 5-15 虧損圖表

圖 5-16 每月報酬、圖 5-17 年報酬、圖 5-18 每月報酬分佈，觀察策略在每一個月及年度的報酬，觀察色溫圖找出此策略在五月、十月獲利較好。

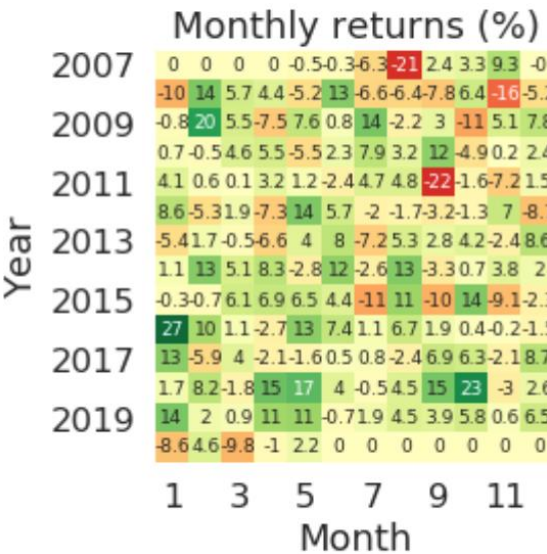


圖 5-16 每月報酬

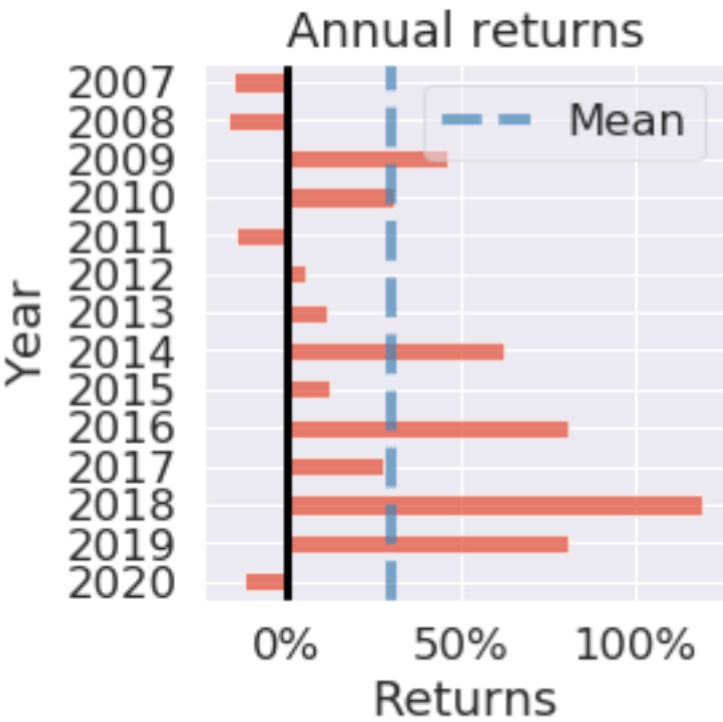


圖 5-17 年報酬

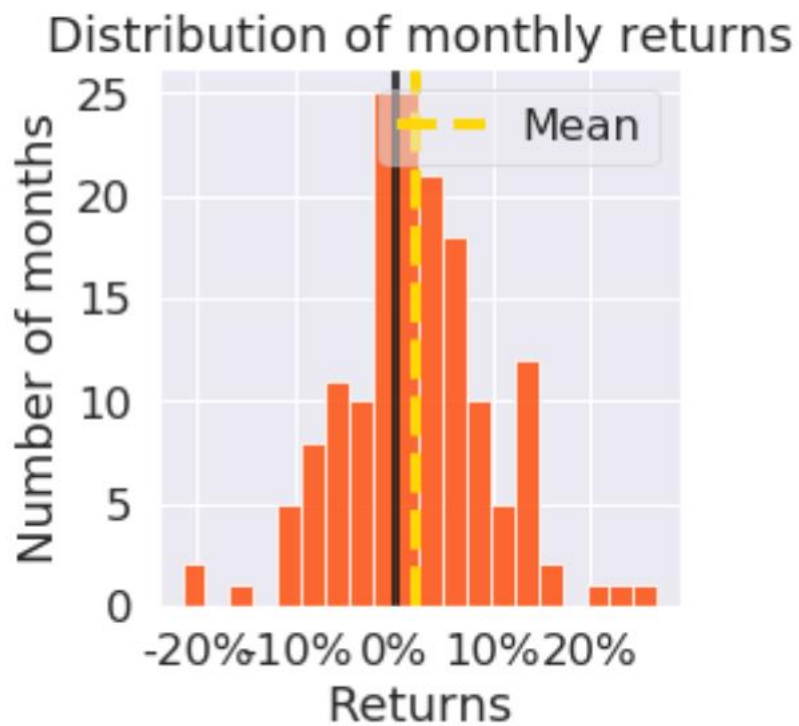


圖 5-18 每月報酬分佈

圖 5-19 報酬分點，找出策略使用時分別集中於日、週、月的區間。

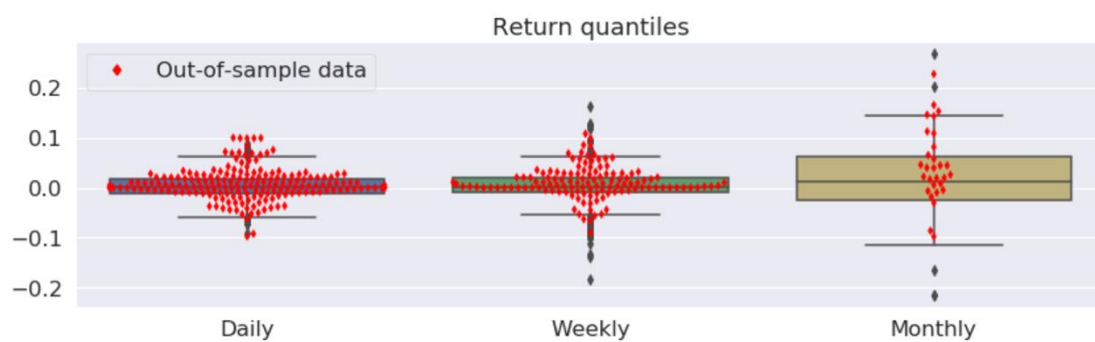


圖 5-19 報酬分點

### (五) 財報評分

如圖 5-20 財報評分將最新一期發佈的財報搜集下來，並針對安全性、成長性、轉機性三個面向進行評比，評分方式為參考財報狗所訂製出來的，能作為使用者選股之參考依據。

```

本次大立光電股份有限公司針對財報，安全性、成長性、轉機性三面向健檢如下
財報日期2020-03-31 00:00:00
=====3008的安全性健檢總分數:6/6=====
3008的條件1分數(自由現金流入近五年有三年大於 0):1
3008的條件2分數(自由現金流入近五年平均大於 0):1
3008的條件3分數(營業現金流入對淨利比近五年有三年大於 100%):1
3008的條件4分數(營業現金流入對淨利比近五年平均大於 100%):1
3008的條件5分數(應收帳款週轉天數小於等於去年同期數據):1
3008的條件6分數(存貨週轉天數小於等於去年同期數據):1
*****我是分隔線*****
=====3008的成長性健檢總分數:5/5=====
3008的條件7分數(月營收 YOY 連續三個月大於 0):1
3008的條件8分數(近一季毛利率年增率大於 0):1
3008的條件9分數:(近一季營業利益年增率大於 0):1
3008的條件10分數:(近一季稅前淨利年增率大於 0):1
3008的條件11分數:(近一季稅後淨利年增率大於 0):1
*****我還是分隔線*****
=====3008的轉機性健檢總分數:8/9=====
3008的條件12分數(當年度的總資產報酬率 > 0):1
3008的條件13分數(當年度的營業現金流 > 0):1
3008的條件14分數:(當年度營業現金流大於淨利):0
3008的條件15分數:(當年度長期負債金額小於上一年度):1
3008的條件16分數:(當年度流動比例(流動資產/流動負債)大於上一年度):1
3008的條件17分數(上一年度沒有發行新股):1
3008的條件18分數(當年度的總資產報酬率大於上一個年度的總資產報酬率):1
3008的條件19分數:(當前毛利率大於上一年度):1
3008的條件20分數:(當前資產週轉率大於上一年度):1
=====3008的安全性、成長性健檢分數:11/11=====
=====3008的安全性、成長性、轉機性健檢總分數:19/20=====

```

圖 5-20 財報評分

### (六) 策略選股

利用 Dash API 觀察我們所設計出的所有策略，在每一天所選出的標的。如圖 5-21 策略監控在 4/8 號時利用玩股網-鳴槍起漲強勢股策略選出 3217、3545、6547 三檔股票，並計算出這三檔股票持有至今的漲跌幅並畫出圖供使用者參考。

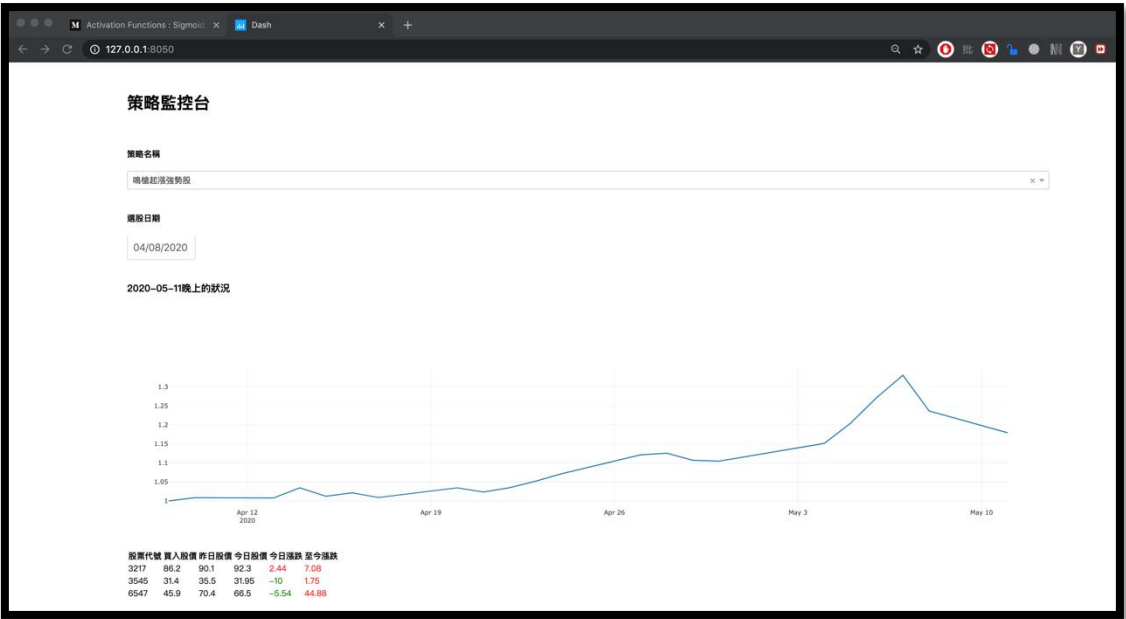


圖 5-21 策略監控

(七) 策略回測

根據圖 5-21 策略監控所使用的鳴槍起漲強勢股策略，進行策略回測，並固定持有 40 天做一次換股，回測後得出圖 5-22 和圖 5-23 每次換手最大報酬為：38.55%，每次換手最少報酬：-7.36%，持股數量，以及圖 5-24 持股清單，單一個股的最大報酬：51.94%，最大虧損來到 14%。根據圖 5-26 凱利公式建議使用這個策略操作每次可投入本金的 27%。

且圖 5-25 回測可根據使用者自行設定條件，包含回測的開始日、結束日、使用的策略、停損、停利、以及可自行選擇標的作為基準與策略進行對比。



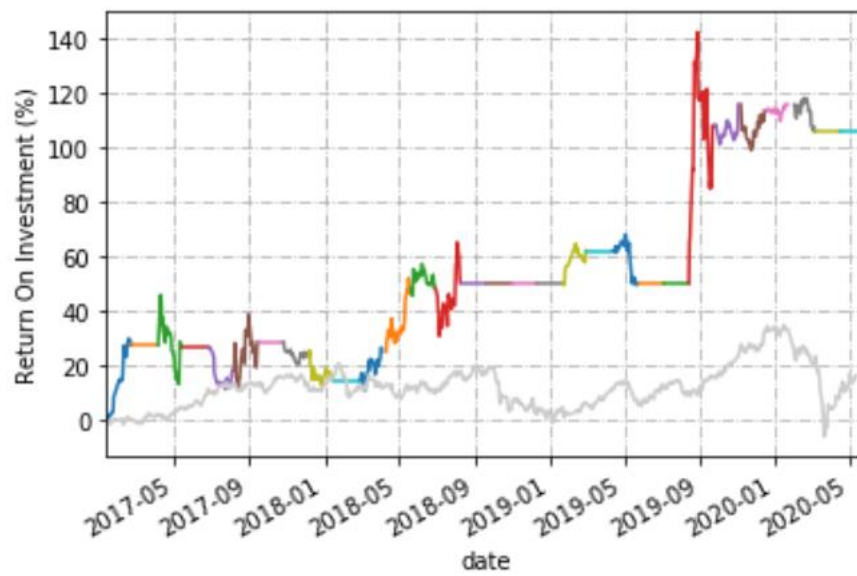


圖 5-22 策略回測

2017-01-11	-	2017-02-21	報酬率: 27.30 %	hold_stock 2
2017-02-21	-	2017-04-03	報酬率: 0.00 %	hold_stock 0
2017-04-03	-	2017-05-14	報酬率: -0.78 %	hold_stock 1
2017-05-14	-	2017-06-24	報酬率: 0.00 %	hold_stock 0
2017-06-24	-	2017-08-04	報酬率: -6.51 %	hold_stock 1
2017-08-04	-	2017-09-14	報酬率: 8.83 %	hold_stock 1
2017-09-14	-	2017-10-25	報酬率: 0.00 %	hold_stock 0
2017-10-25	-	2017-12-05	報酬率: -4.03 %	hold_stock 2
2017-12-05	-	2018-01-15	報酬率: -7.18 %	hold_stock 1
2018-01-15	-	2018-02-25	報酬率: 0.00 %	hold_stock 0
2018-02-25	-	2018-04-07	報酬率: 9.18 %	hold_stock 5
2018-04-07	-	2018-05-18	報酬率: 18.56 %	hold_stock 2
2018-05-18	-	2018-06-28	報酬率: 0.00 %	hold_stock 1
2018-06-28	-	2018-08-08	報酬率: 1.38 %	hold_stock 1
2018-08-08	-	2018-09-18	報酬率: 0.00 %	hold_stock 0
2018-09-18	-	2018-10-29	報酬率: 0.00 %	hold_stock 0
2018-10-29	-	2018-12-09	報酬率: 0.00 %	hold_stock 0
2018-12-09	-	2019-01-19	報酬率: 0.00 %	hold_stock 0
2019-01-19	-	2019-03-01	報酬率: 7.95 %	hold_stock 2
2019-03-01	-	2019-04-11	報酬率: 0.00 %	hold_stock 0
2019-04-11	-	2019-05-22	報酬率: -7.36 %	hold_stock 2
2019-05-22	-	2019-07-02	報酬率: 0.00 %	hold_stock 0
2019-07-02	-	2019-08-12	報酬率: 0.00 %	hold_stock 0
2019-08-12	-	2019-09-22	報酬率: 38.55 %	hold_stock 1
2019-09-22	-	2019-11-02	報酬率: 3.76 %	hold_stock 2
2019-11-02	-	2019-12-13	報酬率: -1.09 %	hold_stock 1
2019-12-13	-	2020-01-23	報酬率: 1.07 %	hold_stock 1
2020-01-23	-	2020-03-04	報酬率: -4.61 %	hold_stock 4
2020-03-04	-	2020-04-14	報酬率: 0.00 %	hold_stock 0
2020-04-14	-	2020-05-25	報酬率: 0.00 %	hold_stock 0
每次換手最大報酬 : 38.55 %				
每次換手最少報酬 : -7.36 %				

圖 5-23 換手報酬

	buy_price	sell_price	lowest_price	highest_price	buy_date	sell_date	profit(%)
stock_id							
2204	28.15	28.90	28.15	29.60	2017-01-12	2017-02-21	2.664298
4903	36.00	54.70	36.00	56.20	2017-01-12	2017-02-21	51.944444
3383	3.84	3.81	3.41	4.40	2017-04-06	2017-05-12	-0.781250
3090	30.70	28.70	27.00	30.80	2017-06-27	2017-08-04	-6.514658
3622	15.85	17.25	14.95	18.60	2017-08-07	2017-09-14	8.832808
1711	19.50	18.40	18.30	19.50	2017-10-26	2017-12-05	-5.641026
2380	7.00	6.83	6.50	7.00	2017-10-26	2017-12-05	-2.428571
3680	41.80	38.80	38.15	42.50	2017-12-06	2018-01-15	-7.177033
2038	14.10	13.10	11.40	15.50	2018-02-27	2018-04-03	-7.092199
2888	12.05	11.35	11.05	12.70	2018-02-27	2018-04-03	-5.809129
4560	61.00	76.70	57.00	81.00	2018-02-27	2018-04-03	25.737705
6141	21.90	30.75	21.90	30.75	2018-02-27	2018-04-03	40.410959
6237	35.40	32.80	32.60	40.35	2018-02-27	2018-04-03	-7.344633
3033	22.15	22.90	21.55	23.15	2018-04-10	2018-05-18	3.386005
3563	339.50	454.00	339.50	470.50	2018-04-10	2018-05-18	33.726068
1217	7.82	7.82	7.68	8.30	2018-05-21	2018-06-28	0.000000
1773	108.50	110.00	95.70	121.00	2018-06-29	2018-08-08	1.382488
2458	82.10	89.40	81.20	93.70	2019-01-22	2019-02-27	8.891596
6146	157.00	168.00	157.00	169.00	2019-01-22	2019-02-27	7.006369
4526	20.65	18.60	18.05	21.00	2019-04-12	2019-05-22	-9.927361
4966	521.00	496.00	492.00	551.00	2019-04-12	2019-05-22	-4.798464
6683	137.50	190.50	137.50	222.00	2019-08-13	2019-09-20	38.545455
2820	14.90	14.90	14.85	15.10	2019-09-24	2019-11-01	0.000000
6165	11.30	12.15	10.55	12.15	2019-09-24	2019-11-01	7.522124
2049	275.50	272.50	254.00	275.50	2019-11-05	2019-12-13	-1.088929
2881	46.60	47.10	45.80	47.15	2019-12-16	2020-01-20	1.072961
2867	12.50	11.55	11.50	12.50	2020-01-31	2020-03-04	-7.600000
2905	22.60	21.25	21.25	22.65	2020-01-31	2020-03-04	-5.973451
5206	18.90	16.25	16.25	18.95	2020-01-31	2020-03-04	-14.021164
8131	37.10	40.50	36.85	41.65	2020-01-31	2020-03-04	9.164420

圖 5-24 持有清單

```
Signature:
backtest(
    start_date,
    end_date,
    hold_days,
    strategy,
    data,
    weight='average',
    benchmark=None,
    stop_loss=None,
    stop_profit=None,
```

圖 5-25 自訂回測條件

我的選股策略在30次的進場數據中  
-----  
成功率為：46.67 %  
  
賠率為：2.79 %  
  
Kelly Formula 建議，若長期想使用這個策略進行操作，每次可投入本金的27 %進入市場

圖 5-26 凱利公式

## 第二節 Telegram Bot 操作流程

本專題另外建立了 Telegram Bot，提供所有使用者更方便快速的選擇。使用者僅須用手機下載 Telegram 並加入本專題的機器人，就能夠使用我們的系統，Telegram Bot 不僅操作簡單，更能提供使用者快速便利的股票資訊及預測結果。

### (1) 個股即時資料

如圖 5-27 在與機器人的聊天室中輸入指令/start 即可開始使用本專題機器人，輸入 help 則可查詢機器人功能，本範例以 3008 大立光為例，輸入@3008 則可呼叫主程式。



圖 5-27 機器人啟動畫面

## (2) 查詢、近一個月之 K 線圖

如圖 5-28 利用指令呼叫主程式後，機器人首先會回傳個股即時基本資訊以及本專題三個模型的預測結果。因模型計算過程較耗費時間，此過程會耗費大約 3~4 分鐘。



圖 5-28 機器人提供資訊結果

### (3) K 線均線成交量繪製與當日頭條新聞推播

如圖 5-29 在最後的部分，機器人則會回傳 K 線圖與利用爬蟲從聯合報爬下的當日個股頭條新聞。



圖 5-29 機器人提供個股頭條

## 附錄

附錄提供本系統計算所使用的技術指標以及參數

### (一) Volume Indicators 成交量指標

#### 1. AD - Chaikin A/D Line 量價指標

函數名稱：AD

名稱：Chaikin A/D Line 累積/派髮線(Accumulation/Distribution Line)

簡介：Marc Chaikin 提出的一種平衡交易量指標，以當日的收盤價位來估算成交量，用於估定一段時間內該證券累積的資金流量。

計算公式：

多空對比 =  $[(收盤價 - 最低價) - (最高價 - 收盤價)] / (最高價 - 最低價)$

#### 2. ADOSC - Chaikin A/D Oscillator

函數名稱：ADOSC

名稱：Chaikin A/D Oscillator Chaikin 震盪指標

簡介：將資金流動情況與價格行為相對比，檢測市場中資金流入和流出的情況。

計算公式：fastperiod A/D - slowperiod A/D

#### 3. OBV - On Balance Volume

函數名：OBV

名稱：On Balance Volume 能量潮

簡介：Joe Granville 提出，通過統計成交量變動的趨勢推測股價趨勢。

計算公式：以某日為基期，逐日累計每日上市股票總成交量，若隔日指數或股票上漲，則基期 OBV 加上本日成交量為本日 OBV。隔日指數或股票下跌，則基期 OBV 減去本日成交量為本日 OBV

計算公式：多空比率淨額 =  $[(收盤價 - 最低價) - (最高價 - 收盤價)] \div (最高價 - 最低價) \times 成交量$

### (二) Overlap Studies Functions 重疊研究指標

#### 1. BBANDS - Bollinger Bands

函數名：BBANDS

名稱：布林線指標

簡介：其利用統計原理，求出股價的標準差及其信賴區間，從而確定股價的波動範圍及未來走勢，利用波帶顯示股價的安全高低價位，因而也被稱為布林帶。

#### 2. DEMA - Double Exponential Moving Average 雙移動平均線



函數名：DEMA

名稱：雙移動平均線

簡介：兩條移動平均線來產生趨勢信號，較長期者用來識別趨勢，較短期者用來選擇時機。正是兩條平均線及價格三者的相互作用，才共同產生了趨勢信號。

### 3. EMA - Exponential Moving Average

函數名：EMA

名稱：指數平均數

簡介：是一種趨向類指標，其構造原理是仍然對價格收盤價進行算術平均，並根據計算結果來進行分析，用於判斷價格未來走勢的變動趨勢。

### 4. HT\_TRENDLINE - Hilbert Transform - Instantaneous Trendline

函數名：HT\_TRENDLINE

名稱：希爾伯特瞬時變換

簡介：是一種趨向類指標，其構造原理是仍然對價格收盤價進行算術平均，並根據計算結果來進行分析，用於判斷價格未來走勢的變動趨勢。

### 5. KAMA - Kaufman Adaptive Moving Average 考夫曼的自適應移動平均線

函數名：KAMA

名稱：考夫曼的自適應移動平均線

簡介：短期均線貼近價格走勢，靈敏度高，但會有很多雜訊，產生虛假信號；長期均線在判斷趨勢上一般比較準確，但是長期均線有著嚴重滯後的問題。我們想得到這樣的均線，當價格沿一個方向快速移動時，短期的移動平均線是最合適的；當價格在橫盤的過程中，長期移動平均線是合適的。

### 6. MA - Moving average 移動平均線

函數名：MA

名稱：移動平均線

簡介：移動平均線，Moving Average，簡稱 MA，原本的意思是移動平均，由於我們將其製作成線形，所以一般稱之為移動平均線，簡稱均線。它是將某一段時間的收盤價之和除以該週期。比如日線 MA5 指 5 天內的收盤價除以 5。

### 7. MAMA - MESA Adaptive Moving Average

### 8. MIDPRICE - Midpoint Price over period

### 9. SAR - Parabolic SAR 拋物線指標

函數名：SAR

名稱：拋物線指標

簡介：拋物線轉向也稱停損點轉向，是利用拋物線方式，隨時調整停損點位置以觀察買賣點。由於停損點(又稱轉向點 SAR)以弧形的方式移動，故稱之為拋物線轉向指標。

10. SAREXT - Parabolic SAR - Extended

11. SMA - Simple Moving Average 簡單移動平均線

函數名：SMA

名稱：簡單移動平均線

簡介：移動平均線，Moving Average，簡稱 MA，原本的意思是移動平均，由於我們將其製作成線形，所以一般稱之為移動平均線，簡稱均線。它是將某一段時間的收盤價之和除以該週期。比如日線 MA5 指 5 天內的收盤價除以 5。

12. T3 - Triple Exponential Moving Average (T3) 三重指數移動平均線

函數名：T3

名稱：三重指數移動平均線

簡介：TRIX 長線操作時採用本指標的訊號，長時間按照本指標訊號交易，獲利百分比大於損失百分比，利潤相當可觀。比如日線 MA5 指 5 天內的收盤價除以 5。

13. TEMA - Triple Exponential Moving Average

14. TRIMA - Triangular Moving Average

15. WMA - Weighted Moving Average 移動加權平均法

(三) Momentum Indicators 動量指標

1. ADX - Average Directional Movement Index

函數名：ADX

名稱：平均趨向指數

簡介：使用 ADX 指標，指標判斷盤整、振蕩和單邊趨勢。

2. ADXR- Average Directional Movement Index Rating

函數名：ADXR

名稱：平均趨向指數的趨向指數

簡介：使用 ADXR 指標，指標判斷 ADX 趨勢。

3. APO - Absolute Price Oscillator

4. AROON - Aroon

函數名：AROON

名稱：阿隆指標

簡介：該指標是通過計算自價格達到近期最高值和最低值以來所經過的期間數，阿隆指標說明你預測價格趨勢到趨勢區域(或者反過來，從趨勢區域到趨勢)的變化。

計算公式：

Aroon(上升)=[(計算期天數-最高價后的天數)/計算期天數]\*100

Aroon(下降)=[(計算期天數-最低價后的天數)/計算期天數]\*100

#### 5. AROONOSC - Aroon Oscillator

函數名：AROONOSC

名稱：阿隆振蕩

#### 6. AROONOSC - Aroon Oscillator

函數名：BOP

名稱：均勢指標

#### 7. CCI - Commodity Channel Index

函數名：CCI

名稱：順勢指標

簡介：CCI 指標專門測量股價是否已超出常態分佈範圍。

#### 8. CMO - Chande Momentum Oscillator 錢德動量擺動指標

函數名：CMO

名稱：錢德動量擺動指標

簡介：與其他動量指標擺動指標如相對強弱指標(RSI)和隨機指標(KDJ)不同，錢德動量指標在計算公式的分子中採用上漲日和下跌日的數據。

計算公式： $CMO=(Su-Sd)*100/(Su+Sd)$

其中：Su 是今日收盤價與昨日收盤價(上漲日)差值加總。若當日下跌，則增加值為 0；Sd 是今日收盤價與昨日收盤價(下跌日)差值的絕對值加總。若當日上漲，則增加值為 0。

#### 9. DX - Directional Movement Index DMI 指標又叫動向指標或趨向指標

函數名：DX

名稱：動向指標或趨向指標

簡介：通過分析股票價格在漲跌過程中買賣雙方力量均衡點的變化情況，即多空

雙方的力量的變化受價格波動的影響而發生由均衡到失衡的循環過程，從而提供對趨勢判斷依據的一種技術指標。

#### 10. MACD - Moving Average Convergence/Divergence

函數名：MACD

名稱：平滑異同移動平均線

簡介：利用收盤價的短期(常用為 12 日)指數移動平均線與長期(常用為 26 日)指數移動平均線之間的聚合與分離狀況，對買進、賣出時機作出研判的技術指標。

#### 11. MFI - Money Flow Index 資金流量指標

函數名：MFI

名稱：資金流量指標

簡介：屬於量價類指標，反映市場的運行趨勢。

#### 12. MINUS\_DI - Minus Directional Indicator

函數名：DMI 中的 DI 指標負方向指標方向指標名稱：下降動向值

簡介：通過分析股票價格在漲跌過程中買賣雙方力量均衡點的變化情況，即多空雙方的力量的變化受價格波動的影響而發生由均衡到失衡的循環過程，從而提供對趨勢判斷依據的一種技術指標。

#### 13. MINUS\_DM - Minus Directional Movement

函數名稱：MINUS\_DM

名稱：上升動向值 DMI 中的 DM 代表正趨向變動值即上升動向值

#### 14. MOM - Momentum 動量

函數名：MOM

名稱：上升動向值

簡介：投資學中意思為續航，指股票(或經濟指數)持續增長的能力。研究發現，贏家組合在牛市中存在著正的動量效應，輸家組合在熊市中存在著負的動量效應。

#### 15. PLUS\_DI - Plus Directional Indicator

#### 16. PLUS\_DM - Plus Directional Movement

#### 17. PPO - Percentage Price Oscillator 價格震盪百分比指數

函數名：PPO

名稱：價格震盪百分

比指數簡介：價格震盪百分比指標(PPO)是一個和 MACD 指標非常接近的指標。

PPO 標準設定和 MACD 設定非常相似：12、26、9 和 PPO，和 MACD 一樣說明了兩條移動平均線的差距，但是它們有一個差別是 PPO 是用百分比說明。

18. ROC - Rate of change： $((price/prevPrice)-1)*100$  變動率指標

函數名：ROC

名稱：變動率指標

簡介：ROC 是由當天的股價與一定的天數之前的某一天股價比較，其變動速度的大小，來反映股票市變動的快慢程度。

19. ROCP - Rate of change Percentage： $(price-prevPrice)/prevPrice$

20. RSI - Relative Strength Index 相對強弱指數

函數名：RSI

名稱：相對強弱指數

簡介：是通過比較一段時期內的平均收盤漲數和平均收盤跌數來分析市場買沽盤的意向和實力，從而作出未來市場的走勢。

21. STOCH - Stochastic 隨機指標，俗稱 KD

函數名：STOCH

名稱：隨機指標，俗稱 KD

22. STOCHF - Stochastic Fast

23. STOCHRSI - Stochastic Relative Strength Index

24. TRIX - 1-day Rate-Of-Change (ROC) of a Triple Smooth EMA

25. ULTOSC - Ultimate Oscillator 終極波動指標

函數名：ULTOSC

名稱：終極波動指標

簡介：UOS 是一種多方位功能的指標，除了趨勢確認及超買超賣方面的作用之外，它的"突破"訊號不僅可以提供最適當的交易時機之外，更可以進一步加強指標的可靠度。

26. WILLR - Williams' %R 威廉指標

函數名：WILLR

名稱：威廉指標

簡介：WMS 表示的是市場處於超買還是超賣狀態。股票投資分析方法主要有如下三種：基本分析、技術分析、演化分析。在實際應用中，它們既相互聯繫，又有重要區別。

#### (四) Volatility Indicator Functions 波動率指標函數

##### 1. ATR - Average True Range

函數名：ATR

名稱：真實波動幅度均值

簡介：真實波動幅度均值(ATR)是以 N 天的指數移動平均數平均後的交易波動幅度。

計算公式：一天的交易幅度只是單純地最大值 - 最小值。而真實波動幅度則包含昨天的收盤價，若其在今天的幅度之外：

真實波動幅度 =  $\max(\text{最大值}, \text{昨日收盤價}) - \min(\text{最小值}, \text{昨日收盤價})$  真實波動幅度均值便是「真實波動幅度」的 N 日指數移動平均數。

##### 2. NATR - Normalized Average True Range

函數名：NATR

名稱：歸一化波動幅度均值

##### 3. TRANGE - True Range

##### 4. Price Transform 價格指標

##### 5. AVGPRICE - Average Price

函數名：AVGPRICE

名稱：平均價格函數

##### 6. MEDPRICE - Median Price

函數名稱：MEDPRICE

名稱：中位數價格

##### 7. TYPPRICE - Typical Price

函數名稱：TYPPRICE

名稱：代表性價格

##### 8. WCLPRICE - Weighted Close Price

函數名稱：WCLPRICE

名稱：加權收盤價

#### (五) Cycle Indicators 週期指標

1. HT\_DCPERIOD - Hilbert Transform - Dominant Cycle Period

函數名：HT\_DCPERIOD

名稱：希爾伯特變換-主導週期

簡介：將價格作為資訊信號，計算價格處在的週期的位置，作為擇時的依據。

2. HT\_DCPHASE - Hilbert Transform - Dominant Cycle Phase

函數名：HT\_DCPHASE

名稱：希爾伯特變換-主導循環階段

3. HT\_PHASOR - Hilbert Transform - Phasor Components

函數名：HT\_DCPHASE

名稱：希爾伯特變換-希爾伯特變換相量分量

4. HT\_SINE - Hilbert Transform - SineWave

函數名：HT\_DCPHASE

名稱：希爾伯特變換-正弦波

5. HT\_TRENDMODE - Hilbert Transform - Trend vs Cycle Mode

函數名：HT\_DCPHASE

名稱：希爾伯特變換-趨勢與週期模式

## 參考文獻

- Eugene Fama (1970), "Efficient Capital Markets : A Review of Theory and Empirical Work," *Journal of Finance*, 25, pp. 383-417.
- 樓禎祺、何培基(2003),「股價移動平均線之理論與實證—以台灣股市模擬投資操作為例」,育達研究叢刊,5&6期,頁27-51。
- 李良俊(2003),「台灣股票市場技術分析有效性之研究」,未出版碩士論文,實踐大學企業管理研究所。
- 陳惠郁(2018),「以技術指標分析股價走勢:以台灣股市為例」,未出版碩士論文,國立政治大學風險管理與保險學系。
- Patel, J., Shah, S., Thakkar, P., K Kotecha (2014), "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques," *Expert Systems with Applications* 42 (2015) 259–268
- Shubharthi, D., Snehanthu, S., Yash, K., and Suryoday, B. (2016), "Forecasting to Classification : Predicting the direction of stock market price using Xtreme Gradient Boosting," Working Paper, PESIT South Campus.
- Suryoday Basak · Snehanthu Saha · Saibal Kar · Luckyson Khaidem · Sudeepa Roy Dey (2019), "Predicting the Direction of Stock Market Price Using Tree Based Classifiers"
- Sepp Hochreiter and Jurgen Schmidhuber (1997), "LONG SHORT-TERM MEMORY," *Neural Computation*, vol. 9, no. 8, 1997.
- Murtaza Roondiwala, Harshal Patel and Shraddha Varma (2015), "Predicting Stock Prices Using LSTM," *International Journal of Science and Research (IJSR)*.
- Xiong et al.(2016), "Learning Stock Volatility with Google Domestic Trends," *arXiv preprint arXiv : 1512.04916*.
- Deng Yue et al.(2017), "Deep Direct Reinforcement Learning for Financial Signal Representation and Trading," *IEEE transactions on neural networks and learning systems* 28.3 (2017) : 653-664.
- Nelson, D., Pereira, A., & Oliveira, R. (2017). "Stock Market's Price Movement Prediction with LSTM Neural Networks," In *Neural Networks (IJCNN)*, 2017



International Joint Conference on (pp. 1419-1426). IEEE.

Fischer Thomas and Krauss Christopher (2017), “Deep learning with long short-term memory networks for financial market predictions,” FAU Discussion Papers in Economics, No. 11.

python 網路爬蟲簡介, available from URL :

<https://freelancerlife.info/zh/blog/python-web-scraping-overview/>

洪志麟、賴冠妤、陳國瑞、呂馥伊(2016),「機器人投資顧問 (Robo-Advisor) 國外實務及相關法令與管理措施之研究」

施育霖(2016),「以機器學習方法預測股價：以台股金融類股為案例」, 未出版碩士論文, 中興大學資訊管理系所。

劉順富(2019),「用 XGBOOST 演算法預測台灣指數期貨」, 未出版碩士論文, 雲林科技大學財務金融所。