

國立臺灣大學工學院工程科學及海洋工程學系



碩士論文

Department of Engineering Science and Ocean Engineering

College of Engineering

National Taiwan University

Master Thesis

結合 LSTM 股價預測與基因模糊交易策略—以台灣 50 為例

Combining LSTM to predict stock price and fuzzy genetic
algorithm to determine trading strategy in the case of Taiwan
ETF50 Stock

呂雅芳

Ya-Fang Lu

指導教授：張瑞益 博士

Advisor: Ray-I Chang, Ph.D.

中華民國 112 年 7 月

July 2023

誌謝



在就讀碩士班這兩年中，很感謝背後支持我的家人、師長、朋友以及同學們的協助，讓我順利完成學業，感謝他們的幫忙。

離開家鄉花蓮，前來北部讀書，我要衷心謝謝家人一直以來默默的支持和陪伴，他們時常關心我，為我打氣，讓我感受到無盡的溫暖與鼓勵。

在研究所學習過程中，我非常感謝我的指導教授張瑞益教授，這一路上不僅在碩論給予我指導，也教導我許多處事方面。每週定時的開會，都會給我許多寶貴的建議。當在實驗上遇到困難時感到茫然無措時，老師總是能迅速指出我的問題點，讓我立即領悟並找出解決的辦法，使我研究進度穩定的前進。

在就讀臺大讀碩士期間，我與許多同學度過了許多難忘的時光，像是一起修課、寫作業甚至熬夜寫到天亮，這些經歷都讓我在碩士生涯中留下許多快樂的回憶。我要感謝學長姐給予我們寶貴的研究經驗以及建議，讓我在研究上能夠順利地進行，Lab 125A 是個溫暖的大家庭，彼此互相扶持，願大家在未來道路上發光發熱。

最後，感謝在這兩年遇到的任何人，祝大家身體健康、快樂。

中文摘要



本研究以 LSTM 預測元大台灣 50 (ETF50) 股價，為了提高模型預測的準確度，提出了「誤差校正法」，進行修正預測股價。我們計算每日預測股價與實際股價之誤差，並判斷誤差值是否超過閾值，若超過閾值則進行校正預測股價。同時，我們使用基因演算法 (Genetic Algorithm, GA) 來調整校正法中的參數，以達到準確預測股價的目的。

在股票交易中，決定股票買賣時間點是交易關鍵之一。過去許多股票研究會使用基本面、消息面或是技術面的技術性指標來判斷最佳買賣點。本研究提出了基因演化之模糊演算法來決定買賣時機，使用多種技術指標應用規則建立模糊系統 (Fuzzy System)，並結合 GA 來演化出最佳的隸屬函數之參數，以改善模糊系統。

本研究收集 2003 年到 2020 年間 ETF50 和 ETF50 中成分股佔比最大的股票作為資料集，去預測 ETF50 股價，其中前段 90% 作為訓練資料，後段 10% 作為測試資料。將訓練資料經過 12 項技術性指標計算以及前處理後，作為模型輸入變數。

模型預測值經過誤差校正法進行校正後，與未校正的股價進行比較，使用均方誤差 (Mean Square Error, MSE) 來評估預測準確性，結果顯示，未經校正股價的 MSE 為 11.5758，而經過校正後的股價 MSE 降至 1.2687，大幅地降低模型預測的誤差。透過誤差校正法校正股價並以基因演算法決定買賣點，本研究最終實驗可獲得 32.0% 的報酬率。

關鍵字：基因演算法、模糊系統、LSTM、校正策略、ETF50、交易策略、技術性指標。

ABSTRACT

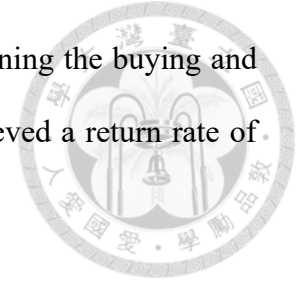
This study employs LSTM to predict the stock prices of Yuanta/P-shares Taiwan Top 50 ETF. In order to improve the accuracy of the prediction model, we propose the "Deviation Calibration" method to adjust the predicted stock prices. We calculate the daily deviation between the predicted stock price and the actual stock price, and determine if the deviation value exceeds a threshold. If it exceeds the threshold, we proceed to correct the predicted stock price. Additionally, we utilize Genetic Algorithm (GA) to adjust the parameters of the deviation calibration method in order to achieve accurate stock price predictions.

In stock trading, determining the timing of buying and selling stocks is one of the key aspects of trading. In the past, many stock studies have used fundamental, news-based, or technical indicators to identify optimal buying and selling points. This study proposes a genetic fuzzy algorithm to determine the timing of trades by utilizing multiple technical indicators to establish a fuzzy system. Additionally, it combines GA to evolve the optimal parameter values for the membership functions in order to improve the fuzzy system.

This study collected the stock data of the ETF50 and the constituents with the highest weighting in the ETF50 as the dataset, which is from 2003 to 2020, to predict the ETF50 stock price. The dataset was split into two parts, with the first 90% used as training data and the last 10% as testing data. The training data underwent calculations of 12 technical indicators and preprocessing, which were then used as input variables for the model.

After applying the deviation calibration method to the predicted values of the model, a comparison was made with the uncorrected stock prices. The mean square error (MSE) was used to evaluate the prediction accuracy. The results showed that the MSE of the uncorrected stock prices was 11.5758, whereas the MSE of the corrected stock prices decreased to 1.2687, and the stock prediction error is reduced significantly. By correcting

the stock prices using the deviation calibration method and determining the buying and selling points through genetic algorithm, this study ultimately achieved a return rate of 32.0% in the experiments.



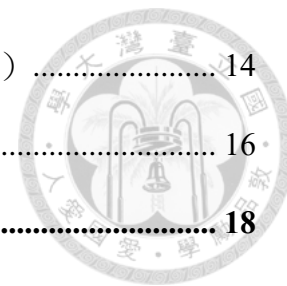
Keywords: Genetic Algorithm, Fuzzy System, LSTM, Calibration Strategy, ETF50, Trading strategy, Technical Indicators.

目錄

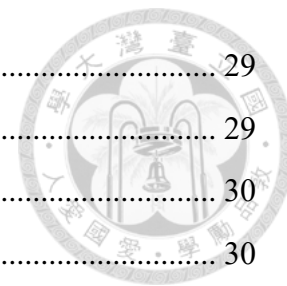


誌謝	i
中文摘要	ii
ABSTRACT	iii
目錄	v
圖目錄	viii
表目錄	ix
第一章 緒論	1
1.1 研究動機	1
1.2 研究目的	1
1.3 論文結構	2
第二章 相關文獻	3
2.1 指數股票型基金 (ETF)	3
2.2 技術面指標	3
2.3 模糊系統	8
2.3.1 模糊理論	8
2.3.2 模糊控制	8
2.4 基因演算法	10
2.4.1 初始族群	11
2.4.2 適應函數 (Fitness Function)	11
2.4.3 選擇 (Selection Scheme)	11
2.4.4 交配 (Crossover)	12
2.4.5 突變 (Mutation)	13
2.5 類神經網路	13
2.5.1 循環神經網路 (Recurrent Neural Networks, RNN)	13

2.5.2	長短期記憶模型 (Long Short-Term Memory, LSTM)	14
2.6	國內外相關研究	16
第三章	研究方法及架構	18
3.1	資料集	18
3.2	研究流程與架構	18
3.3	資料前處理	19
3.3.1	計算技術指標	19
3.3.2	資料正規化 Normalization	19
3.3.3	調整成分股權重	20
3.3.4	切割資料	20
3.4	LSTM 類神經網路模型與校正策略	21
3.4.1	LSTM 類神經網路架構	21
3.4.2	模型訓練參數	21
3.4.3	模型評估方式	22
3.4.4	誤差校正法	22
3.4.5	基因演化之誤差校正法	23
3.5	交易策略	24
3.5.1	買賣規則	24
3.5.2	投資報酬率	25
3.6	基因演化之模糊演算法	25
3.6.1	技術指標之應用規則	25
3.6.2	隸屬函數 (Membership Function)	26
3.6.3	隸屬函數初始值	27
3.6.4	基因演算法輸入變數	27
3.6.5	適應函數	28
3.6.6	基因演化之模糊演算法訓練參數	28
第四章	實驗結果	29



4.1	模型預測	29
4.1.1	模型預測校正比較	29
4.1.2	誤差校正法與基因演算法	30
4.1.3	當日最高價與最低價	30
4.2	基因演化之模糊演算法	30
4.3	交易策略比較	32
4.3.1	股價校正後分析	33
4.3.2	不同的總資金比較	34
4.3.3	基因演化之模糊演算法報酬率比較	34
第五章	結論與未來展望	36
5.1	結論	36
5.2	未來展望	36
參考文獻	37



圖目錄



圖 1 模糊控制架構圖	9
圖 2 定義服務品質的隸屬函數示意圖	9
圖 3 基因演算法流程架構圖	11
圖 4 輪盤式選擇示意圖	12
圖 5 基因單點交配示意圖	13
圖 6 基因雙點交配示意圖	13
圖 7 RNN 展開示意圖	14
圖 8 LSTM 架構圖	14
圖 9 研究流程圖	18
圖 10 切割資料示意圖	21
圖 11 LSTM 架構圖	21
圖 12 誤差校正法虛擬碼	23
圖 13 K 值隸屬函數	27
圖 14 模型預測股價之每日收盤價圖	29
圖 15 模型校正後預測之每日收盤價圖	29
圖 16 K 值最佳隸屬函數	31
圖 17 D 值最佳隸屬函數	32
圖 18 RSI 值最佳隸屬函數	32
圖 19 Covid-19 影響 ETF50 股價圖	33

表目錄



表 1 模型輸入變數	19
表 2 模型訓練參數設定	22
表 3 基因演算法訓練參數	24
表 4 技術指標買賣訊號與 ACTION 之規則	26
表 5 技術指標與輸出之隸屬函數參數設定	26
表 6 技術指標之隸屬函數初始值	27
表 7 基因演算法輸入變數	28
表 8 基因演化之模糊演算法訓練參數	28
表 9 校正閾值對誤差校正法 MSE 比較	30
表 10 基因演算法得出的最佳隸屬函數參數值 (Membership Function)	31
表 11 不同交易策略報酬率比較	33
表 12 不同總資金的報酬率比較表	34
表 13 不同時間區間報酬率比較	35

第一章 緒論



本篇研究主要著重於股票技術面分析，透過技術指標數值的表現，以了解目前股市的狀況。然而，在某些黑天鵝事件或股市劇烈波動的時期，過去的股票技術指標應用規則可能無法達到較好的交易效果，例如 2020 年受到國際疫情的影響，股價突然大幅度波動，以往的股票技術指標應用規則無法預期到這種劇烈震盪。因此，本研究使用校正方法來修正模型對股價的預測，以減少突發外部因素對模型預測的不準確性所造成的影響。

為了獲取更高的報酬率，投資者希望在股票價格低點時買進，並在股價高點時賣出。本研究利用基因演化之模糊演算法決定買賣時機，作為股票交易策略[3]，最後，建立模擬交易並計算交易後的報酬率。

1.1 研究動機


近期股票市場會因外部因素造成巨幅波動，使得投資者難以掌握良好的買賣時機，如近期 2020 年 Covid-19 疫情[18]、2018 年中美貿易戰以及 2011 年美債危機等事件，對金融股市造成了極大的影響，導致投資者難以獲得較好的報酬率。因此，一些投資者會通過消息面來了解公司近日的消息，再決定是否進行股票買入或賣出等操作。

然而，基本面的訊息往往落後於市場，隨著假訊息的生成與散佈越加猖獗，消息面的訊息可信度令人疑慮，因此，本論文採用股票市場真實數據的技術面指標來分析和了解目前股市情況。

1.2 研究目的

本研究提出一套股票交易策略，使投資人能透過模型預測未來股價、得知買賣時間點以及選擇最佳交易策略，以獲得報酬率。本研究的目的如下：

- 一、**預測股價**：本研究使用深度學習長短期記憶模型，透過模型隱藏層多個神經元以及多層神經，訓練過去的股票歷史資料和技術性指標資料，以學習股價走勢，預測出未來股價。

- 
- 二、**降低模型預測的誤差**：為了降低模型預測的誤差並避免股價即時股價劇烈波動而導致預測結果的不準確性，透過誤差校正法去修正股價，以達到降低模型預測誤差的目的。
 - 三、**最佳的買賣點**：利用基因演算法演化出技術性指標最佳隸屬函數之參數，建立隸屬函數以及設定模糊規則，形成一套模糊系統，以決定最佳的買賣點，以獲取最佳買賣時機。
 - 四、**買賣交易策略**：藉由模型預測的最低價和最高價作為買入和賣出的價格，進行交易。

1.3 論文結構

本研究之架構分成四個章節，第一章提及此論文的研究動機。第二章介紹基因演算法、模糊系統以及股市交易策略的相關文獻。第三章則說明本研究的實驗流程與架構、時間序列模型、決定買賣時間點以及買賣決策規則。第四章為實驗結果與分析，找到最佳的買賣決策規則。最後則為結論與未來展望，總結研究成果，探討缺失並提出未來改進及延伸的方向。

第二章 相關文獻



本章節介紹指數股票型基金 (ETF) 以及技術面常用的技術指標，次而介紹模糊系統 (Fuzzy System) 以及基因演算法 (Genetic Algorithm, GA)，接著介紹預測股價的深度學習模型，最後，探討國內外投資股票的相關研究論文。

2.1 指數股票型基金 (ETF)

ETF (Exchange Traded Fund) [1] 又稱為「指數股票型基金」，會追蹤並複製特定指數的基金。

ETF 優點：

1. 交易簡單：如股票相同交易方式。
2. 分散風險：追蹤指數表現，多樣化選擇，降低投資風險性。
3. 投資成本低：ETF 管理費比一般基金管理費較低，ETF 與股票手續費為 0.1425%，ETF 證券交易稅僅 0.1%。
4. 被動式投資：追蹤複製指數表現買入，無須耗費大量的時間管理。
5. 持股透明度高：由於 ETF 為追蹤指數，持股內容可容易得知。

ETF 缺點：

1. 無法自行調整指數：由於複製股票指數，無法客製化調整指數。
2. 不利於頻繁交易：ETF 適合於長期投資，太頻繁交易會有多次手續費的成本。
3. 報酬率較低：ETF 報酬率屬於穩健成長型，如追求短期報酬率高的投資者，反而不適合投資 ETF。

本研究使用「元大寶來台灣卓越 50 基金」(ETF50)，簡稱「元大台灣 50」做為研究目標，為追蹤台灣前 50 大股票指數，並持有與台灣前 50 指數成分股相同。

2.2 技術面指標

在股票市場中，常有以技術面作為分析，對於過往股市歷史資料分析，經過統計與數值的計算，來預測未來股市走勢，以下介紹常見的指標：



1. 隨機指標 (Stochastic Oscillator, KD)

KD 又稱 (Stochastic Oscillator) 由 George Lane [19] 提出，由 K 值與 D 值所組成。KD 值範圍介於 0~100 之間，透過這兩個值評估目前股價脫離股價正常範圍的程度。K 值為「快速平均值」，又稱快線，對於股價變化反應較快和靈敏，D 值為「慢速平均值」，又稱慢線，對於股價變化反應相對於 K 值較為慢。K 值與 D 值是基於 RSV (Raw Stochastic Value) 而算出，RSV 又稱「未成熟隨機」，判斷當日股價走勢。以下為 RSV、K 值與 D 值計算：

RSV 值計算方式：

C_t ：今日收盤價， L_n ：過去 n 日內最低價， H_n ：過去 n 日內最高價

$$RSV = \frac{(C_t - L_n)}{(H_n - L_n)} \times 100 \quad (2-1)$$

KD 值計算方式：

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

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

KD 值的應用規則：

- (1) 當 KD 值 < 20 時，目前股價弱勢，為超賣的狀態，可買入股票。
- (2) 當 KD 值 > 80 時，代表目前股價強勢，超買的現象，可賣出股票。
- (3) K 線由下往上穿越 D 線時，稱之為「黃金交叉」，表示股價可能上漲，建議買進、做多。
- (4) K 線由上往下跌破 D 線時，稱之為「死亡交叉」，表示股價可能會下跌，建議賣出、做空。

2. 移動平均線 (Moving Average, MA)

移動平均線 (Moving Average, MA)，可稱為均線，對過去 n 天到今日的收盤價，除以 n 天做平均，來表示過去一段的平均成交價格，可快速判斷目前股價強勢或弱勢，以下為移動平均線值計算：

$$MA = \frac{\sum_{i=1}^n C_i}{n} \quad (2-4)$$

台灣股市常使用移動平均期間為 5 日均線 (5MA)、10 日均線 (10MA)、20 日均線 (20MA)、60 日均線 (60MA)、120 日均線 (120MA)、240 日均線 (240MA)。

移動平均線 (MA) 的應用規則：

- (1) 當短週期均線由下往上穿越長期均線，稱之為「黃金交叉」，為買入訊號。
- (2) 當短週期均線由上往下跌破長期均線，稱之為「死亡交叉」，為賣出訊號。

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

相對強弱指標為計算一段時間的股價平均漲幅與平均跌幅組成的比例，來評估目前股價為強勢或是弱勢，RSI 值範圍為 0~100 之間，以下為相對強弱指標公式：

PI_t ：每日上漲幅度， PD_t ：每日下跌幅度

PIA ：n 天漲幅平均值， PDA ：n 天跌幅平均值

$$PIA = \frac{\sum_{i=1}^n PI_t}{n} \quad (2-5)$$

$$PDA = \frac{\sum_{i=1}^n PD_t}{n} \quad (2-6)$$

$$RSI = \frac{PIA}{PIA + PDA} \times 100 \quad (2-7)$$

相對強弱指標的應用規則：

- (1) RSI 值 > 50，表示為強勢市場，上漲幅度較大。
- (2) RSI 值 < 50，表示為弱勢市場，下跌幅度較大。
- (3) RSI 值 > 70，有些人會使用 80 值為界線，表示市場過熱，未來股價可能會下跌，為賣出訊號。
- (4) RSI 值 < 30，有些人會使用 20 值為底線，表示市場過冷，未來股價可



能會上漲，為買入訊號。

- (5) 當短週期 RSI 往上突破長週期 RSI 時，稱為「黃金交叉」。
- (6) 當短週期 RSI 往下跌破長週期 RSI 時，稱為「死亡交叉」。

4. 乖離率(Bias)

乖離率顯示當日股價和移動平均線偏離之程度，是否過度上漲或下跌，以收盤價減去移動平均線 (MA) 差，衡量股價偏離的程度，來判斷買入或賣出的時機，以下為乖離率的公式：

$$P_t : \text{當日股價}$$
$$\text{Bias} = \frac{(P_t - MA)}{MA} \quad (2-8)$$

乖離率的應用規則：

- (1) 正乖離：當日股價 > 移動平均值，表示目前市場處於超買的情況，未來股價下跌機率高。
- (2) 負乖離：當日股價 < 移動平均值，表示目前市場處於超賣的情況，未來股價上漲機率高。

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

指數平滑異同移動平均線由 Gerald Appel [20] 所提出，透過兩條不同週期的指數移動平均值 (Exponential Moving Average, EMA)，計算出兩者間的差離狀態 (DIF)，通常短週期為 12 日，長週期為 26 日，接著將把差離狀態 (DIF) 作為指數移動平均值 (EMA)，通常為 9 日，即為指數平滑異同移動平均線 (MACD)，以下為指數平滑異同移動平均線公式：

$$P_t : \text{當日股價}$$
$$EMA_t = EMA_{t-1} + \alpha \times (P_t - EMA_{t-1}) \quad (2-9)$$

$$DIF = EMA(12) - EMA(26) \quad (2-10)$$

$$MACD = EMA(DIF, 9) \quad (2-11)$$

其中， α 平滑指數 $\alpha = 2 / (1 + \text{移動平均天數})$ 。指數平滑異同移動平均線多

會使用柱狀圖計算差離狀態 (DIF) 及指數平滑異同移動平均線 (MACD) 的差值繪製成「柱形圖」(MACD bar /Oscillator, OSC)。

$$OSC = DIF - MACD \quad (2-12)$$



指數平滑異同移動平均線的應用規則：

- (1) 快線 (DIF) 向上突破慢線 (MACD)，為買入訊號。
- (2) 快線 (DIF) 向下跌破慢線 (MACD)，為賣出訊號。

柱狀圖 (OSC) 的應用規則：

- (1) OSC 由負轉正，為買入訊號。
- (2) OSC 由正轉負，為賣出訊號。

6. 威廉指標 (Williams%R)

威廉指標 Williams 是投資者 Larry R. Williams 所開發的技術指標，用於判斷目前股價是否處於過高或是過低的情況，該指標範圍在 0 到 100% 之間。當指標越小，表示收盤價越接近最高價，股市處於超買程度越大的情況，反之，指標越大，表示市場處於超買程度越大的情況。在實際應用中，會將指標乘上 -1 ，使其值範圍變為 -100 到 0% 之間，指標值越大表示股市處於超買程度越大的情況，反之，值越小表示股市處於超賣程度越大的情況，符合我們對平時認知上的直覺理解。

計算方式：

C_t ：當日收盤價， L_n ：過去 n 日內最低價， H_n ：過去 n 日內最高價

$$\text{威廉指標} = \frac{(C_t - H_n)}{(H_n - L_n)} \times -100\% \quad (2-13)$$

威廉指標的應用規則：

- (1) $0 < \text{威廉指標值} < -20\%$ ，表示市場處於超買狀態。
- (2) $-80 < \text{威廉指標值} < -100\%$ ，表示市場處於超賣狀態。

7. 動量指標 (Momentum Index)

動量指標，也稱為動能，基於動力學原理分析股價波動速度，透過股價波動過程中的加速、減速、慣性作用等等現象。一般採用 10 日動量指標以及 10 日移動平均線（10MA）作為輔助，以下為動量指標公式：

C_t ：當日收盤價， C_{t-n} ： n 日前的收盤價

$$\text{Momentum Index} = C_t - C_{t-n} \quad (2-14)$$

動量指標的應用規則：

- (1) 動量指標上升，動量指標的 n 日移動平均線值由負轉正時，為買進訊號。
- (2) 動量指標下降，動量指標的 n 日移動平均線值由正轉負時，為賣出訊號。

2.3 模糊系統

2.3.1 模糊理論

模糊理論[2]於 1965 年由 L.A. Zadeh 教授所提出，以數學方法解決模糊語意的問題。在日常生活中，每個人對於事物的描述程度可能會有所不同，例如，對於天氣的描述，有人認為低於攝氏 18 度才算冷，而有人則認為低於攝氏 20 度才覺得冷。在餐廳的評價中，有些人認為餐點味道好，給予 9 分的評價，而另一些人則認為餐點好吃但只給出 7 分的評價，對於好吃的定義不同。每個人可能有不同的回答方式或模稜兩可的描述，造成語意上的不確定性和模糊性，模糊理論能夠處理這些模糊或不準確性的描述。

2.3.2 模糊控制

模糊控制利用模糊集合、模糊邏輯和模糊推論來處理模糊性和不確定性的問題。相較於傳統的控制理論需要基於精確的數學模型和精確的輸入，可以不被數學模型所限制輸入，而經過專家經驗、規則及知識來設定控制。圖 1 為模糊控制的架構示意圖。將輸入的數值進行模糊化，轉換為模糊集合的形式，接著，通過模糊推論引擎，根據知識庫與模糊規則進行推論，最後將其解模糊化，得出推論的結果。

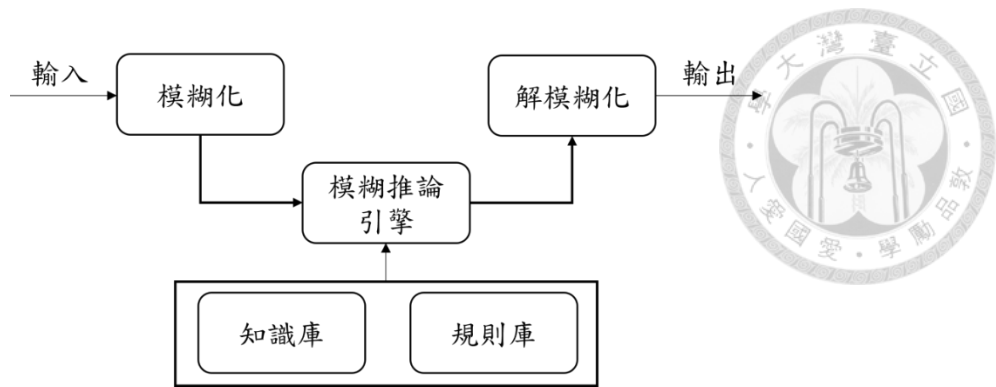


圖 1 模糊控制架構圖

1. 模糊化 (Fuzzify)

模糊化的目的是將輸入的數值映射到模糊集合上，並使用隸屬函數來表示數值對應到每個模糊集合的程度。透過模糊化，可以將具體的數值轉換成模糊集合的形式，進行模糊邏輯處理。

模糊集合用於表示模糊及不確定性的數學函式。在模糊集合中，元素會與函數之間會有關聯，並計算出該集合的隸屬值。隸屬度代表著元素與該集合之間的相關程度，當元素歸屬於該集合越相近，其隸屬值越大。常見的隸屬函數[6]包括高斯型 (Gaussian)、三角形 (Triangular)、梯形 (Trapezoid)、S 函數 (S function)、Z 函數 (Z function)、 Π 函數 (Pi function)，可依需求決定其隸屬函數。

以下介紹本研究會使用的三角形隸屬函數，會以三角形的形狀來描述該模糊集合的隸屬值。如圖 2 所示，有三種模糊集合，分別為服務品質「好」(Amazing)、「壞」(Poor)、「普通」(Acceptable)。

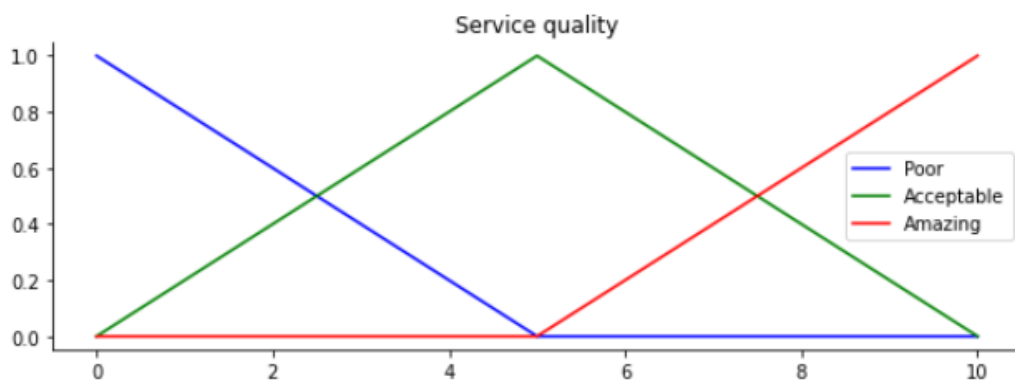


圖 2 定義服務品質的隸屬函數示意圖

2. 模糊推論引擎 (Fuzzy Inference)

當輸入值經過模糊化後，模糊推論引擎會根據規則庫中的規則[5]和知識庫進

行推理計算，引擎再根據輸入值的隸屬函數和條件式，計算該值的隸屬值。

規則庫是由一系列的 IF-THEN 規則所構成，每個規則包含一個「IF」條件式，並用「THEN」描述對應的動作，這些規則是由專家的經驗、知識或是收集大量資料所定義。

3. 解模糊化 (Defuzzify)

模糊引擎使用知識庫和規則進行推論，得出模糊的結果，而解模糊化則是將這些模糊的結果轉換成具體的結果，獲得最合理的結果，以達到建立模糊系統的目的。常見的解模糊化方法有最大中心法 (Center of Maximum)、面積中心法 (Center of Area) 與最大平均法 (Mean of Maximum)。

最大中心法是取最高隸屬值的模糊集合為中心點，將其作為解模糊化的結果。面積中心法則是將模糊集合的隸屬函數視為一個面積，並計算出面積的中心位置，作為解模糊化的結果。最大平均法取最高隸屬值的元素，將最大隸屬值的所有元素取平均值，作為解模糊化的結果。

2.4 基因演算法

基因演算法[4]於 1975 年由 John Holland 於「Adaptation in Natural and Artificial System」中提出[7]，基於達爾文進化論「適者生存，不適者淘汰」的概念，選擇優異的基因進行交配，保留較佳的基因於子代，並在交配過程中固定的機率值發生基因突變，產出最佳的基因，常用於解決最佳化的問題。圖 3 為基因演算法流程架構圖：

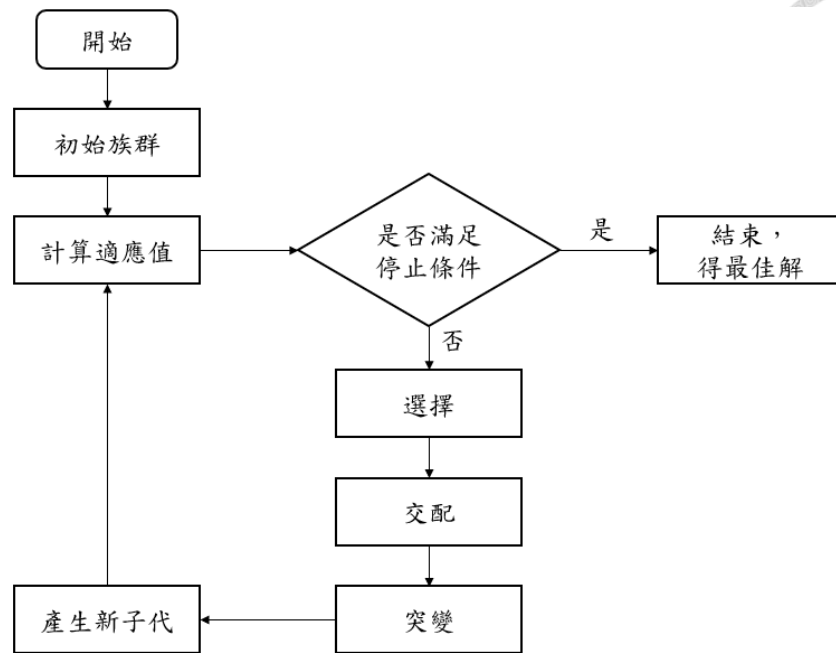


圖 3 基因演算法流程架構圖

2.4.1 初始族群

基因演算法本身是模擬生物演化過程，個體的基因構成染色體。在使用基因演算法之前，需先設定初始族群的大小以及染色體的長度，可由經驗、透過大量數據資料或是隨機亂數來決定，再將要解決此問題的變數編碼轉換成數值。

2.4.2 適應函數 (Fitness Function)

適應函數用來評估每個個體的適應程度即為對環境的適應程度，並且直接影響其存活及繁衍的機會。適應值越高，表示個體的適應程度越好，存活機率相對較高，相反地，適應值越低，則代表個體不適應環境程度越高，越容易被環境所淘汰。

2.4.3 選擇 (Selection Scheme)

基於個體的適應值去決定，適應值越高的個體會較高的機率選中，並進行繁衍產生子代，為了讓優秀的個體有更高的機率遺傳給後者，以提高整體族群的適應度。選擇基因的方法有很多種，常見的有輪盤式選擇 (Roulette Wheel Selection)、競爭式選擇 (Tournament Selection)。

1. 輪盤式選擇 (Roulette Wheel Selection) :



概念來自於射飛鏢的輪盤，有大小不同面積的區域，大區域為適應值高的個體表示越容易選中。隨機選取某一區域去選擇個體，雖然適應度低的個體被選中機率較低，但仍然還是有機率被選中，如圖 4。

F ：被選擇的個體所有適應值總和, f : 適應函數, C_i : 第 i 個個體

N ：個體總數

$$F = \sum_{i=1}^N f(c_i)$$

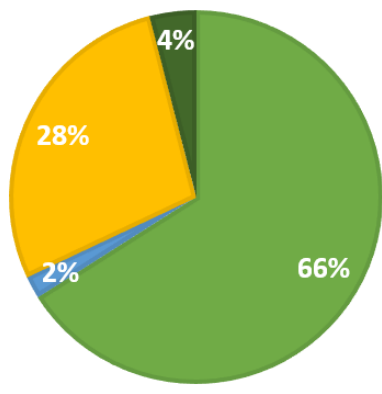


圖 4 輪盤式選擇示意圖

2. 競爭式選擇 (Tournament Selection) :

隨機從族群選出固定大小的子集，在子集中選出兩個個體進行競爭，比較這兩個個體的適應值，適應值高的為勝者，重複選擇的動作，直到選擇足夠數量的個體為止。由於每次只需要比較兩個個體的適應值，可在族群中快速找尋較優的個體，適應值較低的個體也有機會為勝者，故保有族群的多樣性。

2.4.4 交配 (Crossover)

交配指個體交換基因，產生新的子代。選取兩個個體互為交配對象，選擇交配點，在該位置將兩個個體的基因進行切割集劃分，將切割的基因片段互相交換，演化出新的後代，常見的交配方式有單點交配 (Single-Point Crossover)、雙點交配 (Two-Point Crossover)。

1. 單點交配 (Single-Point Crossover) :

如圖 5 選擇一個交配點，將兩個固體的基因在該點並做切割基因，接著，互相交換切割後的基因片段，演化出新的後代。

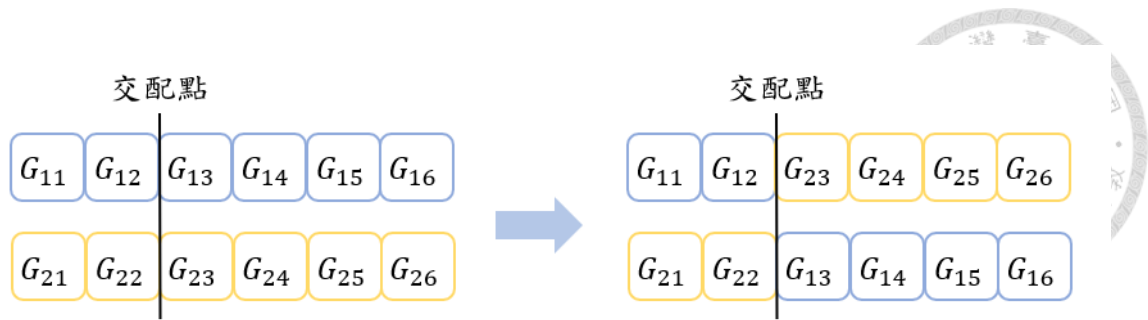


圖 5 基因單點交配示意圖

2. 雙點交配 (Two-Point Crossover) :

圖 6 與單點交配不同於交配點為兩處，先選擇兩個不同的交配點並做切割基因，在兩個交配點之間進行交換基因片段，交換之後演化出新的後代個體。

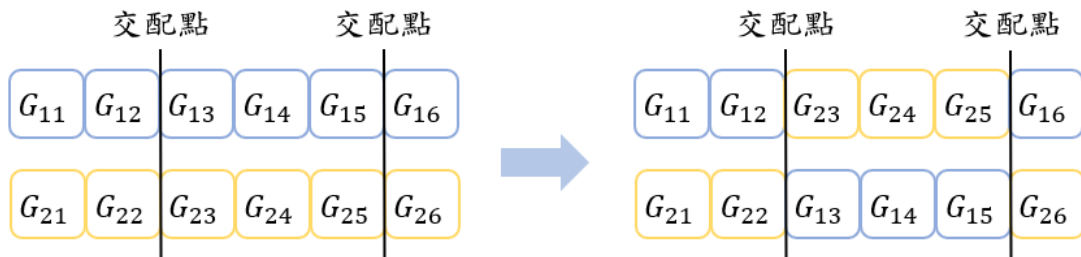


圖 6 基因雙點交配示意圖

2.4.5 突變 (Mutation)

在演化論裡，突變是為了製造適應性更強的個體。在多次的演化過程中，可能因為交配時選擇相似基因的情況，導致下一代的基因組合與父母相似，使其陷入局部最佳解，無法找到最佳的個體。隨機選擇要突變的個體，以固定的機率進行基因突變，以增加基因多樣性。

2.5 類神經網路

2.5.1 循環神經網路 (Recurrent Neural Networks, RNN)

循環神經網路 (Recurrent Neural Networks, RNN) [8]，是時間序列中常見的模型之一，也稱為遞歸神經網路。RNN 圖 3 接收輸入序列中每個元素，將其乘以權重並通過隱藏層的計算，然後將計算結果再傳遞給下一層作為輸入，這個過程會一直重複步驟直到模型收斂結束。在訓練過程中，會計算損失函數值來評估此次

訓練，並將值反向傳播回給每一層，調整每一層的權重。

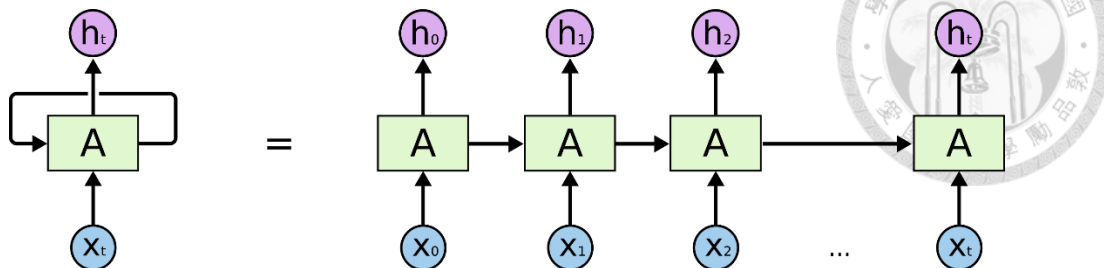


圖 7 RNN 展開示意圖

圖片來源：<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

2.5.2 長短期記憶模型 (Long Short-Term Memory, LSTM)

由於 RNN 有梯度爆炸與梯度消失的問題，導致模型訓練效果不佳。為了解決這個問題，Sepp Hochreiter 和 Jürgen Schmidhuber 於 1997 提出長短期記憶模型 (LSTM) [9]。

LSTM 具備了三種 Gate：Forget gate、Input Gate 以及 Output Gate。會由三種 Gate 來決定記憶單元是否保留或遺忘。

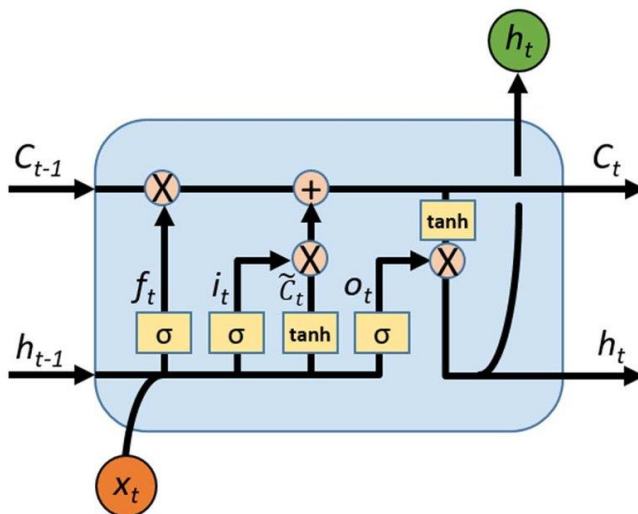


圖 8 LSTM 架構圖

圖片來源：https://www.researchgate.net/figure/Structure-of-the-LSTM-cell_fig1_329902570

Forget Gate f_t 用於判斷上一個時間點的記憶單元 c_{t-1} 是否要保留或是遺忘，將目前輸入值 x_t 與上一個時間點的隱藏狀態 h_{t-1} 乘上一個權重矩陣，並將計算結果輸入至 sigmoid 函式中。Forget Gate 輸出值 f_t 範圍為界於 0 與 1 之間，越大的值代表保留機率越高，而越小的值則代表遺忘機率越高。



$$f_t = \text{sigmoid}(W_f x_t + U_f h_{t-1} + b_f) \quad (2-15)$$

Input Gate 用於控制當前的輸入，將前一時間的隱藏狀態 h_{t-1} 與目前輸入值 x_t 乘上權重矩陣，輸入至 \tanh (Hyperbolic Tangent) 得到 \tilde{c}_t ，藉此同時，將 \tilde{c}_t 輸入到 sigmoid 函數中，得到 i_t ， i_t 會決定 \tilde{c}_t 中哪些訊息應該需被保留。

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (2-16)$$

$$i_t = \text{sigmoid}(W_i x_t + U_i h_{t-1} + b_i) \quad (2-17)$$

$$h_t = o_t \times \tanh(c_t) \quad (2-18)$$

更新長期單元狀態 c_t ，將先前計算得到的 i_t 和 \tilde{c}_t 相乘，以判斷單元是否要進行更新。同時，Forget Gate 的輸出值 f_t 與前一個時間點的單元狀態 c_{t-1} 相乘，再加上當前 i_t 乘以新單元狀態 \tilde{c}_t 。

$$c_t = i_t \times \tilde{c}_t + f_t \times c_{t-1} \quad (2-19)$$

Output Gate 決定下一層隱藏狀態值 h_t ，並將目前輸入值 x_t 和前一個時間點的隱藏狀態 h_{t-1} 經過 sigmoid 激勵函式計算，以決定是否將當前記憶單元 c_t 作為輸出 o_t ，最後得到 Output Gate 輸出值 o_t 。

$$o_t = \text{sigmoid}(W_o x_t + U_o h_{t-1} + b_o) \quad (2-20)$$



2.6 國內外相關研究

張凱婷 [10] 使用線性切割法去切割歷史資料以獲取股價轉折的時機點，並計算每日技術指標再結合支持向量迴歸 (Support Vector Regression, SVR) 與 TS 模糊規則 (Takagi-Sugeno fuzzy rule-based)，學習股票轉折時機點。然而，研究只參考經濟面技術指標，未考慮其他技術指標，因此在有外在因素而影響股市，模型預測效果可能會不太準確。

何公皓 [11] 使用常見的技術性指標：移動平均線 (Moving Average, MA)、隨機指標 (Stochastic Oscillator) 以及布林通道 (Bollinger Band)，由於有時股價會突然地下跌，加上停損點及時停損的交易方式，再搭配多種參數組合成 47 種不同交易策略，以預測台灣 50 (ETF 50)。

吳宜謙 [12] 利用紐約證券交易所及道瓊共十一檔股票的歷史資料，將一維數值資料轉成二維的圖像資料，包含了波峰、波谷的位置以及交易訊號標記。其中交易訊號標記分別為：買入、賣出以及持有，並且在數據轉換過程中使用了 15 種技術指標以及 15 種不同的區間，將數據輸入至深度卷積神經網路進行模型訓練，接著提取重要特徵轉化並將轉化為矩陣，以作為模型訓練的資料，訓練模型學習如何辨識波峰和波谷之訊號。實驗結果比較了 LSTM、CNN 以及其他模型，最終 LSTM 的年化報酬率略勝其他模型。

黃華山、邱一薰 [13] 使用台灣 50 (ETF 50) 歷史資料，計算 12 項技術至指標，再按照台灣 50 成分股中前 20 名股票的佔比重新計算權重，作為模型訓練的資料。倒傳遞類神經網路 (Back Propagation Neural Network, BPNN) 進行股價的預測，使用交易策略獲取報酬率以評估模型。沈沛瑄 [17] 則以台灣 50 以及台灣 50 中佔比率最大的股票台積電的歷史資料作為訓練資料，計算 12 項技術性指標，並使用 LSTM 模型進行股價預測，接著使用校正策略校正模型的預測值，最後結合二次交易策略和價格區間修正法，作為最佳的交易組合策略。

陳柏年 [14] 透過基因演算法 (Genetic Algorithm, GA) 選出最佳股票的組合，計算報酬率並使用夏普率 [23] 作為適應值，作為適應值函數值，進行演化，其中使用 70 種基本面財務指標，該研究結果顯示，最終演化出的投資組合表現優於大盤的報酬率。

Longbing Cao 等人 [15] 運用資料探勘過去金融市場資料，挖掘出股票與市場

之間的實體關係，結合模糊集合和遺傳演算法，以股票規則、股票市場應用規則等知識建立模糊集合，並透過基因演算法演化，最後使用夏普率作為適應函數的適應值去評估。實驗結果發現有 13 個高度相關的股票，有助於實際股市上進行股票交易。





第三章 研究方法及架構

本章節首先介紹本研究使用資料集，第一節將說明整體實驗流程及架構，接著，第二節將列舉資料前處理規則以及使用的技術指標，第三節介紹類神經網路模型，包括調整模型參數以及校正模型等內容，第四節中我們會討論模擬交易規則，最後一節使用基因演算法與模糊理系統，即為基因演化之模糊演算法，去決定買賣時機。

3.1 資料集

資料來自於台灣證券交易所，收集 ETF50 以及 ETF50 成分股佔比最大股票—台積電作為本研究的資料集，研究期間收集了 2003 年 6 月 30 日至 2020 年 12 月 31 日的資料，其中前段 90% 作為訓練資料、後段 10% 作為測試資料，扣除計算技術指標空值，最終共獲得 4297 筆資料。

訓練資料開始時間為 2003 年 8 月 18 號至 2019 年 4 月 8 號，測試資料開始時間為 2019 年 4 月 9 號至 2020 年 12 月 31 號。

3.2 研究流程與架構

本研究首先對資料進行個別的技術指標計算，並扣除初期計算所產生的空值資，再依照台積電佔 ETF50 的成分股比例，計算出總佔比並調整資料權重。接著進行模型訓練並調整模型參數。此外，為了提高模型預測的準確性，我們會對模型預測值進行校正，以符合當前股市的趨勢。最後，結合基因演算法和模糊系統來計算出最佳模糊隸屬函數的參數，並作為買賣策略。圖 9 為本研究架構。

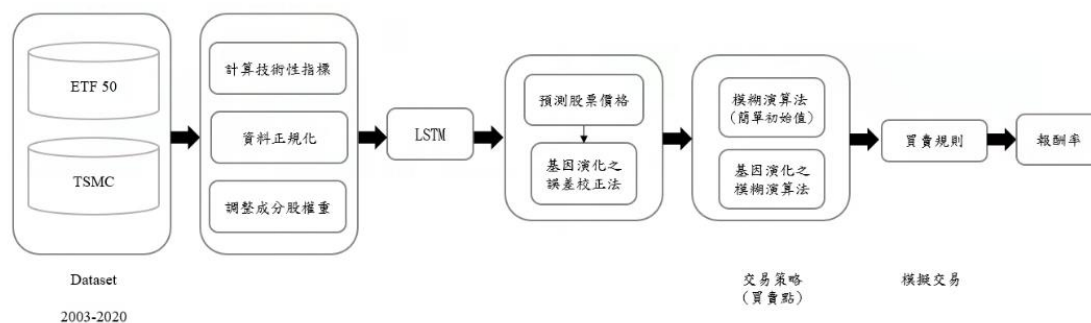


圖 9 研究流程圖



3.3 資料前處理

3.3.1 計算技術指標

使用每日股價、成交量以及 9 種技術指標，共有 14 個變數，計算每日的技術指標並將其作為類神經網路訓練資料輸入。模型的預測目標為每日股價的最高價、最低價以及收盤價，共 3 個模型。以下表 1 列出為預測最高價模型輸入變數，而最低價以及收盤價的預測也是依此類推。

表 1 模型輸入變數

變數		說明
X_1	High	最高價
X_2	V	成交量
X_3	K	9 日 K 值指標
X_4	D	9 日 D 值指標
X_5	MA 6	6 日移動平均線
X_6	MA 12	12 日移動平均線
X_7	Bias 6	6 日乖離率
X_8	Bias 3-6	3-6 日乖離率 (3 日乖離率-6 日乖離率)
X_9	RSI 6	6 日強弱指標
X_{10}	W%R 12	12 日威廉指標
X_{11}	MTM 6	6 日動量指標
X_{12}	MTM (avg 6)	6 日動量指標均線
X_{13}	MACD	9 日平滑異同移動平均線指標
X_{14}	DIF	差離值 (短期 12 日，長期 26 日)
X_{15}	TSMC High	台積電最高價

3.3.2 資料正規化 Normalization

Min-max scaling 是一種常見的正規化方法，將數值特徵資料縮放於 [0,1] 之間，以限制數值在一定範圍內並加速模型訓練速度。以下為正規化的公式：



$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \in [0,1] \quad (3-1)$$

3.3.3 調整成分股權重

資料中會使用到 ETF50 最大成分股的股票歷史資料—台積電，可以根據台積電在 ETF50 成分股中的百分比比例，調整整個股票資料的權重。首先，先將 ETF50 權重設為 100%，接著根據台積電在 ETF50 成分股中的比例，計算出調整後的股票權重。最後，將正規化後的資料乘以調整後的權重，以考慮到 ETF50 包含多種不同比例的股票。根據沈沛瑄 [17] 的實驗結果顯示，只加入第一大成分股納入訓練模型可以提高模型的穩定，台積電在 ETF50 成分股佔比百分比最大，故本研究只考慮台積電股票作為訓練資料。

W_i ：股票之權重比例， W_T ：股票總權重

$W_T : W_{i/T}$ ：調整後的股票之權重（百分比）

$$W_T = \sum W_i \quad (3-2)$$

$$W_{\frac{W_i}{W_T}} = \frac{W_i}{W_T} \times 100\% \quad (3-3)$$

3.3.4 切割資料

使用長短期記憶模型 (LSTM) 模型進行訓練時，需要將要預測的每個時間點切割成固定長度時間的時間序列資料，使用滑動視窗概念去實作。切成長度相同的時間序列資料，設定時間長度為 20 日，股票資料第 1 天至第 20 天的資料作為輸入資料，去預測第 21 天股價作為模型預測目標，如圖 10，深色框中的資料為輸入資料，紅字為此次模型的預測目標。最後將切割完成的資料分為訓練資料集和測試資料集，其中 90% 為訓練資料集，10% 為測試資料集，驗證資料集佔訓練資料集的 10%。

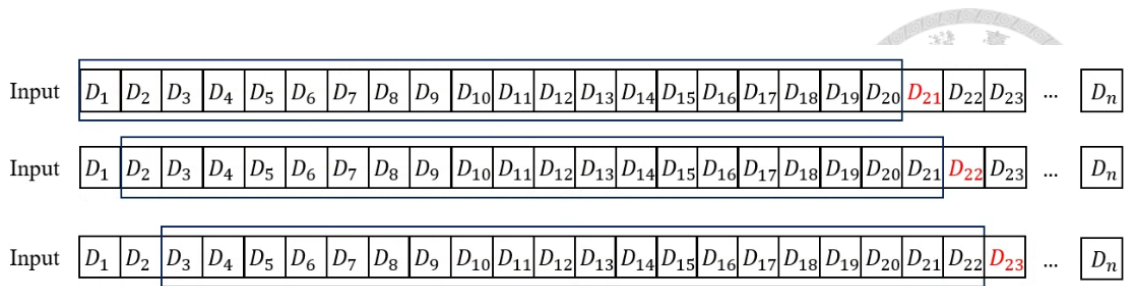


圖 10 切割資料示意圖

3.4 LSTM 類神經網路模型與校正策略

3.4.1 LSTM 類神經網路架構

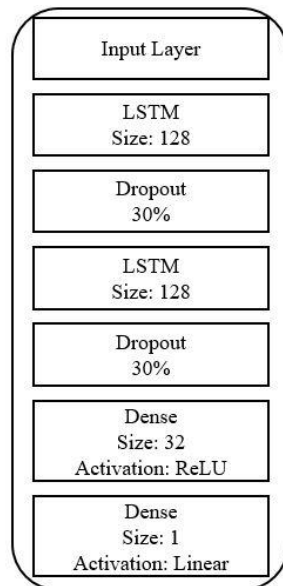


圖 11 LSTM 架構圖

圖 11 為本研究 LSTM 類神經網路模型的架構圖，使用 Keras [21]作為 LSTM 的實作框架。該模型包含兩層 LSTM Layer，為了預防模型有過度擬合 (overfitting) 的問題，加了 Dropout Layer 來解決，最後，加上兩層 Dense Layer，用於輸出預測的結果。

3.4.2 模型訓練參數

下表 2 為本研究實驗訓練模型參數[17]，使用 Adam 作為梯度下降的方法，最後以 MSE 作為損失函數：

表 2 模型訓練參數設定

層數	Layer Name/參數	說明	參數設定
1	LSTM Layer	神經元個數	128
2	Dropout Layer	丟棄神經元之比例	0.3
3	LSTM Layer	神經元個數	128
4	Dropout Layer	丟棄神經元之比例	0.3
5	Dense Layer	激活函數	ReLU
5	Dense Layer	神經元個數	32
6	Dense Layer	激活函數	Linear
6	Dense Layer	神經元個數	1
	optimizer	梯度下降法	Adam
	Loss function	損失函數	MSE

3.4.3 模型評估方式

本研究使用 MSE (mean-square error) 作為模型的損失函數 (Loss Function)，用來評估模型。在每次訓練後，會計算出 MSE 會反饋給類神經網路模型，以調整每一層的權重。當 MSE 越小代表模型誤差越小為模型預測精準度越高，越大則代表模型誤差越大為模型預測的精準度較低，以下為 MSE 計算公式：

n ：樣本大小， y ：實際值， \hat{y} ：預測值

$$MSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 / n \quad (3-4)$$

3.4.4 誤差校正法

模型預測股價可能與當日實際股價存在微小的誤差，然而該誤差可能會影響後續的買賣時機以及交易時的最高價和最低價。為了解決這個問題，會使用校正方法去修正模型預測值 [17]，雖然校正方法可以降低預測誤差，但僅能對隔日的預測進行校正，無法改善未來股價趨勢。因此，本研究提出誤差校正法，利用當日預測股價的誤差來判斷預測值是否需要進行校正，並校正日後的股價。以下圖 12 是計算誤差值並判斷是否進行校正的虛擬碼：



R_t ：今日實際股價， P_t ：模型預測今日股價， δ : 校正閾值

C_{last} ：上次誤差值， C_t ：今日誤差值

Algorithm 1 Deviation Calibration

```

1:  $P_t = P_t + C_{last}$ 
2:  $C_t = R_t - P_t$ 
3: if  $|C_t| > (R_t \times \delta)$  then
4:    $C_{last} = C_{last} + C_t$ 
5: end if

```

圖 12 誤差校正法虛擬碼

當今日股市交易結束時，會計算今日股價與模型預測今日股價差，作為預測誤差值，當預測誤差值超過 ETF50 漲跌 $\pm 8\%$ ，也就是超過校正閾值時，則會對日後預測的股價進行校正。

3.4.5 基因演化之誤差校正法

為了降低模型預測的誤差並找出最佳的校正閾值 δ ，本研究使用基因演算法進行演化以獲得最佳值。

1. 適應函數：

在基因演算法中，適應函數用於計算適應值，以評估個體在環境的適應程度，適應值越大表示個體存活機率越高。我們會計算經過誤差校正法後 MSE 值作為適應函數來計算誤差的適應值。

2. 基因演化之誤差校正法訓練參數及初始值：

表 3 為基因演算法訓練參數設定及初始值，校正閾值初始值設為 0.05，總共演化 15 次作為停止條件，由於本研究的問題複雜程度不高，可使用 Steady State Selection [22] 方法去選擇個體進行繁衍且計算成本較低，選擇後進行單點交叉交配以演化下一代個體，交配過程中設定 1% 突變率。

表 3 基因演算法訓練參數

參數名稱	值/方法
初始值	0.05
演化次數	15
選擇方法	Steady State Selection
突變率	1%
交配方法	單點交叉
sol_per_pop	48
init_range_low	0.0001
init_range_high	0.1
num_parents_mating	9

3.5 交易策略

3.5.1 買賣規則

1. 買入以一股為單位買入，賣出則以一股為單位賣出。
2. 每次買入或賣出時的交易手續費為成交價的 1.425%。
3. 每次賣出時所需支付的證券交易稅為成交價的 1%。
4. 本研究無融資融券交易且不考慮放空回補。
5. 必須在持有該股的情況下才能執行賣出股票的動作。
6. 模擬交易則為一買一賣的方式作為一次完整的交易，若最後一次交易動作為買入，則扣除該買入所花費之成本。
7. 當模擬交易結束時，如投資者手上還持有股票，為了計算最後的報酬率，則將以當日收盤價計算成成本。
8. 每次交易的買價為模型預測校正後當日的最高價，賣價為模型預測校正後當日的最低價。
9. 交易失敗即為當日實際最低價 \leq 買價或賣價 \leq 當日實際最高價，當今日交易失敗時則改以隔日進行交易。
10. 若連續兩日交易失敗，則停止交易動作。
11. 本研究以 1,000,000 為股票買賣的實驗總資金。



3.5.2 投資報酬率

本研究採用投資報酬率作為評估交易策略和買賣策略的指標，投資報酬率計算方式是將賣出的總收益扣去買入的總成本，再除以買入總成本，最後以取百分比表示。如下為報酬率公式：

$$\text{投資報酬率(\%)} = \frac{\text{賣出總收益} - \text{買入總成本}}{\text{買入總成本}} \times 100 \quad (3-5)$$

3.6 基因演化之模糊演算法

本研究使用 K、D 以及 RSI 技術指標，建立了買入 (BUY)、賣出 (SELL) 以及持有 (HOLD) 隸屬函數。為了找出最佳的隸屬函數參數值，故結合基因演化之模糊演算法找出最佳值，以加強買賣策略。本研究使用 pygad 套件 [16] 來實作基因演算法。

3.6.1 技術指標之應用規則

我們可設計技術指標應用規則，當小於或大於門檻時，做買入賣出或持有的訊號，以下列出三個技術指標應用規則：

1. K 指標之應用規則：
 - a. 若 K 值 \leq KB_H 時，則為 BUY 訊號。
 - b. 若 K 值 \geq KS_L 時，則為 SELL 訊號。
 - c. 若 $KH_L < K$ 值 $< KH_H$ 時，則為 HOLD 訊號。
2. D 指標之應用規則：
 - a. 若 D 值 \leq DB_H 時，則為 BUY 訊號。
 - b. 若 D 值 \geq DS_H 時，則為 SELL 訊號。
 - c. 若 $DH_L < D$ 值 $< DH_H$ 時，則為 HOLD 訊號。
3. RSI 指標之應用規則：
 - a. 若 RSI 值 \leq RB_H 時，則為 BUY 訊號。
 - b. 若 RSI 值 \geq RS_H 時，則為 SELL 訊號。



c. 若 $RH_L < RSI \text{ 值} < RH_H$ 時，則為 HOLD 訊號。

利用 K、D 以及 RSI 三個技術指標建立規則[24]，如表 4：

表 4 技術指標買賣訊號與 ACTION 之規則

RSI 值	BUY			HOLD			SELL		
D 值									
K 值	BUY	HOLD	SELL	BUY	HOLD	SELL	BUY	HOLD	SELL
BUY	A.BUY	A.BUY	A.BUY	A.BUY	A.HOLD	A.HOLD	A.BUY	A.SELL	A.SELL
HOLD	A.BUY	A.HOLD	A.BUY	A.HOLD	A.HOLD	A.HOLD	A.SELL	A.HOLD	A.SELL
SELL	A.BUY	A.HOLD	A.BUY	A.HOLD	A.HOLD	A.SELL	A.SELL	A.SELL	A.SELL

由於 HOLD 表示無做任何動作，我們可以從邏輯上優化表 4 的規則，可以簡化為：

1. 若 K 指標與 D 指標皆為 BUY 訊號時，則 ACTION 為 A.BUY。
2. 若 K 指標與 D 指標的皆為 SELL 訊號，則 ACTION 為 A.SELL。
3. 若 K 指標與 RSI 指標皆為 BUY 訊號時，則 ACTION 為 A.BUY。
4. 若 K 指標與 RSI 指標皆為 SELL 訊號時，則 ACTION 為 A.SELL。

3.6.2 隸屬函數 (Membership Function)

將上述技術指標應用規則去建立隸屬函數，如表 5，共有 16 個隸屬函數值之參數，以及輸出之隸屬函數。

表 5 技術指標與輸出之隸屬函數參數設定

	BUY	HOLD	SELL
K 指標	$0-0-KB_H$	$KH_L - ((KH_L + KH_H)/2) - KH_H$	$KS_L - 100 - 100$
D 指標	$0-0-DB_H$	$DH_L - ((DH_L + DH_H)/2) - DH_H$	$DS_L - 100 - 100$
RSI 強弱指標	$0-0-RB_H$	$RH_L - ((RH_L + RH_H)/2) - RH_H$	$RS_L - 100 - 100$
ACTION	$0-0-AB_H$	$AH_L - ((AH_L + AH_H)/2) - AH_H$	$AS_H - 100 - 100$



當 K 值 $\leq KB_H$ 時，則 ACTION 為 BUY，當 K 值介於 KH_L 與 KH_H 之間時，則判定為 HOLD；當 K 值 $\geq KS_L$ 時，則判為 SELL。對於其他指標，採用如同上述的方式進行判定，如圖 13。

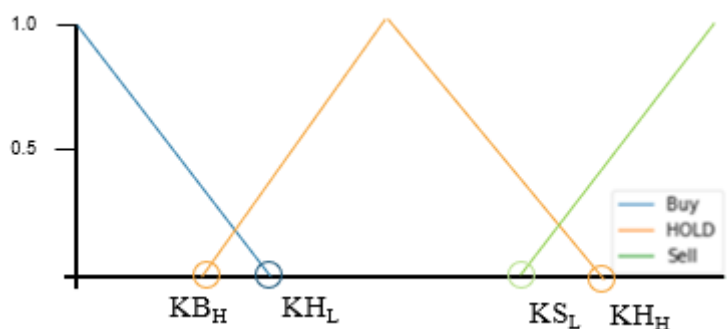


圖 13 K 值隸屬函數

3.6.3 隸屬函數初始值

定義隸屬函數初始值可以根據經驗去定義或是以亂數的方式產生。表 6 中 K 值介於 0 至 100 之間， K 值 80 以上代表市場過熱，會有機會反轉下跌，可以將股票賣出 (SELL)， K 值 20 以下表示市場過冷，會有機會反轉上漲，可以將股票買入 (BUY)， K 值介於 20 至 80 之間則不做任何動作即為持有 (HOLD)，表 6 為 K 、 D 以及 RSI 指標隸屬函數參數初始值所建立的隸屬函數。

表 6 技術指標之隸屬函數初始值

	BUY	HOLD	SELL
K 指標	0-0-20	20-50-80	80-100-100
D 指標	0-0-20	20-50-80	80-100-100
RSI 強弱指標	0-0-30	30-50-70	70-100-100
ACTION	0-0-20	20-50-80	80-100-100

3.6.4 基因演算法輸入變數

如表 7， G_1 至 G_4 為 K 指標隸屬函數之參數， G_5 至 G_8 為 D 指標隸屬函數之參數， G_9 至 G_{12} 為 RSI 指標隸屬函數之參數。

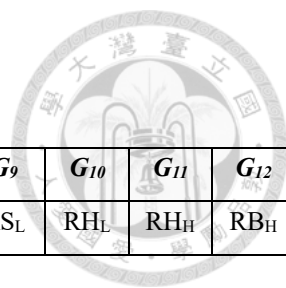


表 7 基因演算法輸入變數

基因變數	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9	G_{10}	G_{11}	G_{12}
隸屬函數參數	KS _L	KH _L	KH _H	KB _H	DS _L	DH _L	DH _H	DB _H	RS _L	RH _L	RH _H	RB _H

3.6.5 適應函數

演化出新的參數去以輸入至隸屬函數中，並建立模糊系統，透過模糊系統去決定買賣時機，進行交易接著計算報酬率作為適應函數，評估此次新演化的參數是否為最佳解。

3.6.6 基因演化之模糊演算法訓練參數

表 8 為基因演化之模糊演算法訓練參數設定與方法，其說明如下：訓練演化總共 50 次，使用 Steady State Selection 方法去選擇個體進行繁衍，選擇後進行單點交叉交配以演化下一代個體，交配過程中設定 1% 突變率。

表 8 基因演化之模糊演算法訓練參數

參數名稱	值/方法
演化次數	50
選擇方法	Steady State Selection
突變率	1%
交配方法	單點交叉
sol_per_pop	48
init_range_low	1
init_range_high	99
num_parents_mating	32



第四章 實驗結果

4.1 模型預測

4.1.1 模型預測校正比較

模型預測股價的 MSE 為 11.5758，經過誤差校正後，MSE 降至 1.2687。透過誤差校正法成功降低了 10.3071 的誤差，有效地改善未來的股價。圖 14 展示了預測股價的每日收盤價，圖 15 則展示了經過校正後的每日收盤價。在圖中，灰色點代表當日實際股價，而黑色點代表預測股價。可以觀察到在紅框標記的區間內，即為 2020 年 3 月至 2020 年 4 月，實際股價與預測股價之間在明顯的分離，即存在很大的誤差。

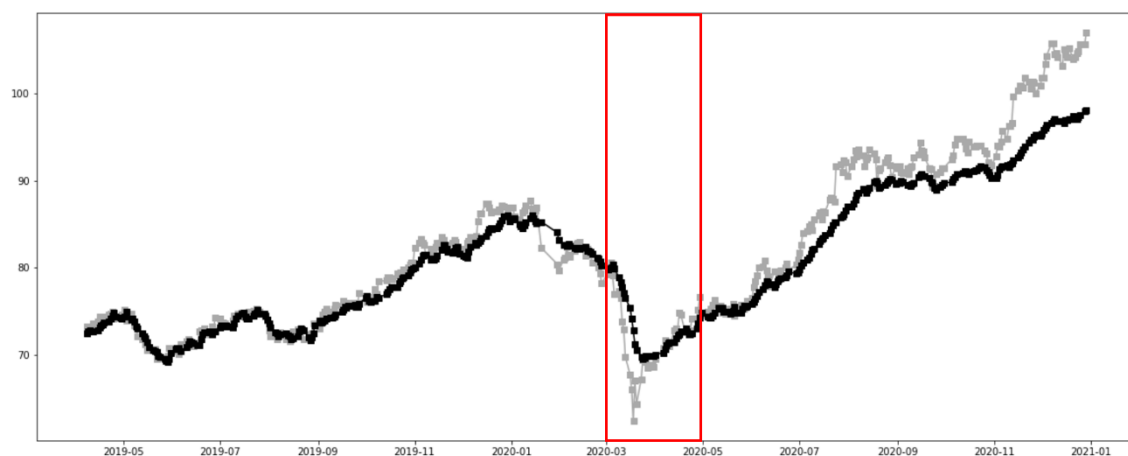
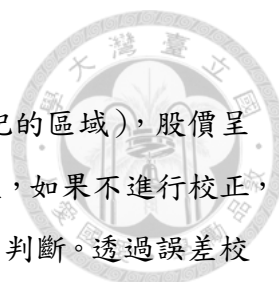


圖 14 模型預測股價之每日收盤價圖



圖 15 模型校正後預測之每日收盤價圖



從圖 14 與圖 15 觀察到，在 2020 年 3 月開始（紅框標記的區域），股價呈現持續下跌的趨勢，模型此時預測股價與實際股價開始產生誤差，如果不進行校正，未來的股價誤差值可能會越來越大，進而影響到決定買賣時機的判斷。透過誤差校正法，我們可及時校正股價。例如 2020 年 12 月 9 日預測值為 109.53369，而當日的實際股價為 120.04999，預測值與實際值之間存在偏差，經過校正後，修正後的值為 120.38567，校正了 10.85198364 的誤差，使預測結果變得更精準。

4.1.2 誤差校正法與基因演算法

為了決定誤差校正法中的校正閾值，我們使用基因演算法去演算最佳校正閾值，表 9 比較了誤差校正法與結合基因演算法後的 MSE 值：

表 9 校正閾值對誤差校正法 MSE 比較

	誤差校正法	基因演化之誤差校正法
最高價 (MSE)	1.30058	1.26394
收盤價 (MSE)	1.05160	0.99267
最低價 (MSE)	1.21169	1.16202

可從表 9 中發現，在經過基因演算法獲得最佳的校正閾值後，最高價、收盤價和最低價的 MSE 均有下降。

4.1.3 當日最高價與最低價

實驗中共有三個模型，分別針對最高價、最低價以及收盤價進行預測。為了確認三個模型預測出來的股價是否符合股價實際情況，在校正後進行檢測。在這裡，我們觀察當日最低價是否大於最高價的情形，結果顯示只有在第一天發生此情況。

4.2 基因演化之模糊演算法

模型預測值經過校正後，使用技術指標之應用規則去建立模糊系統，決定買賣時機，為了找出最佳隸屬函數值，使用基因演算法找出最佳參數值。將技術指標模

糊隸屬函數參數輸入到基因演算法進行訓練，並使用基因演算法得出的參數建立模糊系統，經過股價計算技術指標透過模糊系統去決定買賣時機，以模擬交易進行交易，計算報酬率作為適應函數，經過幾次演化，獲得最佳隸屬函數之參數。

表 10 基因演算法得出的最佳隸屬函數參數值 (Membership Function)

INPUT/ACTION	BUY	HOLD	SELL
K 指標	0-0-31	9-41-74	68-100-100
D 指標	0-0-18	8-51-95	71-100-100
RSI 指標	0-0-54	32-48-64	62-100-100
ACTION	0-0-17	8-51-95	71-100-100

從表 10 可以觀察到，K 指標和 D 指標的 BUY 與 SELL 隸屬函數參數與原先初始值相比差異不大。K 指標 HOLD 隸屬函數參數 KH_L 與初始值 20 有些差異，經過基因演算法的調整後， KH_L 的值變為 9。D 指標的情況與 K 指標相似，初始值為 20 的 DH_L 參數經過演算後變為 8。RSI 指標 BUY 的隸屬函數參數 RB_H 與初始值 30 存在有比較大的差異。整體而言，16 個參數經過基因演算法訓練，確實發揮了優化參數的作用。

圖 16、圖 17 以及圖 18 為 K 指標、D 指標以及 RSI 指標得到最佳參數所建的隸屬函數。

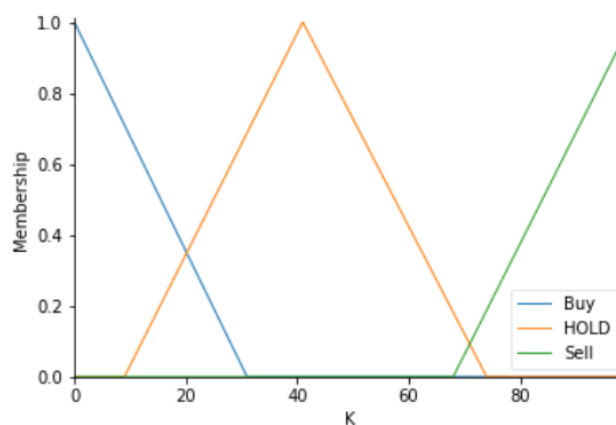


圖 16 K 值最佳隸屬函數

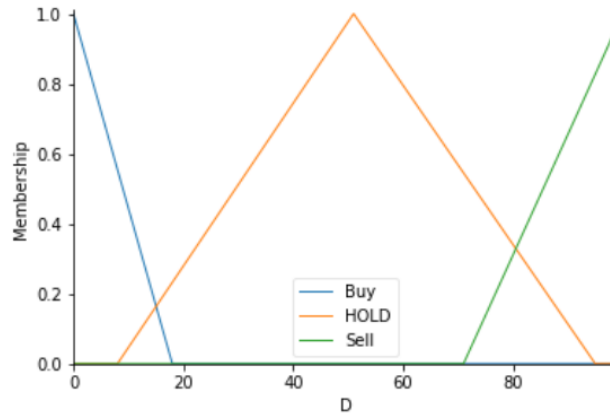


圖 17 D 值最佳隸屬函數

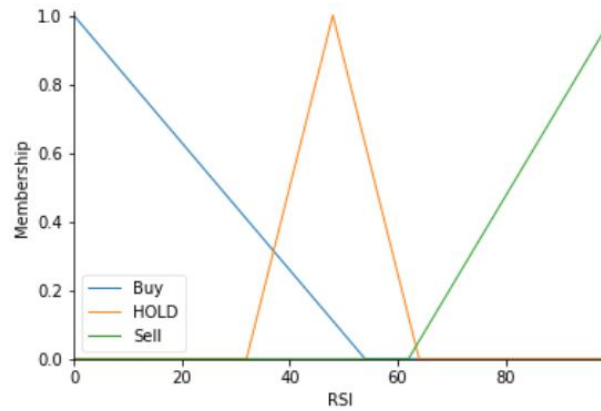


圖 18 RSI 值最佳隸屬函數

4.3 交易策略比較

交易策略是決定適當的進出場或是買賣時機，本研究會比較「一次轉換」[13]、「模糊系統(使用初始值)」、「基因演化之模糊演算法」三種交易策略，如表 11。

隔日預測收盤價減去今日實際收盤價去除以今日收盤價作為一次轉換的成本，當買入成本超過上限門檻 1.425% (手續費) 則買入；反之，當賣出成本超過 2.425% (手續費 + 證交稅) 則賣出。

P_{t+1} : 隔日預測收盤價， R_t : 今日實際收盤價

$$\text{一次轉換成本} = \frac{P_{t+1} - R_t}{R_t} \quad (4-1)$$

表 11 不同交易策略報酬率比較

交易策略	投資報酬率(%，百分比)
一次轉換	4.7
模糊系統(使用初始值)	4.72
基因演化之模糊演算法	<u>32.0</u>

由表 11 觀察可以得知使用基因演化之模糊演算法可以得到 32.0% 的報酬率，使用一次轉換和模糊系統作為交易策略都可以有 4.7% 的報酬率，最終，我們選擇「基因演化之模糊演算法」作為交易策略。

4.3.1 股價校正後分析

測試集資料從 2019 年 4 月 9 號開始預測股價，在 2020 年上半年遇到 Covid-19 事件，影響了金融市場，導致模型預測股價誤差值越來越大，經過校正之後，如圖 19 黑點所示，紅點表示股票買入的時間點，綠點為股票賣出的時間點，黃框所框的時間區間為 Covid-19 在台灣影響最嚴重的時間區間。

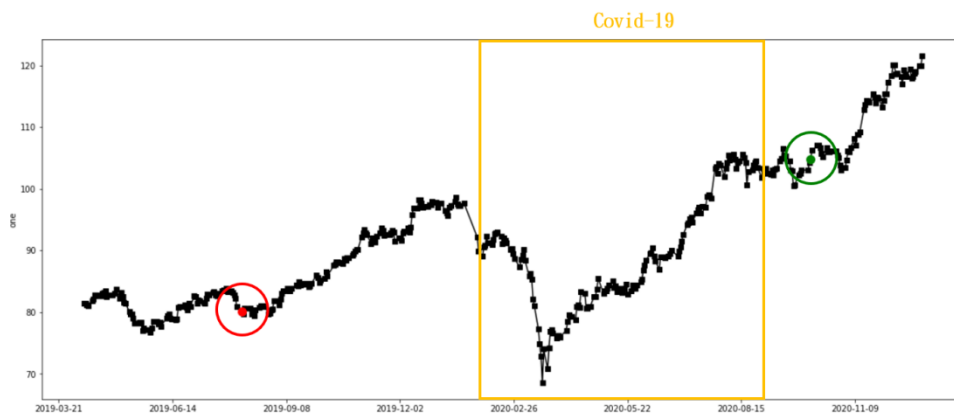


圖 19 Covid-19 影響 ETF50 股價圖

可以發現雖然 Covid-19 影響股價非常多，但是透過基因演化之模糊演算法去找到最佳的買賣點，雖然只有一個回合的交易，但賣點正處於上漲的趨勢，也是我們希望的高點賣出，故報酬率相對的高。



4.3.2 不同的總資金比較

表 12 列出了不同的總資金所得到的報酬率。當總資金為 500,000、100,000 時，報酬率皆為 32.0，不同的總資金下，報酬率皆相同。

表 12 不同總資金的報酬率比較表

總資金	報酬率(%)
1,000,000	32.0
500,000	32.0
100,000	32.0

4.3.3 基因演化之模糊演算法報酬率比較

表 13 深入觀察不同時間區間的報酬率表現，以半年為實驗區間，比較買入持有與基因演算法兩種不同策略。最佳(理想)報酬率為第一天用最低價買入股票，最後一天以最高價賣出計算報酬率。買入持有是在第一天買入，最後一天賣出，中間不做任何交易。表 13 為不同時間區間報酬率表，以黑底線劃記表示基因演化之模糊演算法的報酬率高於買入持有策略。在 2019/10/1-2020/3/31、2019/11/1-2020/4/30、2019/12/2-2020/5/29、2020/1/2-2020/6/30 的區間內，買入持有的報酬率皆為負值，但基因演化之模糊演算法仍能獲得獲利率，這表示即使在難以取得獲利率的情況下，基因演化之模糊演算法能穩定獲利。在 2020/5/4-2020/10/30、2020/6/1-2020/11/30、2020/7/1-2020/12/31 的區間，股價開始大幅上漲，買入持有的報酬率高於基因演算法，在大漲的情況下技術指標可能沒抓取最佳時間點做交易。

表 13 不同時間區間報酬率比較



時間區間	最佳(理想報酬率)	買入持有報酬率	基因演化之模糊演算法報酬率
2019/4/9-2019/9/27	5.38	3.93	2.72
2019/5/2-2019/10/31	9.28	8.55	6.82
2019/6/3-2019/11/29	19.57	16.48	10.62
2019/7/1-2019/12/31	18.46	17.30	<u>21.60</u>
2019/8/1-2020/1/31	10.41	9.29	<u>19.76</u>
2019/9/2-2020/2/27	9.61	8.24	<u>9.18</u>
2019/10/1-2020/3/31	-8.96	-11.40	<u>18.02</u>
2019/11/1-2020/4/30	-4.73	-5.10	<u>6.05</u>
2019/12/2-2020/5/29	-7.95	-8.25	<u>0.46</u>
2020/1/2-2020/6/30	-7.31	-7.93	<u>4.29</u>
2020/2/3-2020/7/31	19.06	16.56	<u>19.44</u>
2020/3/2-2020/8/31	20.03	16.54	<u>25.27</u>
2020/4/1-2020/9/30	36.38	35.70	25.27
2020/5/4-2020/10/30	26.03	24.09	0.0
2020/6/1-2020/11/30	36.43	33.45	0.0
2020/7/1-2020/12/31	35.18	35.08	0.0

第五章 結論與未來展望



5.1 結論

雖然股市趨勢變化很快可能因為多種原因而造成股價波動大，而此時只要使用本研究所提出的誤差修正法，不但可以即時修正未來股價的趨勢，也可以讓預測的誤差值降低，達到進一步預測股價精準的目的。

此外我們利用模糊隸屬函數透過基因演算法，經過多次的演化，演算出最佳的技術性指標的隸屬函數參數，以建立模糊系統，並結合技術指標應用規則，進行推論出股價最佳的買賣時機，進而實際進行模擬交易，得到報酬率。

5.2 未來展望

此章節會進一步討論校正方法改善以及改善模糊系統，並未來期望改善的部分有兩點整理以下兩點：

首先，我們可以增加更多技術性指標的隸屬函數與應用規則，來判斷何時為最佳買賣點，以達到最佳的成果。


其次，雖然誤差校正法能降低每日預測誤差值，但校正次數太頻繁，可能會是因為某幾天的校正值異常很大或很小，導致未來預測股價的校正值有很大的波動，因此應該要限制校正值在一個合理的範圍內，可以透過前幾日的資料和誤差值，推斷未來校正值應該為多少，以減少校正次數。透過這兩點的改善，可以讓我們的模型更精確地預測股價，提高報酬率。

最後，基因演化模糊演算法在股市大漲時期雖仍有報酬率，但沒辦法獲利太高，較為保守的交易策略，可加入布林通道等指標輔助進場時機。

參考文獻



- [1] 臺灣證交所 - 指數股票型基金 (ETF) 簡介。檢自 <https://www.twse.com.tw/zh/page/ETF/intro.html>
- [2] Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8(3), 338-353.
- [3] Allen, F., & Karjalainen, R. (1999). Using genetic algorithms to find technical trading rules. *Journal of Financial Economics*, 51(2), 245-271.
- [4] Beasley, D., Martin, R. R., & Bull, D. R. (1993). An overview of genetic algorithms: Part 1. Fundamentals. *University Computing*, 15, 58-58.
- [5] L. X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 6, pp. 1414-1427, 1992.
- [6] J. S. R. Jang, "Self-learning fuzzy controller based on temporal backpropagation," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 714-723, 1992.
- [7] J. H. Holland. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- [8] Elman, J. L. (1990). Finding structure in time. *Cognitive Science* 14(2), 179-211.
- [9] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- [10] 張凱婷(2011)，應用支撐向量迴歸及模糊規則於股價買賣點之預測，元智大學資訊管理學系碩士論文。
- [11] 何公皓(2016)，技術分析投資績效之實證分析—以台灣 50 ETF 為例，國立台灣大學管理學院國際企業學系暨碩士論文。
- [12] 吳宜謙(2021)，使用卷積-長短期記憶神經網路進行股票交易，國立清華大學計算與建模科學研究所碩士論文。
- [13] 黃華山、邱一薰(2005)，類神經網路預測臺灣 50 股價指數之研究，國立彰化師範大學資訊管理研究所碩士論文。
- [14] 陳柏年(2001)，應用遺傳演算法於財務指標選股策略之探討，國立中央大學資訊管理研究所碩士論文。
- [15] L.Cao, C. Luo, J. Ni, D. Luo, C. Zhang. (2006). Stock Data Mining through Fuzzy Genetic Algorithm. In *Proceedings of the 2006 Joint Conference on Information Science (JCIS 2006)*, Kaohsiung, Taiwan: DBLP.

- 
- [16] PyGAD. Retrieved from <https://pygad.readthedocs.io/en/latest/>
- [17] 沈沛瑄(2020)，以 LSTM 結合二次交易策略預測 ETF 50股價趨勢，國立台灣大學工程科學及海洋工程所碩士論文。
- [18] Lee, K. J., & Lu, S. L. (2021). The Impact of COVID-19 on the Stock Price of Socially Responsible Enterprises: An Empirical Study in Taiwan Stock Market. *International Journal of Environmental Research and Public Health*.
<https://doi.org/10.3390/ijerph18041398>
- [19] Lane, G. (1950). The Stochastic Indicator. *Technical Analysis of Stocks & Commodities Magazine*.
- [20] Appel, G. (1985). The moving average convergence-divergence trading method: advanced version. *Scientific Investment Systems*.
- [21] Keras. Retrieved from <https://keras.io/>
- [22] Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley.
- [23] William F. Sharpe, "A Theory of Market Equilibrium under Conditions of Risk", *The Journal of Finance*, vol. 19, no. 3, Sep. 1964, pp. 425-442.
- [24] 葉文宏(2022)，模糊決策系統應用於台灣股市實證研究，國立宜蘭大學電機資訊學院碩士論文。