

國立東華大學資訊工程系

National Dong Hwa University

112 學年度大學部畢業專題研究報告

112 CSIE Undergraduate Project Report

利用股價預測模型與投資組合管理之多元股票資訊網站



指導教授 Advisor :	高韓英教授
專題參與人員 Team Member :	鄭羽農
	陳仁傑
	吳育臻
	高鈺雯

中 華 民 國      113      年      5      月      20      日

# 國立東華大學資訊工程學系

## 專題報告原創性聲明

### National Dong Hwa University

#### Department of Computer Science and Information Engineering

#### Statement of Originality

本人鄭重聲明：

所呈交的專題報告是在指導老師指導下進行的研究工作及取得的研究成果。除文中已經註明引用的內容外，本報告不包含任何其他個人或集體已經發表或撰寫過的研究成果。對本文的研究做出重要貢獻的個人與集體，均已在文中以明確方式標明。若有違上述聲明，願依校規處分及承擔法律責任。

I hereby affirm that the submitted project report is the result of research under the supervision of my advisor. Except where due references are made, the report contains no material previously published or written by another person or group. All significant facilitators to the project have been mentioned explicitly. Should any part of the statement were breached, I am subject to the punishment enforced by the University and any legal responsibility incurred.

學號 Student No.	學生姓名 Name	親筆簽名 Signature
410942017	鄭羽農	
410942019	陳仁傑	
411021245	吳育臻	
411021248	高鈺雯	

中 華 民 國      113      年      5      月      20      日

## 摘要

本專題針對台灣股票市場，對一個月後和四個月進行股價預測，此模型結合長短期記憶網絡（LSTM）、自回歸整合移動平均模型（ARIMA）和馬爾可夫鏈蒙特卡羅方法（MCMC）的股價預測模型，通過線性規劃以最小化均方根誤差（MSE）為目標，相比使用單一模型進行預測達到減少 RMSE 的效果。

另外，針對使用者所輸入之投資組合，也建立一個模型，此模型參考了馬可維茲模型（Markowitz Model），找出最小平方法和最大夏普率之投資組合權重供使用者參考。

本專題最終成果以網站形式呈現，功能包括風險評估、投資小學堂、股票基本面資訊、自選股票管理、系統推薦和投資組合管理，為使用者提供全面的投資建議和預測，並提供一個直觀且易於操作的用戶界面。

關鍵詞：LSTM、時間序列、股價預測、投資組合、股票網站

# 目錄

## 摘要

一、前言.....	5
二、研究動機和目的 .....	6
三、文獻回顧 .....	8
3.1 股價預測.....	8
3.2 投資組合.....	8
3.3 研究方法.....	9
3.3.1 長短期記憶模型(LSTM) .....	9
3.3.2 自回歸整合移動平均模型(ARIMA).....	10
3.3.3 馬可夫鏈蒙地卡羅(MCMC) .....	11
3.3.4 現代投資組合理論(MPT).....	12
四、研究方法和結果.....	13
4.1 資料處理 .....	13
4.2 模型設定 .....	13
4.2.1 LSTM.....	13
4.2.2 ARIMA .....	13
4.2.3 MCMC.....	14
4.2.4 線性規劃(LP).....	14
4.3 實驗結果.....	15
五、系統架構和實作.....	19
5.1 系統功能 .....	19
5.2 系統實作.....	23
5.3 資料庫設計.....	24
5.4 系統展示 .....	26
5.4.1 「登入及註冊」及「風險評估」頁面.....	26
5.4.2 「投資小學堂」頁面.....	28
5.4.3 「首頁」頁面.....	29
5.4.4 「股票列表」頁面.....	30
5.4.5 「你的股票」頁面.....	31
5.4.6 「股票預測」頁面.....	32
六、結論.....	33

七、参考文献 .....	34
--------------	----

## 一、前言

在進入股票市場之前，會需要做很多事前的準備，包含：對於股票的基本認知、K線圖所代表的意思、明白自己適合的股票類型、了解該公司的基本運作及營運方式等等。任何的投資都會有一定的風險存在，投資者在追求收益的同時，也需要考慮到自身所要承擔的風險，其中就會需要到處搜尋有需要的資料並記錄下來。

投資者需要蒐集並儲存有關股票的所有資料，接著要針對每一個資料進行整理以及分析，還要去更新部分的資料，包含：股價、圖表等等，透過這些複雜的過程之後，才可以做出較為適合投資者本身的投資方法及手段。

投資者在評估投資方式的時候，往往會需要各項指標及資料來輔助，如何即時更新所需的資料是大部分的投資者都會遇到的問題，而使用我們的系統可以讓用戶不用再四處找尋需要的資訊，也可以透過簡單的小測驗增加對股票的基本認識，同時也能初步判斷用戶偏好於那些股票並推薦給他們。我們的系統可以讓投資者加快速度過初入股市的徬徨期，並給予他們一點選擇上的建議

## 二、 研究動機和目的

在市場上，大多數股票應用程式針對的是專業投資者，但我們意識到對股市尚未有深入了解的新手投資者也需要一個友好且易於使用的平台。因此，我們決定設計一個針對新手投資者的股票應用程式，提供他們所需的基本資料以及額外的教育資源。

除了基本的股票資訊外，我們還設計了一個「投資小學堂」，希望能幫助使用者在進行投資之前建立基本的投資知識和理解。這個小學堂將提供有關股市、投資策略、風險管理等方面的題目，讓新手投資者能夠更加自信地進入股市。

另外，我們理解風險是投資會進行考慮的重要因素。因此，在使用者註冊後，他們將會填寫一份風險問卷，以評估他們對風險的接受程度。這將有助於我們更好地了解使用者的投資目標和風險偏好，從而為他們推薦合適的投資選擇。

在查看股票資料的同時，我們也為使用者設計了一個預測模型。我們搜集了股票的歷年盤面資料，這些資料包含了過去的價格變動以及市場的各種動態因素，幫助使用者更好地掌握市場情報並作出適當的投資決策，不只如此，在使用者設計完屬於自己的投資組合時，我們設計了投資組合模型來找出最佳投資權重，讓使用者落實「雞蛋不放在同一個籃子裡」的概念。

我們的目標是提供一個綜合性的投資平台，不僅讓使用者能夠獲取市場資訊和進行預測，還能夠透過預測和投資組合模型實現更有效的風險管理和資產配置，讓新手投資者建立信心並提高他們的基本知識，使其在股票市場中取得更好的投資表現。

## 三、 文獻回顧

### 3.1 股價預測

股價預測技術可分為時間序列分析和傳統機器學習的方法和基於深度學習的方法。如 ARIMA、MCMC、LSTM，為我們專題所挑選出來的三種模型方法。

S Siarni-Namini[1]研究結果表明，LSTM 模型在金融數據預測方面顯著優於傳統的 ARIMA 模型，具體而言，LSTM 相較於 ARIMA 的平均誤差率減少了 84-87%，深度學習中的訓練次數（即 epoch）對訓練模型的性能沒有影響，這表現出一種隨機行為。

Mochammad Hariadi [2] 研究結果表明通過結合貝葉斯統計、MCMC 和 Apache Spark 平行計算技術，提供了一種有效的股票價格預測方法。

Adebiyi[3] 研究結果表明 ARIMA 模型在短期預測中的表現優異，能夠有效幫助投資者進行決策。實驗結果顯示，對於 Nokia 和 Zenith Bank 的股票數據，ARIMA 模型提供了準確的短期預測。

李培煜[4]研究結果顯示，ARIMA 預測隔一天股價時有較好的表現。循環預測架構下，LSTM 有相對低的 MSE 值，適合預測中短期間內的股價。時間拉長後的結果則不具參考價值。

### 3.2 投資組合

投資組合方法選擇應用 Harry Markowitz[4]的 MPT 理論進行投資組合最佳化，將風險量化為資產收益的方差或標準差，提供了衡量投資風險的標準化方法。強調了分散投資（Diversification）可以降低投資組合的整體風險。指出個別資產的風險可以通過持有多種不同資產來減少，因為資產之間的相關性（Correlation）不同。



### 3.3 研究方法

#### 3.3.1 長短期記憶模型(LSTM)

LSTM 是 S. Hochreiter and J. Schmidhuber[8]所提出，為 RNN 的一種衍伸型態，傳統的 RNN 在訓練過程中會有梯度消失亦或者梯度爆炸的問題，而 LSTM 正可以藉由遺忘門(forget gate)、輸入門(input gate)和輸出門(output gate)的控制來解決梯度消失的問題，但仍無法解決梯度爆炸。

1. 透過 sigmoid function 可以選擇上一層的輸出帶進來的資訊( $h_{t-1}$ )，以及這一層新增加的資訊( $x_t$ )，有多少比例要進入遺忘門捨棄（如圖 3-1）。(3.1)

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.1)$$

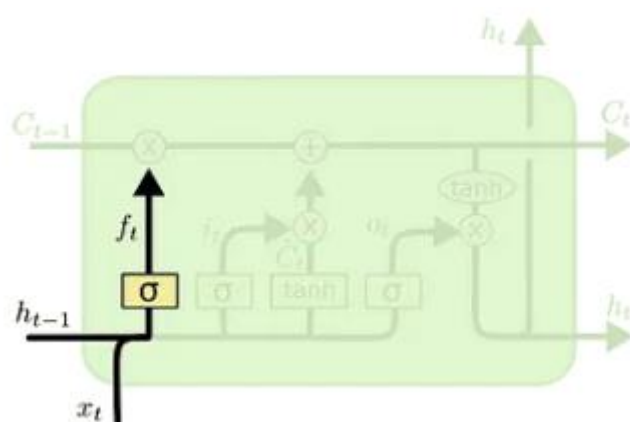


圖 3-1 遺忘門

2. 決定有多少新的資訊要儲存在 cell state，這邊分成兩部分，第一部份是 sigmoid function(3.2)，作用跟第一步一樣，第二部份是 tanh(3.3)，決定有多少資訊要儲存在 cell state（如圖 3-2）。接下來將更新 cell state，也就是把儲存在 cell state 的資訊用來取代被忘記的資訊。

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.2)$$

$$\tilde{C}_t = \sigma(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3.3)$$

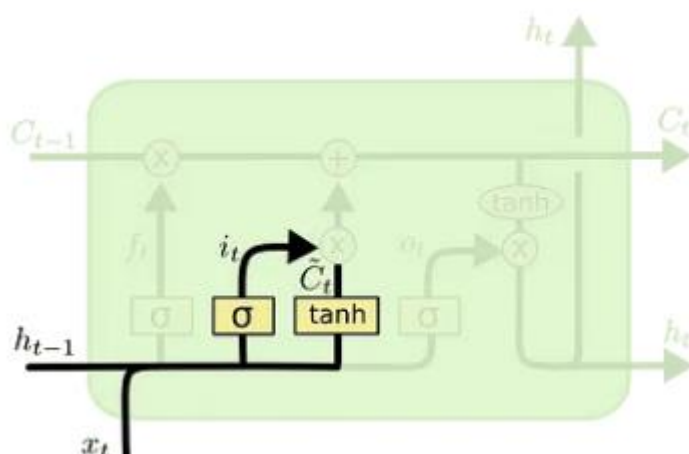


圖 3-2 輸入門

3. 最後，我們決定有多少資訊要被輸出。將  $h_{t-1}$  與  $x_t$  透過 sigmoid function 轉換，然後將上一個步驟產生的結果  $C_t$ ，透過 tanh 的轉換(3.4)，再將兩者相乘(3.5)，來決定主要單元的輸出，以用作下一次預測的資料（如圖 3-3）。

$$h_t = o_t * \tanh(C_t) \quad (3.4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.5)$$

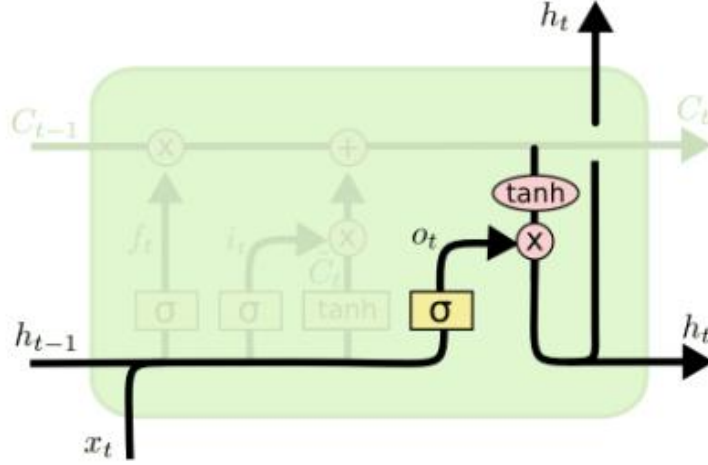


圖 3-3 輸出門

### 3.3.2 自我迴歸整合移動平均模型(ARIMA)

ARIMA 是由 Box Jenkins[5]所提出，自迴歸移動平均模型  $ARMA(p, q)$ ，以下用 ARMA 代稱。結合了 AR 及 MA 的模型， $AR(p)$ 模型表示過去  $p$  天內的觀測值與未來的觀測值之間存在相關性， $MA(q)$  模型是討論  $q$  天內殘差項對固定的平均值  $\mu$  造成擾動的關係式。簡化公式的同時，減少參數的估計量。(3.6)

$$X_t = c + \varepsilon_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (3.6)$$

$C$  is constant,  $p \in \mathbb{Z}$ ,  $\varepsilon_t$  is white noise,  $\phi_i = \arg \min \phi_i (X_t - \hat{X}_t)$

$\mu$  is constant,  $q \in \mathbb{Z}$ ,  $\varepsilon_{t-i}$  is white noise  $i = 1, 2, \dots, q$ ,  $\theta_i = \arg \min \theta_i (X_t - \hat{X}_t)$

$B$  為向後推移算子，並定義  $B_i$  為  $X_t - X_{t-i}$  帶入上式得出(3.7)

$$(1 - \sum_{i=1}^p \phi_i B_i) X_t = C + (1 + \sum_{i=1}^q \theta_i B_i) \varepsilon_t \quad (3.7)$$

$X_t$  與  $\varepsilon_t$  的係數可以視為  $\phi$  和  $\theta$  的函數，可寫成(3.8)

$$\phi(B)X_t = C + \theta(B)\varepsilon_t \quad (3.8)$$

若想獲得非平穩型時間序列模型，可由原始序列經過  $d$  次差分 ( $d > 0$ ) 後轉換成平穩型序列，此平穩型序列可以用 ARMA 模型來表示。如此模型稱為  $(p, d, q)$  階自我迴歸整合移動平均模型 (Autogressive Integrated Moving Average of Order  $(p, d, q)$ ), 簡稱 ARIMA( $p, d, q$ ) 模型，以  $d$  階差分序列將  $X_t$  改為  $(1 - B) dX_t$ ，可得(3.9)。

$$\phi(B)(1 - B)dX_t = C + \theta(B)\varepsilon_t \quad (3.9)$$

Augmented Dickey Fuller Test(ADF)是 Dickey [6]所提出檢定資料是否存在單位根，也是判斷資料是否平穩的依據。ARIMA 建模時，用 ADF 檢定資料以估計參數  $d$ 。資料不平穩則做一階差分，若仍不平穩則持續做差分直到 ADF 通過， $d$  通常為 1。

Akaike Information Criterion (AIC)是由 Akaike[7]提出，此方法是用來衡量統計模型的準度及模型複雜度的指標函數(3.10)，AIC 值與模型參數數量  $K$  呈正相關，與最大似然函數  $L$  呈負相關，AIC 值越低表示模型預測能力越好且有相當的解釋力。

$$AIC = 2k - 2\ln(L) \quad (3.10)$$

### 3.3.3 馬可夫鏈蒙地卡羅法(MCMC)

MCMC 法最早可追溯至 Metropolis, Rosenbluth and Teller[9] 對於統計物理學上的研究，也就是所謂 Metropolis 演算法。MCMC 法的主要概念是經由重覆抽取參數樣本，而建構出一長串的馬可夫鏈，以求得參數的近似分配，MCMC 法由兩部分的步驟所組成，一個為蒙地卡羅積分(Monte Carlo integration)，另一個為馬可夫鏈過程(Markov Chain process)。

原始的蒙地卡羅方法是用來計算複雜的積分。所謂的蒙地卡羅積分方法(Monte Carlo integration)是先從後驗機率分配中抽取樣本，再利用樣本平均數去逼近參數  $\theta$  的後驗機率分配期望值。(3.11)

$$E[f(\theta)|x] = \frac{\int f(\theta)P(x|\theta)d\theta}{\int P(\theta)P(x|\theta)d\theta} \quad (3.11)$$

假設  $x_0, x_1, x_2 \dots$  為一序列變數，且  $x_{t+1}$  來自  $P(x_{t+1} | x_t)$  的機率分配，換句話說，每一個變數僅與前一個變數有關，這樣的序列變數稱為馬可夫鍊(Markov Chain)。

Hastings [10]修正了 Metropolis 演算法中對樣本的產生步驟與認定準則，完成整個運算架構，即所謂的 Metropolis-Hastings 演算法。

M-H Algorithm 的基本步驟如下：

Initialize: Set  $x_0$   
For  $j = 1$  to  $N$  do:  
Generate  $y$  from  $q(x)$   
Generate  $u$  from  $U(0,1)$   
If  $u < \alpha(x' | x_t)$  then:  
Set  $x_{t+1} = x'$   
Else:  
Set  $x_{t+1} = x_t$   
Return the values  $\{x_1, \dots, x_n\}$

1. 初始化：選擇一個任意的點  $x_t$  作為樣本中的第一個觀察值，並選擇一個提議函數  $q(x | y)$ 。在本節中，假設  $q$  是對稱的；換句話說，它必須滿足  $q(x | y) = q(y | x)$ 。步驟
2. 建議分配 (proposal density)：從條件概率分布  $q(x_{t+1}, x_t)$  中抽樣一個候選狀態  $x'$ 。
3. 接受或拒絕步驟：計算接受概率(acceptance probability)(3.12)

$$\alpha(x' | x_t) = \min(1, \frac{p(x_t)q(x' | x_t)}{p(x')q(x_t | x')}) \quad (3.12)$$

其中  $p(x)$  是目標分布。然後，從均勻分布  $U(0,1)$  中抽樣一個隨機數  $u$ 。如果  $u < \alpha(x' | x_t)$ ，則接受候選狀態，即設定  $x_{t+1} = x'$ ；否則，拒絕候選狀態，即設定  $x_{t+1} = x_t$ 。

### 3.4.4 現代投資組合理論(MPT)

貝他值 (Beta Value) 是一項衡量產品或投資組合之系統性風險的指標，定義如(3.13)所示，用於評估該金融產品或投資組合與總體市場波動性的比較。若貝他係數等於 1，其波動性與總體市場波動性一致，若該係數小於 1，表示波動性較總體市場小，而若該係數大於 1，則其波動性較總體市場大。

$$B_a = \frac{\text{Cov}(r_m, r_a)}{\sigma_m^2} \quad (3.13)$$

其中  $\text{Cov}(r_m, r_a)$  是證券  $a$  的收益與市場收益的共變異數； $\sigma_m^2$  是市場收益的變數。

最大夏普值(3.14)：

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_P} \quad (3.14)$$

其中  $R_p$  是投資組合的平均報酬率， $R_f$  是無風險利率（如國庫券的報酬率）， $\sigma_P$  是投資組合回報率的標準差（即投資的風險）。

## 四、研究方法和結果

### 4.1 資料處理

本模型使用 Yahoo Finance[17]中所有台灣上市公司之股票自 2020/01/01 以來至今之每日股票交易資料中的收盤價格，並將資料前 80% 作為訓練集，後 20% 作為測試集。針對資料的處理，在 LSTM 中，將資料標準化處理，將每個特徵的數據縮放到 0 和 1 之間，以防止數據中的特徵值範圍過大影響模型的訓練，進而加快收斂速度，提高模型效能，而對於 ARIMA 和 MCMC 來說，標準化資料並非必要。ARIMA 模型要求數據需為平穩狀態，若數據非平穩應使用差分處理，而對於 MCMC 模型，數據的範圍控制對於機率分布的影響並不大，因此不特別對此二模型資料進行處理。

### 4.2 模型設定

使用三個模型、LSTM、ARIMA、MCMC，最後建立優化問題來混合三個模型的預測，目標函數是最小化預測與過去四個月歷史數據之間的均方誤差 (MSE)。

#### 4.2.1 LSTM

本模型使用多組輸入一組輸出(many-to-one)的架構，並建立兩個 LSTM 層，為防止過度擬合，皆在該層之後分別應用了 35% 和 30% 的 dropout 率；輸出層根據 LSTM 層的結果產生一個預測，模型的編譯過程中，本模型使用 ReLU 激勵函數、Adam 優化器和均方誤差(mean squared error)損失函數，同時使用準確性(accuracy)作為評估指標。

#### 4.2.2 ARIMA

在 ARIMA 當中，auto\_arima 的參數設定優化的是 lbfgs，全稱是 Limited-memory Broyden-Fletcher-Goldfarb-Shanno 算法。它是一種擬牛頓法(quasi-Newton method)，用於在優化問題中尋找局部最優解。

時間序列的前移驗證 (Walk Forward Validation) 是一種交叉驗證技術，特別適用於時間序列數據的建模和預測。這種驗證方法將數據集分成訓練集和測試集，並使用滾動的方式進行模型訓練和預測。

以下是時間序列的前移驗證的步驟：

1. 初始化訓練集和測試集：從時間序列數據的開頭選擇一個固定大小的窗口作為初始訓練集，再選擇一個窗口作為初始測試集。
2. 訓練模型：使用初始訓練集來訓練模型。
3. 預測下一個時間步：使用訓練好的模型來預測下一個時間步的值。

4. 更新訓練集和測試集：將窗口向前移動一步，將先前的測試集的時間步加入到訓練集中，同時將下一個時間步添加到測試集中。
5. 重複步驟 2-4：重複上述步驟，直到測試集中沒有更多的時間步。
6. 計算模型的性能指標：使用所有測試集上的預測結果來計算模型的性能指標，例如均方根誤差(RMSE)。

### 4.2.3 MCMC

在 MCMC 當中，每一個時間步驟中根據幾何布朗運動模型計算股票價格的更新，並使用馬爾可夫鏈蒙特卡洛方法生成隨機數來模擬股票價格的隨機波動。

標準正態分佈的概率密度函數(4.1)：

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-u)^2}{2\sigma^2}} \quad (4.1)$$

Black-Scholes 公式中的股票價格變化(4.2)：

$$S_t = S_0 \cdot e^{\left(\left(r - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t} \cdot Z_t\right)} \quad (4.2)$$

$S_t$  是時間  $t$  的股票價格

$S_0$  是初始股票價格

$r$  是無風險利率

$\sigma$  是波動率

$\Delta t$  是每個時間步的時間間隔

$Z_t$  是標準正態分佈的隨機變數，由 MCMC 方法生成

### 4.2.4 線性規劃(LP)

最後將線性規劃用於在三個不同的時間序列預測模型（LSTM、ARIMA、MCMC）之間分配權重，以最小化均方誤差（MSE），並將此模型稱為 LP。

1. 定義線性規劃變數：定義了三個變數，分別代表 LSTM 模型、ARIMA 模型和 MCMC 方法的權重。這些權重變數的值將在求解過程中被優化。
2. 定義目標函數：將均方誤差的表達式添加到模型中作為目標函數。目標是最小化均方誤差，以使模型的預測與實際數據之間的差異最小化。
3. 添加約束條件：添加約束條件，確保權重變數的取值合理。其中包括權重之和為 1，以及權重的範圍限制(4.3)。

$$\begin{cases} W_{ARIMA} * S_{AIMA} + W_{LSTM} * S_{LSTM} + W_{MCMC} * S_{MCMC} = \hat{S} \\ W_{ARIMA} + W_{LSTM} + W_{MCMC} = 1 \\ MSE(S, \hat{S}) \rightarrow \min \end{cases} \quad (4.3)$$

其中， $S$ 代表實際預測期間的數據； $\hat{S}$ 代表該期間的預測數據； $W$ 代表每個模型預測的權重。這個模型的輸出是加權的預測數據。

### 4.3 實驗結果

我們分別對統一「1216.TW」、中鋼「2002.TW」、台積電「2330.TW」和鴻海「2317.TW」進行一個月的回測和四個月的回測來呈現結果（圖 5-1 至圖 5-8），並以表格比較其 RMSE 結果（表 5-1、表 5-2）。

LSTM 和線性規劃(LP)模型在一個月和四個月內的預測表現良好，ARIMA 模型在一個月和四個月的預測中表現一般，RMSE 值相對較高，MCMC 模型只有在波動性低的股價表現良好，其餘在四個月的預測中情況下表現不佳。

可以發現在四個月的預測底下，透過線性規劃調整權重後的模型預測結果和單用一個模型的 RMSE 平均減少 36%，在一個月的預測，則因為鴻海的近期股價波動劇烈，最終反而使平均 RMSE 上升 38%。

表 4-1 一個月的個別模型預測結果的 RMSE

一個月	1216.TW	2002.TW	2330.TW	2317.TW
ARIMA	1.288	4.763	212.348	38.837
LSTM	0.871	0.515	25.850	5.700
MCMC	0.461	0.458	101.632	22.331
線性規劃(LP)	0.415	0.295	19.33	18.47
Reduce RMSE(%)	10%	36%	25%	-224%
Average RMSE	-38			

表 4-2 四個月的個別模型預測結果的 RMSE

四個月	1216.TW	2002.TW	2330.TW	2317.TW
ARIMA	0.801	0.255	30.795	3.4307
LSTM	1.02	0.336	14.795	4.0764
MCMC	5.248	0.811	248.424	59.0331
線性規劃(LP)	0.196	0.241	11.324	2.990
Reduce RMSE(%)	81	28	23	13
Average RMSE(%)	36			

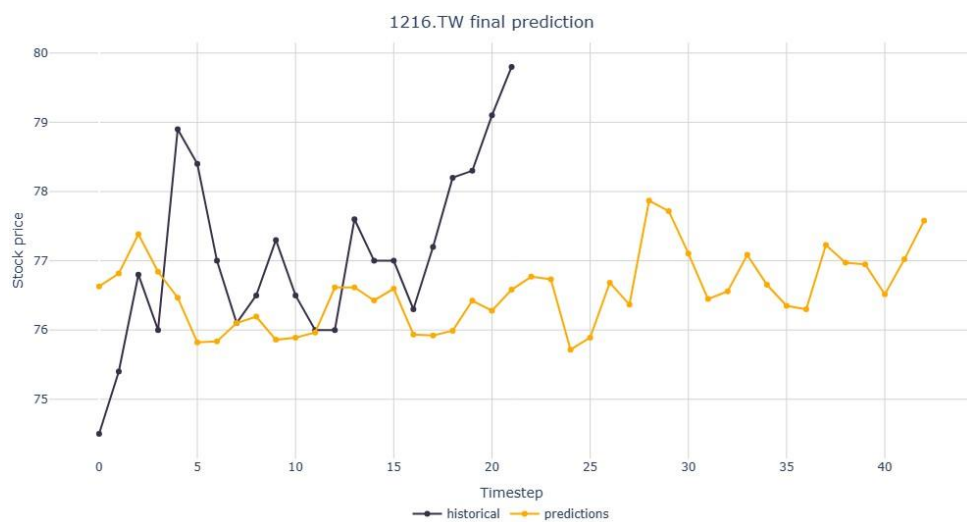


圖 4-1 1216.TW 一個月回測

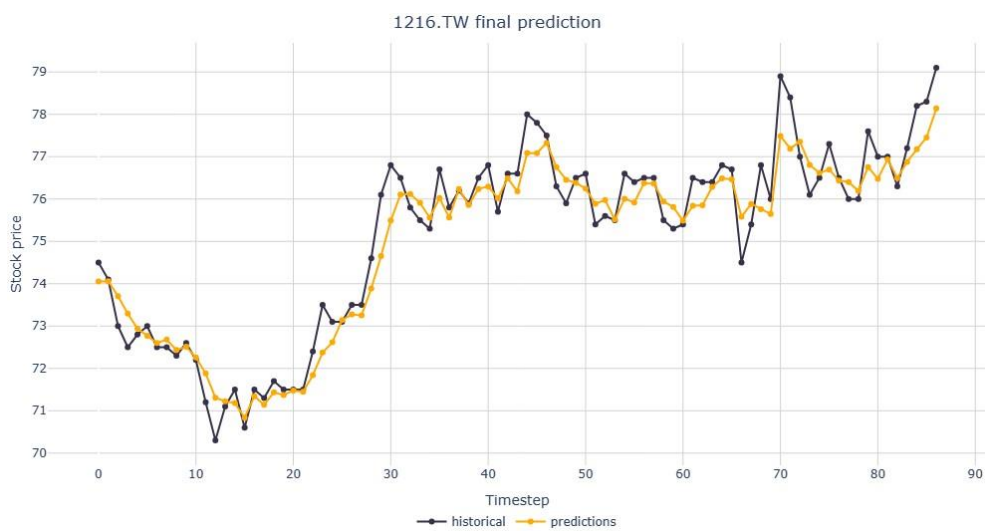


圖 4-2 1216.TW 四個月回測

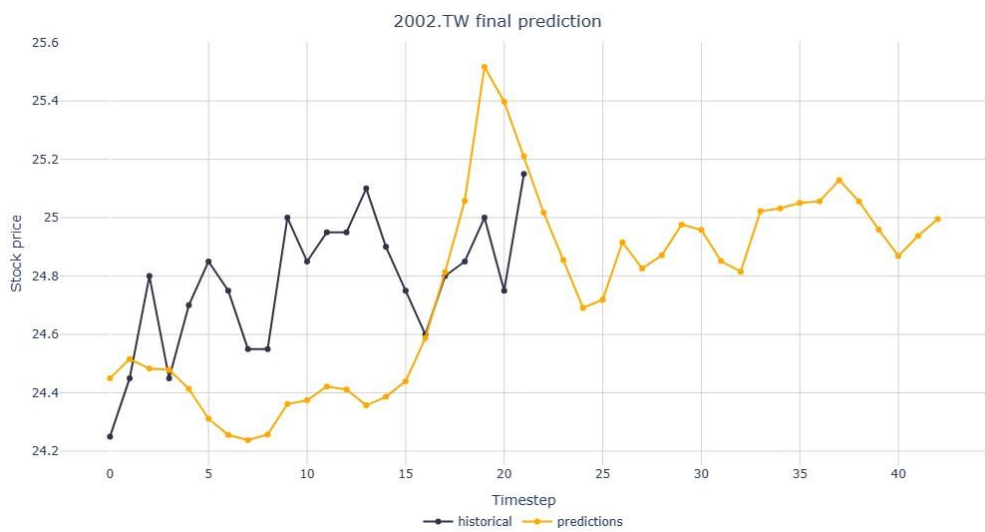


圖 4-3 2002.TW 一個月回測



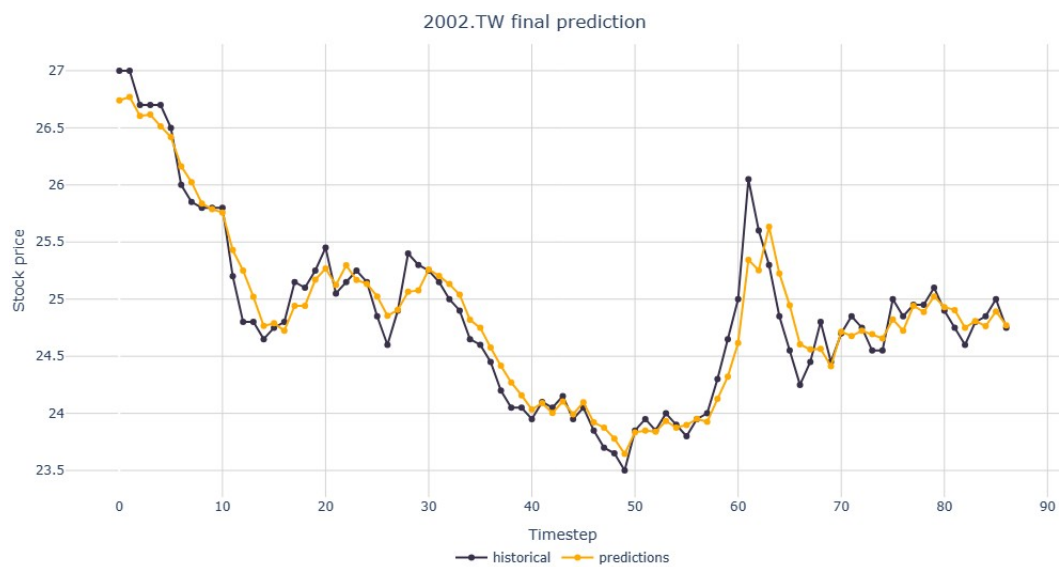


圖 4-4 2002.TW 四個月回測

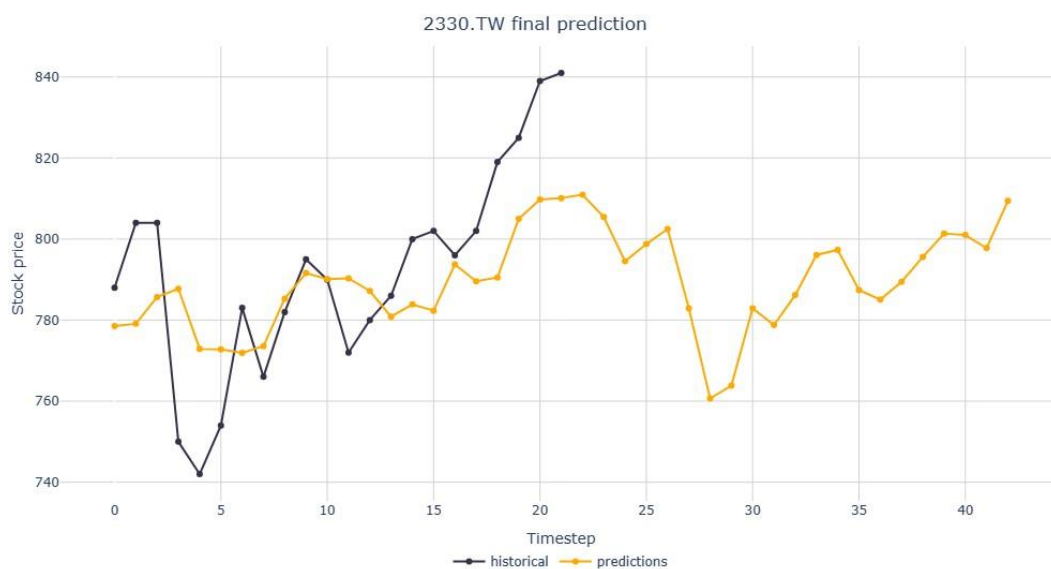


圖 4-5 2330.TW 一個月回測

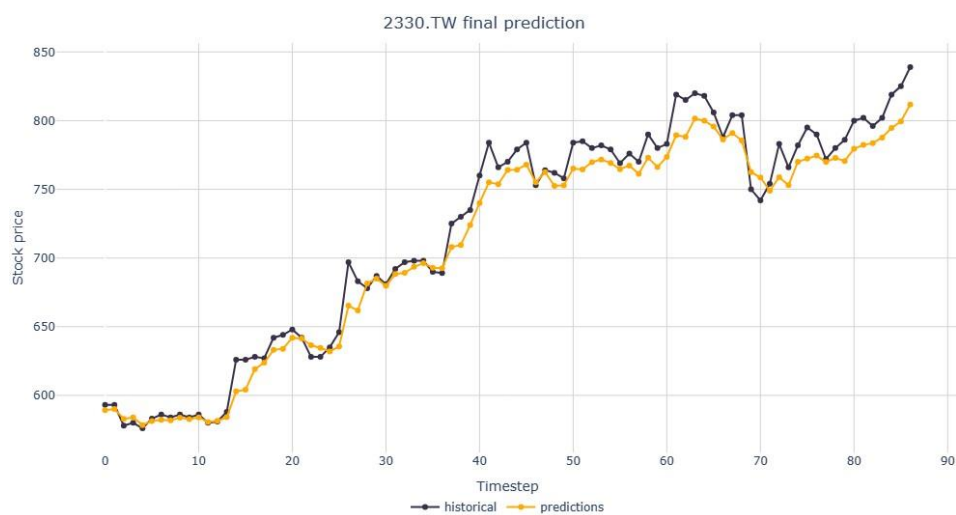


圖 4-6 2330.TW 四個月回測

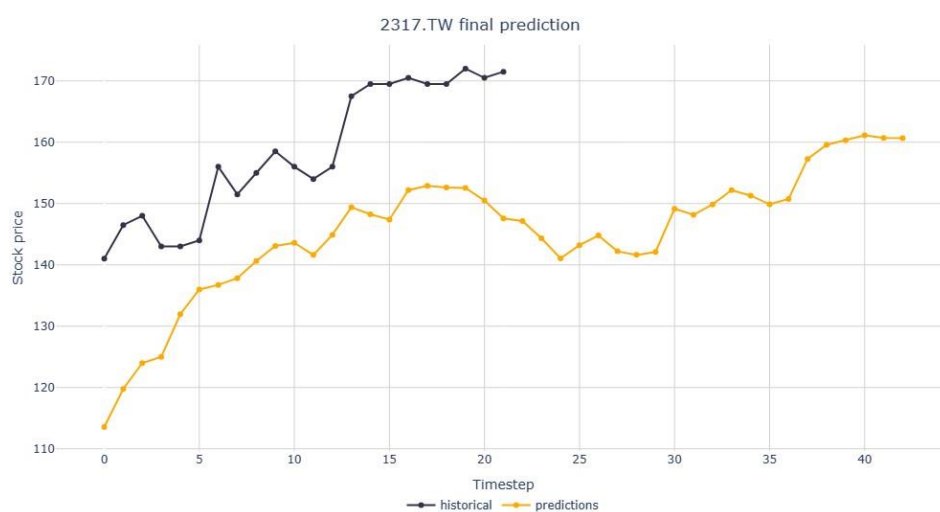


圖 4-7 2317.TW 一個月回測

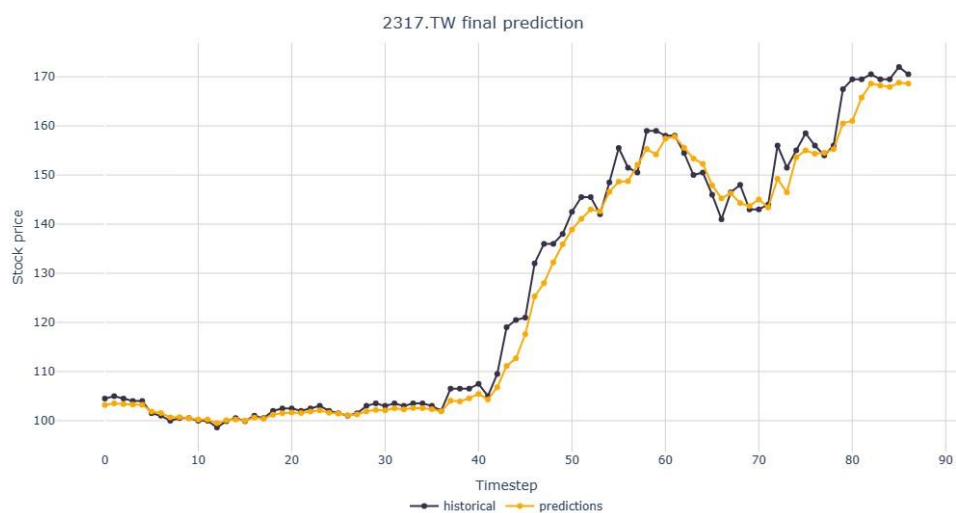


圖 4-8 2317.TW 四個月回測

## 五、系統架構與實作

### 5.1 系統功能

本系統架構（如圖 5-1）可分為五種主要功能：「風險評估」、「投資小學堂」、「股票列表」、「你的股票」和「股票預測」（圖 5-2 中可總覽使用者可使用之功能），使用者經過註冊後會進行風險評估測驗，系統依照成績判斷風險等級，並依照貝他值推薦的股票，可於「風險評估」頁面查詢；「投資小學堂」內設有多種題目可讓使用者練習並在作答完畢後查看解答以增進投資知識。「股票列表」中，使用者可查看所有股票，並利用搜尋功能直接找尋特定股票，另外，使用者也可以點選股票以加入個人的投資組合。加入後的股票可於「你的股票」中查看，此頁面提供了投資組合模型，計算出投資組合中每支股票的最佳權重供使用者參考。最後，若使用者想查看投資組合中各支股票未來之預測，可點選「股票預測」以獲得模型預測結果。

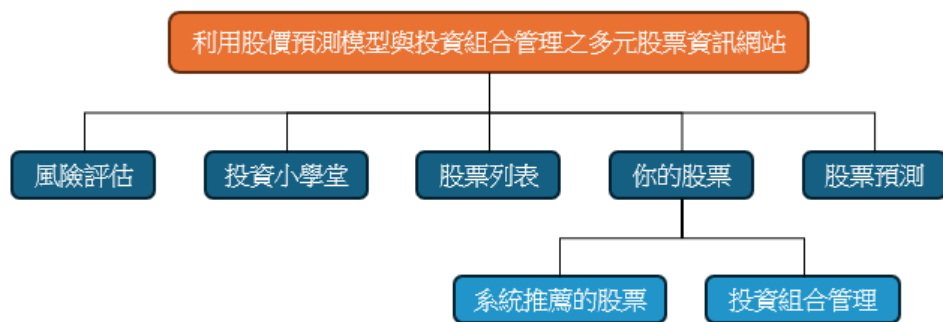


圖 5-1 系統分支圖

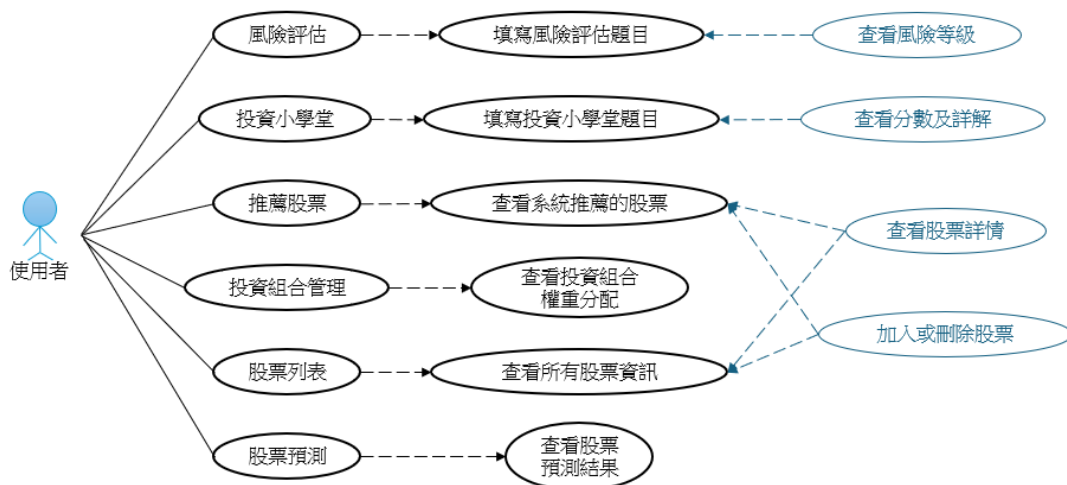


圖 5-2 系統流程圖

接下來依序介紹「登入及註冊」頁面、「投資小學堂」頁面、「你的股票」頁面、「股票預測」頁面。

「登入及註冊」頁面（操作流程如圖 5-3）為使用者進到系統後第一個會看到的頁面，如果該使用者註冊過了，系統會從 user 資料庫中判斷帳號是否存在，如果存在再判斷密碼是否正確，否則就會請使用者重新輸入或註冊；註冊會先填寫想要的使用者帳號及密碼後，會填寫相關基本資料，最後會請使用者填寫風險評估測驗，以讓系統判斷使用者的風險等級並且推薦給使用者適合的股票；本系統設置了一個門檻：使用者需要完成「股票基本」全部填寫完畢後會再請使用者重新登入此系統以便資料管理。

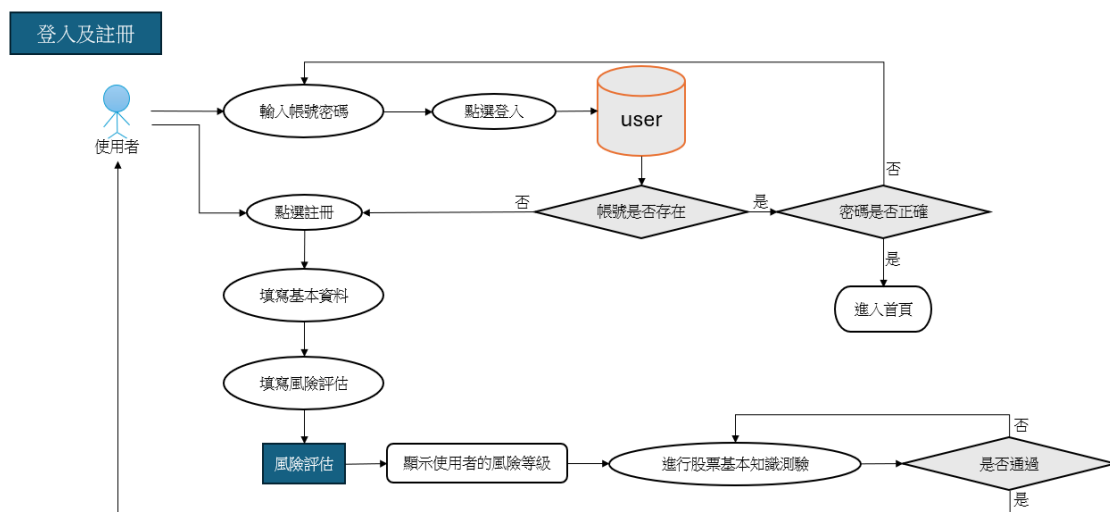


圖 5-3 「登入及註冊」頁流程圖

「風險評估」頁面（操作流程如圖 5-4 所示），風險評估的題目主要是依照可以了解使用者想要得到的報酬即能接受的損失來設計題目，在使用者填寫完題目後，系統會判斷其風險等級，並顯示給使用者看，而後系統會依照 beta 值篩選適合使用者的股票，並存入資料庫中。

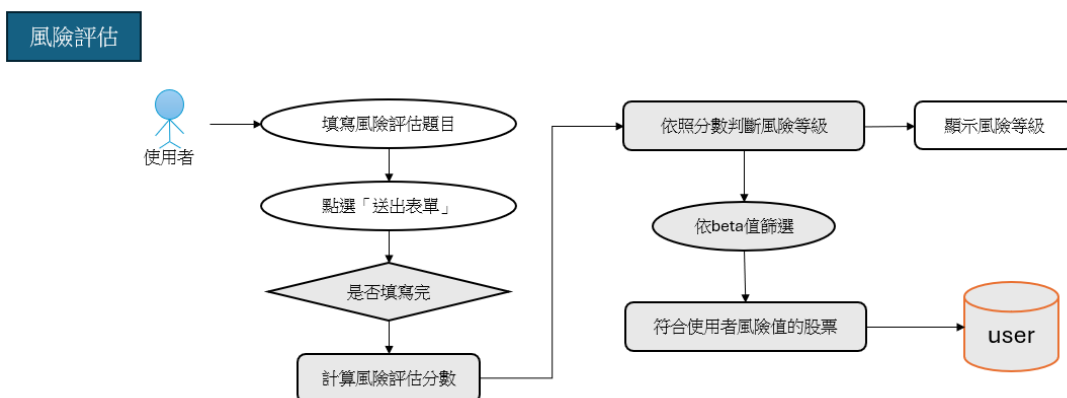


圖 5-4 「風險評估」與推薦股票流程圖

在「投資小學堂」頁面（操作流程如圖 5-5），我們會提供使用者一些題目，題目內容主要為一些股票的基本知識，目的是讓使用者可以更熟悉股票的相關資訊，或者新手也可以藉此更了解股票。

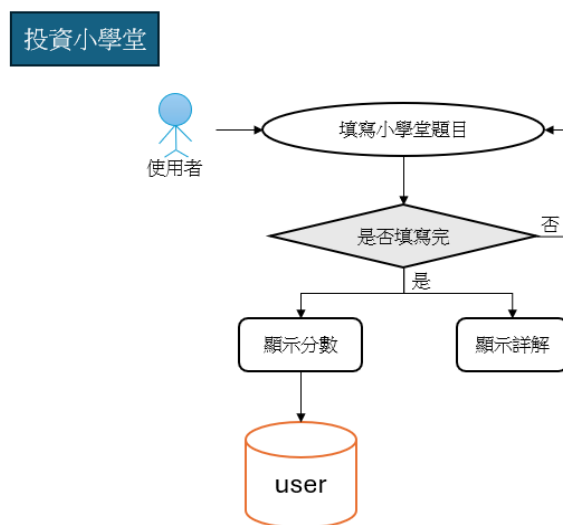


圖 5-5 「投資小學堂」流程圖

在「你的股票」頁面（操作流程如圖 5-6）中主要會分成三個部分：

1. 「推薦給您的股票」（操作流程如圖 5-4）：此部分會顯示我們依照風險評估後推薦給使用者的股票，因為風險評估後會將推薦的股票代碼存到 user 資料庫，因此會先從 user 資料庫中抓取股票代碼後再從 stock 資料庫中抓取我們想呈顯給使用者看的資訊，會設計一個按鈕讓使用者可以點選加入或刪除股票，在此加入的股票會先存在 user 資料庫中再顯示在「您已加入的股票中」。
2. 「您已加入的股票」：此部分會顯示使用者自行點選加入的股票，在此可以刪除已加入的股票，在此操作的結果也會存入 user 資料庫中。
3. 「股票投資組合」（操作流程如圖 5-7）：使用者點選「顯示股票投資組合」後，系統會從 user 資料庫中抓取使用者加入的股票並且將它們給予「投資組合計算模型中計算」，最後會將具有最小波動率的風險投資組合和具有最大夏普率的風險投資組合提供給使用者參考。

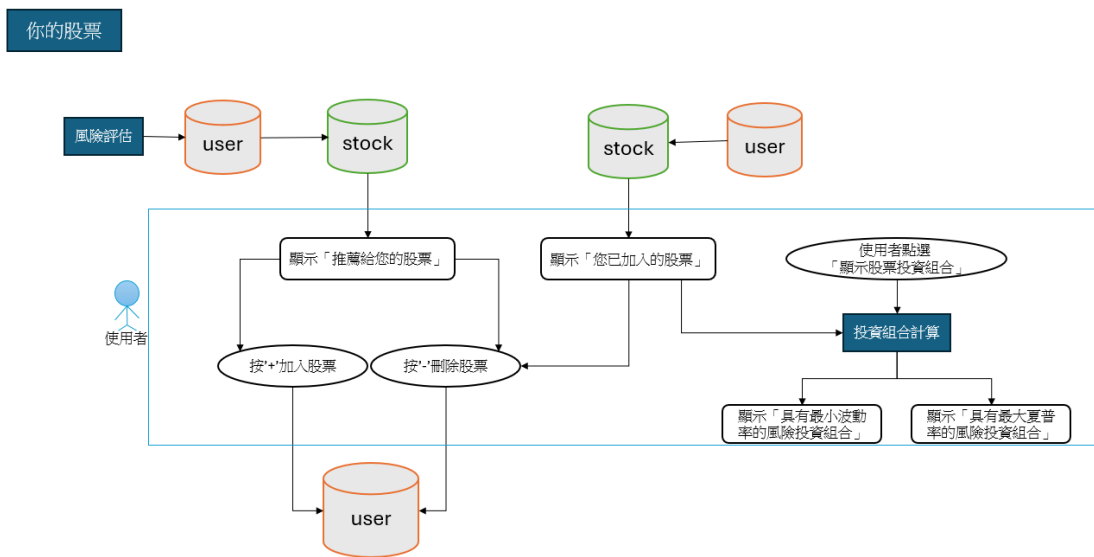


圖 5-6 「你的股票」流程圖

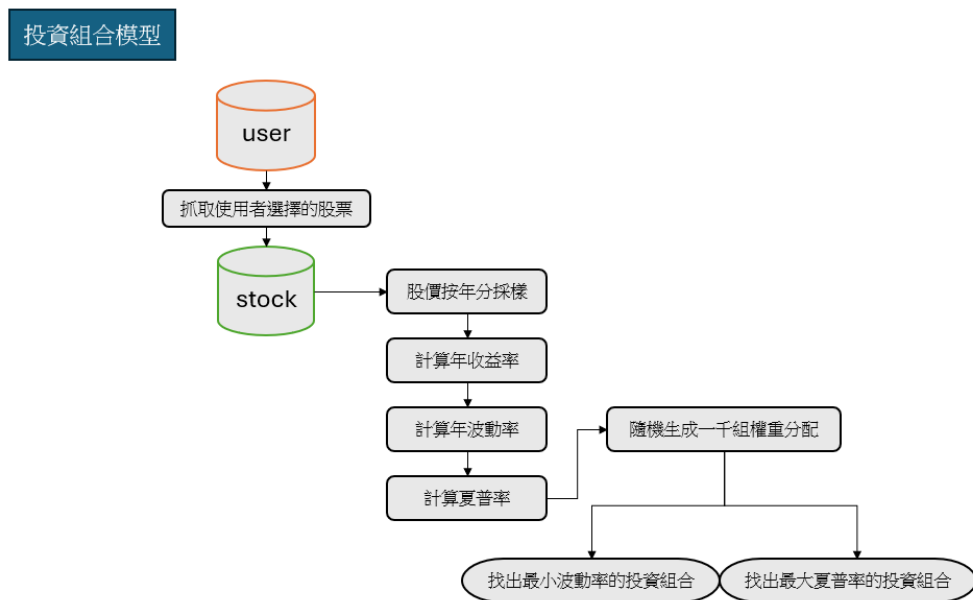


圖 5-7 投資組合流程圖

在「股票預測」頁面（操作流程如圖 5-8）會顯示使用者已加入的股票，在每支股票的最後會放上一個箭頭，讓使用者可以點選，點選後我們會從 stock 資料庫中抓取需要用到的資料丟到預測模型中，最後將結果存到 predict 資料庫，再畫出圖表提供使用者參考。

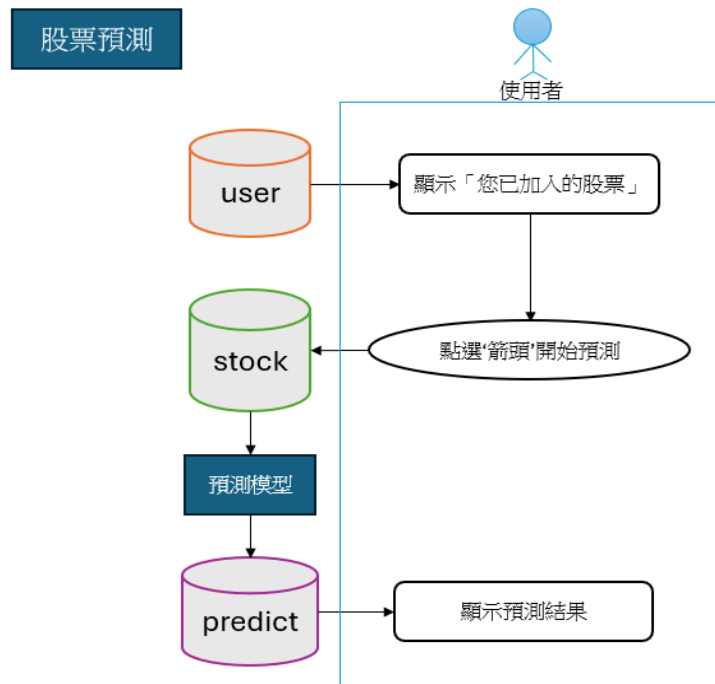


圖 5-8 「股票預測」流程圖

## 5.2 系統實作

系統實作主要分成四大部分（如圖 5-9 所示）。

1. 用戶端：本系統使用 HTML、CSS 及 JavaScript，供使用者在瀏覽器上執行與操作。
2. 伺服器端：使用 php 的 Apache 伺服器來連接資料庫及操作 python 程式後，進行資料處理再連結到用戶端。
3. MySQL 資料庫：主要存放用戶資料及系統會使用到的股票相關資訊，資料庫的設計於章節 5.3 詳細說明。
4. python 推薦與預測模型：利用 python 執行推薦與預測的模型，並將執行結果傳給伺服器或資料庫。

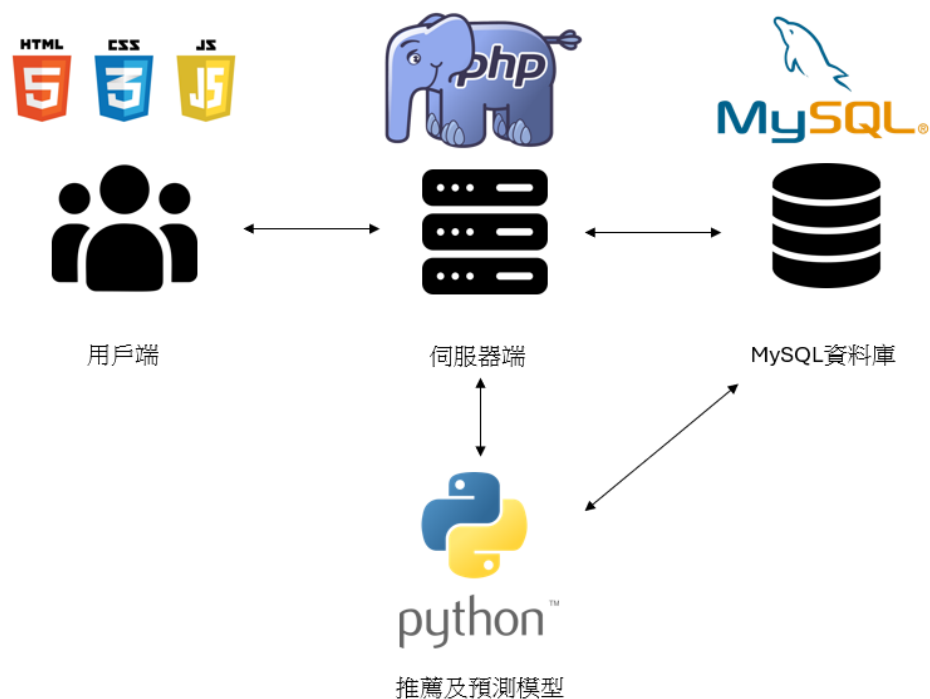


圖 5-9 系統實作概念圖

### 5.3 資料庫設計

資料庫的設計（如圖 5-10 所示），主要有四個表格來儲存系統會使用到的資料：

1. “member”：儲存已註冊使用者資料。
  - (1)Username：使用者的帳號
  - (2>Password：使用者的密碼
  - (3)Sex：使用者的性別
  - (4)age：使用者的年齡
  - (5)risk\_level：使用者的風險等級
  - (6)s\_score：使用者的小學堂成績
  - (7) recommend\_stock\_id []：推薦給使用者的股票代碼(有多支股票)
  - (8) your\_stock\_id[]：使用者自行加入的股票(有多支股票)
2. “stock”：儲存股票資料。
  - (1) Stock\_id：股票代碼
  - (2) Date：資料日期
  - (3) Adj\_Close：股價
  - (4) High：當日股價最高點
  - (5) Low：當日股價最低點
  - (6) Open：開盤價
  - (7) Close：收盤價
  - (8) Volume：成交量



- (9) Spread：價差
- (10) Spread\_Month：月平均價差
- (11) Growth\_Rate：漲跌幅(%)
- (12) Growth\_Rate\_Month：月平均漲跌幅(%)
- (13) Pe\_ratio：本益比(%)
- 3. “predict”：儲存股價預測的結果。
  - (1) Stock\_id：股票代碼
  - (2) Predict：預測股價
  - (3) History：歷史股價
- 4. “company”：儲存股票的公司資訊。
  - (1) Stock\_id：股票代碼
  - (2) company\_name：公司名稱
  - (3) company\_website：公司網站
  - (3) company\_description：公司簡介

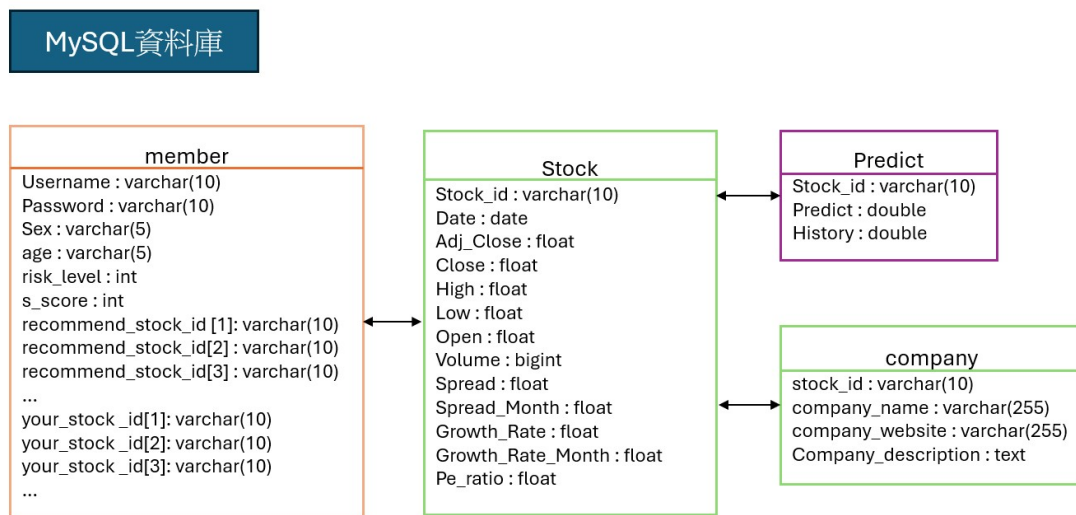


圖 5-10 資料庫設計圖

## 5.4 系統展示

### 5.4.1 「登入及註冊」及「風險評估」頁面

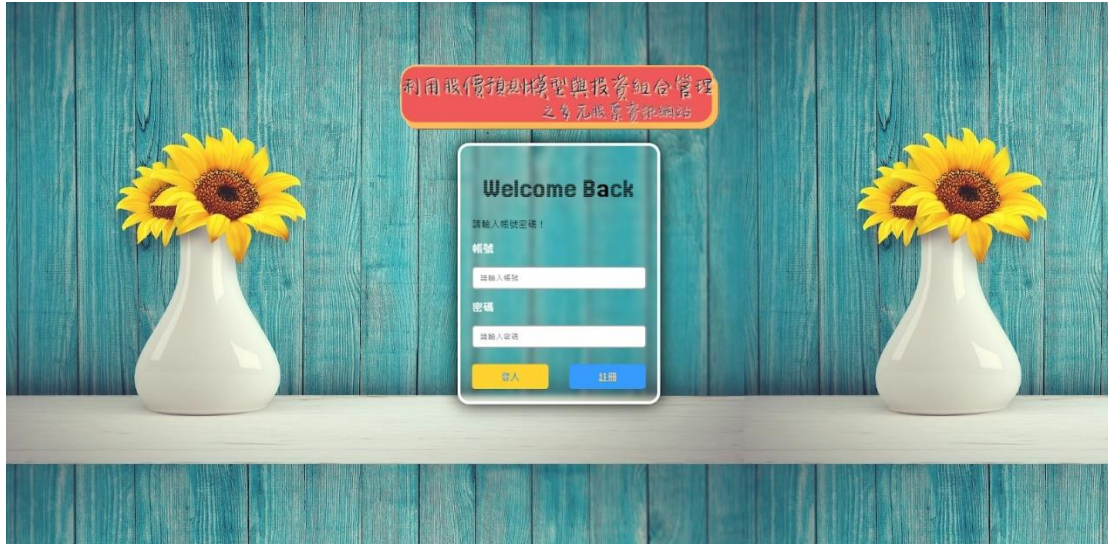


圖 5-11 「登入」頁面

於「登入」頁面（如圖 5-11）輸入帳號密碼，會由伺服器抓取資料庫中已存在的資料來判斷：帳號是否存在以及密碼是否正確，皆正確使用者即會進入到「首頁」，帳號不存在則會請使用者進行「註冊」。

請填寫下列資訊以完成註冊

請輸入使用者名稱

十個字以內的英文及數字組合

請設定密碼

請輸入密碼

請問您的性別？

☐ 男  
☐ 女  
☐ 不方便透露

請問您的年齡？

☐ 7歲以下 ☒ 7~17歲  
☐ 18~29歲 ☐ 30~39歲  
☐ 40~49歲 ☐ 50~59歲  
☐ 60~64歲 ☐ 65歲以上

請問您的教育程度為何？

☐ 碩士以上  
☐ 大學/專科  
☐ 高中職  
☐ 國中(含)以下

[返回登入畫面](#) [送出資料並進入風險評估](#)

圖 5-12 「註冊」頁面

「註冊」頁面（如圖 5-12）會請使用者設定帳號及密碼，以及填寫一些基本資訊，送出資料後伺服器即會將資料存入資料庫中，而後使用者會進入「風險評估偏好測驗」頁面。

**Q6：請問哪一項描述比較符合您的投資理財專業度？**

- ☐ 非常熟悉；對金融商品，包括複雜投資商品(如：結構性商品)有豐富的知識和理解
- ☒ 有些熟悉；對大部分的金融商品，包括複雜投資商品(如：結構性商品)有相當知識和理解
- ☐ 具基本知識；對一般的金融商品(如：股票、債券、基金)有基本的知識和理解
- ☐ 不太熟悉；對金融商品及投資理財的瞭解有限

**Q7：請問哪一項描述比較符合您的投資目的？**

- ☒ 仍以「保障」為主，可接受「些微」本金損失
- ☐ 追求「小幅」收益，可接受「低度」本金損失
- ☐ 追求「較高」收益，可接受「中度」本金損失
- ☐ 追求「大幅」收益，可接受「高度」本金損失

圖 5-13 「風險評估」題目範例

## 您的風險等級為：等級 4



圖 5-14 「風險評估」等級顯示畫面

此頁面會讓使用者填答一些相關問題（如圖 5-13 所示），其中包含是否有投資經驗、期望的報酬率以及能接受多少的損失等等。填答完畢後會將答案

回傳給伺服器端，計算出分數後，系統會顯示使用者的風險等級供其查看（如圖 5-14 所示），做完此測驗後會進入投資小學堂進行測驗。

#### 5.4.2 「投資小學堂」頁面



**Q1：台灣股票市場交易時間為何？(單選題)**

- ☐ 週一至週五9:00~13:00
- ☐ 週一至週六9:00~13:00
- ☐ 週一至週五9:00~15:00
- ☐ 週一至週五10:00~14:00

**Q2：關於股票買賣單位，下列敘述何者錯誤？(單選題)**

- ☐ 1張股票是由1000股所組成
- ☐ 購買股票都是以一張為單位計算
- ☐ 購買零股只能採用電子下單
- ☐ 以上皆正確

**Q3：關於購買股票的手續費及交易稅，下列敘述何者錯誤？(單選題)**

- ☐ 股票交易手續費為購買金額的0.1425%
- ☐ 除了手續費在買入股票時還會收取0.3%的交易稅
- ☐ 股票買進及賣出時都會收取手續費
- ☐ 以上皆正確

圖 5-15 「投資小學堂」題目範例

在「投資小學堂」中，本系統提供使用者 10 題基本的股票知識測驗（如圖 5-15），送出表單後，系統會顯示使用者的作答分數以及每題的詳解供使用者參考，而本系統會設置一個門檻，使用者需要答對七題以上，才可以登入並使用本系統，否則就需要重新進行測驗。

### 5.4.3 「首頁」頁面

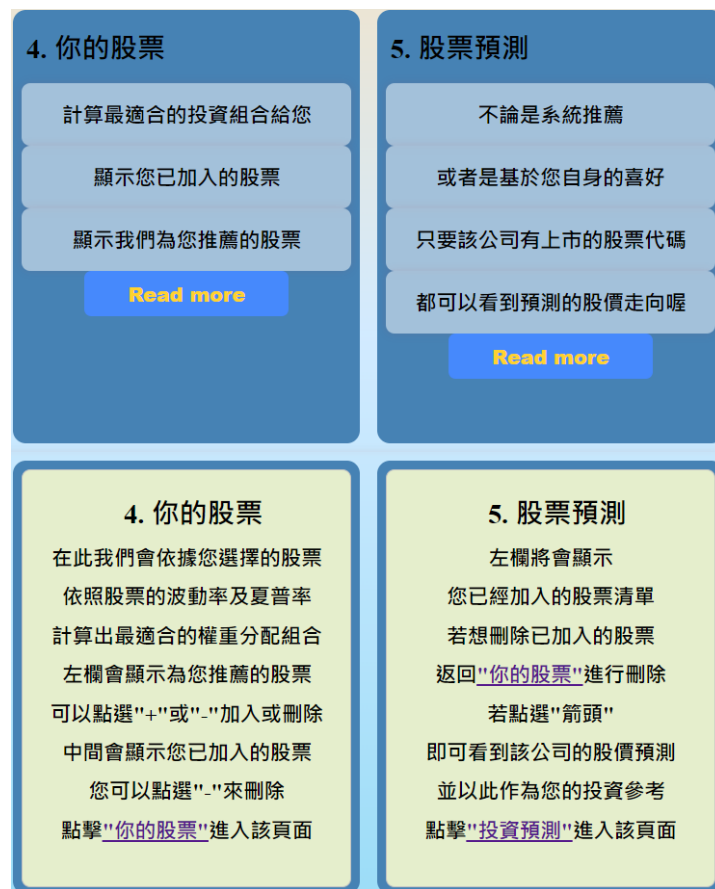


圖 5-16 「首頁」部分畫面

在「首頁」中，主要是系統全部功能的說明（如圖 5-16），讓使用者可以初步的了解此系統。

## 5.4.4 「股票列表」頁面

股票列表						
搜尋股票:		<input type="text"/>	<input type="button" value="Search"/>			
利用產業別篩選	顯示原始排序	股價排序(高->低)	股價排序(低->高)	按照交易量排序(高->低)	按照漲跌幅排序	
	股票代碼	公司名稱	產業別	股價	漲跌幅	收盤股價
+	<a href="#">1101.tw</a>	臺灣水泥	水泥工業	33.9	0.89 %	33.9
+	<a href="#">1102.tw</a>	亞洲水泥	水泥工業	44.6	-2.51 %	44.6
+	<a href="#">1103.tw</a>	嘉新水泥	水泥工業	18.45	-0.27 %	18.45
+	<a href="#">1104.tw</a>	環球水泥	水泥工業	33.3	0.6 %	33.3
+	<a href="#">1108.tw</a>	幸福水泥	水泥工業	17.05	0 %	17.05
+	<a href="#">1109.tw</a>	信大水泥	水泥工業	20	0 %	20
+	<a href="#">1110.tw</a>	東南水泥	水泥工業	20.15	-0.74 %	20.15

圖 5-17 「股票列表」部分畫面

在此頁中，伺服器會向資料庫索取所有股票資料整理後，顯示系統中所有股票的基本資訊（如圖 5-17），使用者可以依照需求點選不同的按鈕來改變排序，方便使用者找到想要的股票，在此也可以點選加號將股票添加到「您已加入的股票」當中，頁可以利用搜尋功能直接查看單支股票的詳細資訊，如公司簡介、公司網址、股票資訊及 K 線圖（如圖 5-18）。



圖 5-18 「2330.TW」詳細資訊頁面

在此使用者可以看到該支股票的公司簡介，以及在此會顯示該股票近兩百天的 K 線圖供使用者參考。使用者也可以利用公司名稱旁的加號將股票添加到「您已加入的股票」當中。



### 5.4.5 「你的股票」頁面

使用者在完成「風險評估」後，系統會推薦符合使用者風險等級的股票，伺服器會從資料庫中提取相關資訊，顯示在網頁上提供給使用者選擇（如圖 5-19 所示）；使用者點選加入後會出現在「您已加入的股票」中（如圖 5-20 所示）；在「股票投資組合」中當使用者點選顯示後，系統需要 30 秒的時間做計算後，會顯示系統推薦給使用者的投資組合權重分配（如圖 5-21 所示）。

推薦給您的股票					
	股票代碼	公司名稱	股價	成交量	本益比
+	<a href="#">6776.tw</a>	展碁國際	63	60092	12.986
+	<a href="#">1789.tw</a>	台灣神隆	28.6	2221281	57.551
+	<a href="#">3024.tw</a>	億豐電子	20.85	23481375	51.4103
+	<a href="#">6573.tw</a>	虹揚發展科技	12.85	37000	-4.41581
+	<a href="#">6285.tw</a>	啟暎科技	155	3278553	17.7907
+	<a href="#">2887.tw</a>	台新金融控股	18.9	15514214	18.5149
+	<a href="#">3607.tw</a>	谷崧精密工業	17.3	601100	-19.9419
+	<a href="#">2913.tw</a>	台灣農林	21.55	2725141	-60.2778
+	<a href="#">6671.tw</a>	三能集團控股	45.1	42000	15.1742
+	<a href="#">6165.tw</a>	浪凡網路科技	39.35	420000	157.308
+	<a href="#">5522.tw</a>	遠雄建設事業	81.2	5072040	18.3579
+	<a href="#">1201.tw</a>	味全	19.15	516308	35.6604
+	<a href="#">1516.tw</a>	川飛能源	19.85	237649	38.4314
+	<a href="#">6184.tw</a>	大豐有線電視	54.5	201252	15.584
+	<a href="#">9940.tw</a>	信義房屋	33.5	415124	13.9749
+	<a href="#">2889.tw</a>	國票金融控股	15.7	10386282	26.8103
+	<a href="#">1434.tw</a>	福懋興業	23.9	812434	71.875
+	<a href="#">1460.tw</a>	宏遠興業	8.49	2563628	-4.50556
+	<a href="#">2331.tw</a>	精英電腦	32.7	7782246	16.6111

股價排序 成交量排序

圖 5-19 「推薦給您的股票」畫面

您已加入的股票		
	股票代碼	公司名稱
-	<a href="#">2317.tw</a>	鴻海
-	<a href="#">2330.tw</a>	台灣積體電路製造
-	<a href="#">2610.tw</a>	中華航空
-	<a href="#">2884.tw</a>	玉山金融控股
-	<a href="#">2887.tw</a>	台新金融控股

圖 5-20 「您已加入的股票」畫面

股票投資組合	
具有最小波動率的風險投資組合：	
收益率	0.16
波動率	0.15
夏普率	0.97
股票代碼	權重
2610.tw	0.03
2317.tw	0.14
2330.tw	0.1
2884.tw	0.39
2887.tw	0.34
具有最大夏普率(最優)的風險投資組合：	
收益率	0.21
波動率	0.17
夏普率	1.12
股票代碼	權重
2610.tw	0
2317.tw	0.08
2330.tw	0.44
2884.tw	0.21
2887.tw	0.27

圖 5-21 「股票投資組合」畫面

#### 5.4.6 「股票預測」頁面

左側會顯示使用者加入的股票，當使用者點選箭頭後，伺服器會發訊息給股票預測的模型，模型會從資料庫中提取該股票預測時需要用到的資訊，最後會將結果繪製成圖表顯示給使用者參考（如圖 5-22 所示），分別為 History 跟 Prediction 兩條線供使用者參考。

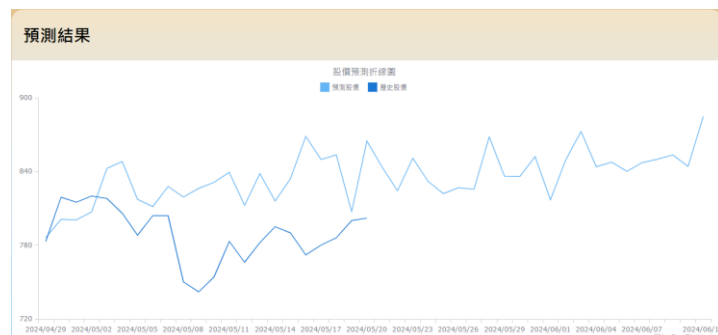


圖 5-22 股價預測結果顯示畫面



## 六、 結論

本專題探討將不同模型進行線性規劃組合模型權重是否會使預測誤差降低，我們使用 2020 年 1 月 1 日到 2024 年 5 月 17 日的收盤價進行股價模型預測。

本文的研究結果顯示在四個月的預測下，經過線性規劃調整模型權重後，LP 模型的平均 RMSE 有相較於使用個別模型來的要更低，同時在統一(食品業)和中鋼(鋼鐵業)相對台積電(半導體)和鴻海(電子工業)有著更好的準確度，推測原因是食品和鋼鐵業本身的股價波動性沒有半導體和電子工業來的高，因此在預測上的誤差較小。

在一個月的預測結果，LSTM 能更好的預測股價波動性相對大的股票，LSTM 在四個月的預測結果明顯的比 ARIMA 和 MCMC 來的要好，符合 Siarni-Namini[1]的研究結果，MCMC 在中長期預測結果下，是最為差勁的。

最後從實驗結果可以得知，經過 LP 模型後，在一個月的預測結果表現上平均 RMSE 增加 38%，是因為鴻海近期的股價波動劇烈，導致在個股預測的 RMSE 極度不準確，反之在四個月的預測結果，LP 模型平均 RMSE 減少了 36%，意味著在四個月的預測下，此模型的準確度比一個月的預測結果還要精準。

未來在模型的特徵值會在增加基本面和技術面資訊，讓模型有更好的準確度，會建議增加新聞資訊作為特徵值，近期鴻海股價上漲，拿到 GB200 伺服器的單，這是在基本面和技術面看不到的，以及將表現最差的 MCMC 方法替換成其他模型。

在更新系統的方向，未來將朝向更便利的設計，而非傳統的 K 線圖或基本面指標，而是直接做出統計數居分析，利用歷史回測數據和同業財報比較，去告知使用者某檔股票的價格是否有被高估，試圖從更簡單的角度讓使用者對股票市場有更低的門檻去做參與。

## 七、 參考文獻

- [1] Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series.
- [2] Hariadi, M., Muhammad, A. A., & Nugroho, S. M. S. (2020). Prediction of Stock Prices Using Markov Chain Monte Carlo.
- [3] Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock Price Prediction Using the ARIMA Model.
- [4] 李培煜 (2021) 。時間序列 ARIMA 與深度學習 LSTM 預測模型之比較：以台灣股票市場為例。
- [5] Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77-91.
- [6] Box, G.E., G.M.Jenkins, and G.C.Reinsel (1994), *Time Series Analysis : Forecasting and Control*. new jersey: prentice-Hall Inc.
- [7] Dickey and Said E (1984), Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order. *Biometrika* Vol. 71, No. 3 (Dec., 1984), pp. 599-607.
- [8] Akaike, Hirotugu.(1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6): 716–723.
- [9] S. Hochreiter and J. Schmidhuber.(1997), Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [10] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. and 53 Teller E. (1953), “Equations of State Calculations by Fast Computing Machines,” *J. Chemical Physics*, Vol. 21, pp.1087-1091.
- [11] Hastings, W. K. (1970), “Monte-Carlo Sampling Methods Using Markov Chains and Their Applications,” *Biometrika*, Vol.57, pp. 97-109.
- [12] Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3), 637-654.
- [13] [Python|選擇權定價 Monte Carlo simulation | by kuan\\_min | Medium](#)
- [14] [淺談遞歸神經網路 \(RNN\) 與長短期記憶模型 \(LSTM\). 介紹遞歸神經網路 \(Recurrent Neural Network\)··· | by TengYuan Chang | Medium](#)
- [15] [時間序列分析 1 - ARIMA 基本理論\(學習筆記\) | by Jason | Medium](#)
- [16] [Stock prediction using 3 methods \(LSTM+ARIMA+MCMC\)](#)
- [17] [Yahoo Finance - Stock Market Live, Quotes, Business & Finance News](#)