



嶺東科技大學  
LING TUNG UNIVERSITY

資訊管理系

# 應用類神經網路探討影響 5G 類 股投資報酬率之關鍵因素

指導教授：謝淑玲 教授

組員名單：  
楊舒涵 A68C025  
梁譽薰 A68C048  
林靖祥 A68C047  
黃啟竣 A68C052  
陳品翰 A68C016

中 華 民 國 一 一 〇 年 五 月



## 謝 誌

本專題報告得以順利完成，首先要感謝 謝淑玲 老師細心引導及協助我們，從剛開始完全不懂，到克服研究過程中所遇到大大小小的問題，慢慢進入狀況，指導老師都不厭其煩的幫助我們尋找解決方法，才使得本專題可以順利完成。

接著非常感謝口試期間的馮曼琳老師、陳健忠老師辛苦的細心審閱，並且提供寶貴的建議，使我們的專題內容更加完善，在此由衷的感謝。

也感謝所有組員的互相扶持，在遇到困難時紛紛提出自己的解決方法互相指教，讓我們能得以做出更完整的結果，以及黃國華主任在百忙之中抽空來當主持人，感謝這麼多幫助過我們的老師及同學們，在此一併致上最高謝意。

楊舒涵  
梁譽薰  
林靖祥 謹誌  
黃啟竣  
陳品翰

中華民國一一〇年五月於嶺東科技大學

# 摘 要

隨著資訊科技的進步，許多資訊科技資料與日俱增地累積下來，伴隨而來的是累積了未經過整理的大量資料，面對這些資料，我們可以運用資料探勘技術尋找深藏在資料內容的資訊，從大量的數據資料中找出適合的演算法，做更進一步之解釋及分析，便能解決傳統繁瑣的計算工作，成為一個龐大的資料庫。

本專題主要目的是應用類神經網路探討影響 5G 類股投資報酬率之關鍵因素，透過文獻資料中找出重要的 11 項財務指標，接著從台灣經濟新報 TEJ 資料庫中，取得台灣 5G 類股作為研究資料集，再利用 SPSS 決策樹篩選出更重要的指標，最後用類神經的方法驗證準確度，進而找出影響投資報酬率的關鍵因素。希望能輔助投資人對於財務指標的選擇以及投資策略的擬定，並降低持有的風險。

關鍵詞：資料探勘、類神經網路、決策樹、5G 類股

# 目 錄

摘 要 .....	I
目 錄 .....	II
表目錄 .....	III
圖目錄 .....	IV
第壹章 緒論 .....	1
1.1 研究背景 .....	1
1.2 研究目的 .....	1
第貳章 文獻回顧與探討 .....	3
2.1 資料探勘 .....	3
2.2 決策樹 .....	5
2.3 類神經網路 .....	7
2.4 財務指標 .....	8
2.5 應用類神經網路之相關文獻 .....	10
第參章 研究方法 .....	12
3.1 研究流程 .....	13
3.2 篩選重要變數 .....	14
3.3 資料預處理 .....	14
3.4 決策樹 .....	16
3.5 類神經網路 .....	17
3.6 研究範圍 .....	17
3.7 研究環境 .....	19
3.7.1 台灣經濟新報 TEJ .....	19
3.7.2 統計產品與服務解決方案 SPSS .....	20
第肆章 實驗結果與分析 .....	21
4.1 決策樹篩選財務指標之重要變數 .....	22
4.2 類神經網路模型 .....	25
第伍章 結論 .....	27
參考文獻 .....	28

## 表目錄

表 1 資料探勘四種主要模組 .....	4
表 2 類神經網路探討股票之相關文獻 .....	10
表 3 篩選重要變數 .....	14
表 4 股票公司名稱 .....	17
表 5 類神經網路模型輸入變數 .....	25
表 6 類神經網路實驗成果 .....	26

## 圖目錄

圖 1 CRISPD-DM 標準流程圖.....	5
圖 2 類神經網路架構圖 .....	7
圖 3 類神經網路架構模型圖 .....	8
圖 4 研究流程圖 .....	13
圖 5 資料探勘執行流程圖 .....	15
圖 6 台灣經濟新報(TEJ)系統圖 .....	19
圖 7 SPSS MODELER 操作介面 .....	20
圖 8 TEJ 匯出 EXCEL .....	21
圖 9 匯入 EXCEL 資料集 .....	22
圖 10 K-MEAN 集群分析 .....	22
圖 11 設定分割區 .....	23
圖 12 設定欄位 .....	23
圖 13 建立 C5.0 決策數模組.....	24
圖 14 決策樹實驗結果 .....	24
圖 15 類神經網路模型摘要 .....	25

# 第壹章 緒論

## 1.1 研究背景

近年來，經濟不穩定、定存利率低，大多人會選擇投資報酬率較高的理財方式，例如：房地產、基金、期貨、或是資本的增值股票...等，其中投資股票可說是相當普遍及重要的一種金融活動，可小額投資，門檻也相對較低，使得股票成為眾多投資者的首要選擇，投資人買賣股票皆希望能夠有高報酬，並且能進一步去降低風險。

面對經濟環境變遷，投資理財成為一項重要的課題，為了瞭解公司營運狀況，投資人也會運用各種層面的分析來選擇買賣股票。因此，當公司營運策略轉變時，投資人便會釋出或購買該公司股票，由此可知股票價值可以反映公司價值，使得在投資時能夠做出最好的決策，能有效增進股票投資報酬能力。

不過，科技日新月異，行動通訊產業鏈包含甚廣，5G(5 Generation)世代慢慢開始嶄露頭角，5G 為發展到第五代的無線通訊技術。從 1G 到 5G 的演進，時代的變遷，訊息的產生、傳輸到接收資料，網路通訊的背後蘊含著無法計算的行動智慧。每一個新的通訊世代，代表的除了更快的網路速度外，5G 最大的改變就是實現從人與人之間走向人與物、物與物之間的行動通訊，實現萬物互聯，並且推動社會發展。例如:移動醫療(Mobile Health， M-Health)、全自動無人駕駛(Fully Automatic Unmanned， UTO)、車聯網(Internet of Vehicles， IoV)、物聯網(Internet of Things， IoT)等。

本研究將以利用決策樹及類神經網路等資料探勘技術，探討投資人投資台灣 5G 股票投資報酬率之關鍵因素，從眾多的財務指標中快速且有效地做篩選，以輔助投資人對於財務指標的選擇以及投資策略的擬定，並降低持有的風險，進而獲取有用的數據訊息作為制定決策之根基，以獲取更高的投資報酬率。

## 1.2 研究目的

目前投資市場上有許多的股票選擇，若能運用資料探勘的功效找出其中潛存的規律性與行為模式，並鑑別出隱藏在資料庫中的樣式，則可藉以預測未來可能發生的現象，挖掘出過去以大量人力資源才能揭露的隱藏樣式，本研究運用了決

策樹及類神經網路來探討股票投資報酬率之關鍵因素，並以台灣最近興起的 5G 網路較具代表性的 47 家公司為研究對象，並找出之間影響 5G 股票投資報酬率的關鍵因素，且透過這 47 家公司的股票互相分析比較得知各個公司股價預測的準確性以及報酬率，本研究目的如下：

- (1) 了解 5G 類股各公司的財務狀況概況，提供給投資人投資相關 5G 類股的參考因素，以期提高投資報酬率。
- (2) 應用類神經網路及決策樹技術，探討提升 5G 類股投資報酬率之關鍵因素，讓往後更多投資者已可以更了解其中的產業發展趨勢、走向以及往後對股價市場的影響力。
- (3) 利用各項財務指標，更深入的了解其未來走向趨勢以及預測未來股價的漲跌，讓投資者可以更有效的、精準的從中獲得收益。

## 第貳章 文獻回顧與探討

為探討 5G 股票的研究方法及解決問題模式，本章節蒐集相關理論及過去之研究成果，內容涵蓋資料探勘、決策樹、類神經網路與財務分析及探討 5G 股票之相關文獻。

### 2.1 資料探勘

在資訊科技逐漸發達的時代，人工流程大多邁入電腦化，伴隨而來的問題即是累積了大量、複雜，未經過整理的資料，為了快速且有效率地處理這些數據資料，以往繁瑣的計算與傳統統計報告已無法滿足現階段使用者的需求了。

運用資料探勘的技術，找出適合的演算法，就能進一步分析大量的數據資料，此技術能解決傳統繁瑣的計算工作。

資料探勘也稱為數據挖掘，資料挖掘，資料採礦。在儲存資料庫中，從大量歷史資料，挖掘出有用的資訊。近幾年，資料探勘技術相當地熱門，許多商業、金融業、製造業、開發業都會積極利用資料探勘的方法，掌握所挖掘出的資訊，再選擇其投資，能有效降低風險並從中獲取利益[1]。

資料探勘可分為四種主要的模組：資料分群、資料關聯、資料分類及循序樣式探勘，如表一所示。

表 1 資料探勘四種主要模組[2]

模組	敘述
資料分類	針對資料集，分析資料的屬性和特徵分類加以定義，以建立類組(class)。ex:用於已經分類的資料，分析顧客的共同點，再依據這些共同點對尚未分類的資料做預測分類。
資料關聯	探討資料項目之間的關係，找出資料、其他項目集合相關性的關聯規則。ex：顧客在購買服飾時通常還會再購買鞋子，這樣服飾與鞋子之間就構成關聯性，可依照這樣的關聯性去擺放商品位置來刺激消費者選購，達到最高獲利。
資料分群	主要將資料相似與相異之處分別出來，分別歸類到數個群集。ex：把商場內顧客購買服飾的品牌及顧客資料逐一分群，統計銷量後，即可預測出某一客群偏好使用某一品牌商品，再加以推銷。
循序樣式探勘	主要分析一些按次序排好的行列轉變，所得到的結果作為趨勢，預測未來的狀態。與資料關聯最大不同點的是循序樣式探勘以時間區分。ex：分析顧客在購買商品時，會先購買哪些商品，以何種目標做為優先考量，藉由顧客購買商品的序列模式，再採用更能獲利的行銷方式。

在資料探勘分析上一個標準作業的流程，該流程涵蓋了所有資料探勘研究步驟。Cross-Industry Standard Process for Data Mining SPSS 和 NCR 在 1996 年為克萊斯勒做資料採礦時制訂。CRISPD-DM 標準流程，所需的步驟及如圖 1 所示。

- (1) 商業理解 (Business Understand)：資料挖掘最重要的部分。此流程包括了解探勘的方向及目標，規劃如何收集資料，分析顧客給予的訊息及制定用何種方式呈現報告。
- (2) 資料理解 (Data Understand)：此流程是從初始的資料收集、整理資料後，對資料做分析，過濾出相關資料，從資料中取得獨立相關變數，其目的是可以降低訊息的重複性，提高變數最終的效果。
- (3) 資料預備 (Data Preparation)：當確定資料來源分類之後，將資料清理、重構、整合及缺漏值的填補，使得資料完整性更加的完善。
- (4) 塑模 (Modeling)：將大量的資料分割成訓練與測試資料集，選用探勘演算法及參數設定，完善的分析方法將用在資料中提取資訊，依分析結果來回修正。
- (5) 評估 (Evaluation)：評估當前的分析結果是否符合企業的目標與預期結果，

獲取資料中的隱藏資訊是否具有價值性。

- (6) 部署 (Deployment)：若適用於企業，則繼續此項計畫，且須監控與維護該資訊，然而產生最終報告及回顧整個專案。

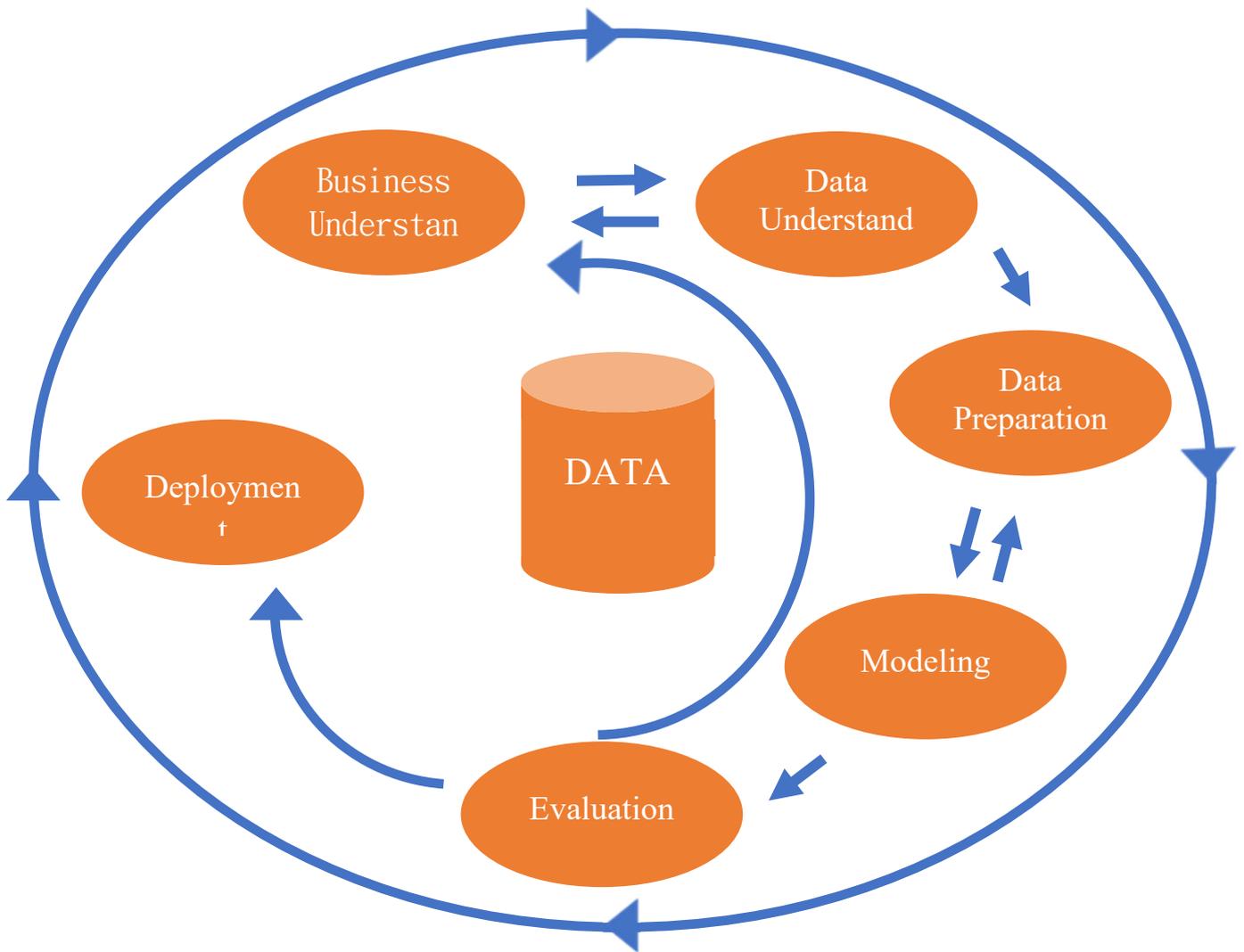


圖 1 CRISPD-DM 標準流程圖[3]

## 2.2 決策樹

決策樹又稱分類樹，是一種監督式特徵萃取的演算法，廣泛的應用在解決分類型態上的問題，讓使用者依照其目的將資料依據特性來加以分類，使用者可以利用決策樹之各項不同的變數來判斷以及預測可能之結果。

決策樹的主要功能，是藉由分類來建立出樹狀結構，所產生出來的決策樹具有規則。依照各種方法的不同，步驟也不同，要產生決策樹，須先以訓練資料建

立決策樹的模組，再進行測試資料及修剪決策樹，以及提升正確率，最後才可預測新的資料[4]。

如果決策樹的樹狀圖過多，代表每個葉節點包含的樣本資料數太少，會產生過度配適現象，過度配適是模型對於範例的過度訓練，導致模型記住的不是訓練資料的普通特性，是資料的局部特性。對測試樣本的分類反而會變得不精確[17]。

如果原始的訓練資料呈現不完整或稀疏、有雜訊時，可能會導致產生的決策樹太過複雜，因此必須再做適當的修剪[17]。一般分為以下兩種修剪方式：

- 1.事前修剪：運用統計門檻值加以衡量，例如：卡方值或資訊獲得值等技術，評估是否該繼續分割某內部節點成數個子分支或是應該立刻停止。
- 2.事後修剪：允許決策樹超適情形的合理存在，當完成決策樹的建立之後，再來進行修剪的程序。

### K-means 集群分析法

屬於資料探勘領域，本次研究使用 K-means 演算方法[5]，固分別提出：

- (1) 選定變數 K 個初始集群的中心，K 為欲分群的數目。
- (2) 計算各個觀察值到每集群中心的遠近距離，再將每一個觀察值逐一分配到離其最近的集群內。
- (3) 在事先假設的調整規則中，重新分配每一個觀察值到 K 組集群中。
- (4) 如果重新分配的資料點能滿足規則條件，則重複步驟(2)、(3)，直到資料點無法重新配置為止。

## 2.3 類神經網路

類神經網路又稱為人工神經網路，是一種由許多個處理單元以平行計算所構成的，包括了硬體和軟體，它使用大量的相連人工神經元來模仿生物的能力。

歷經長時間不斷進化的人腦，目前學習能力和應變能力遠超過所有的數位電腦和生物。因此我們猜測人腦中有某種優秀的運算邏輯，才能使人腦完成困難的學習和辨識工作。

基本原理為使用最除坡降法之觀念，將誤差函數給予最小化，倒傳遞類神經網路之架構，而架構之功能分述及模型主要有網路連接、神經元、學習規則等，如圖 2 所示。

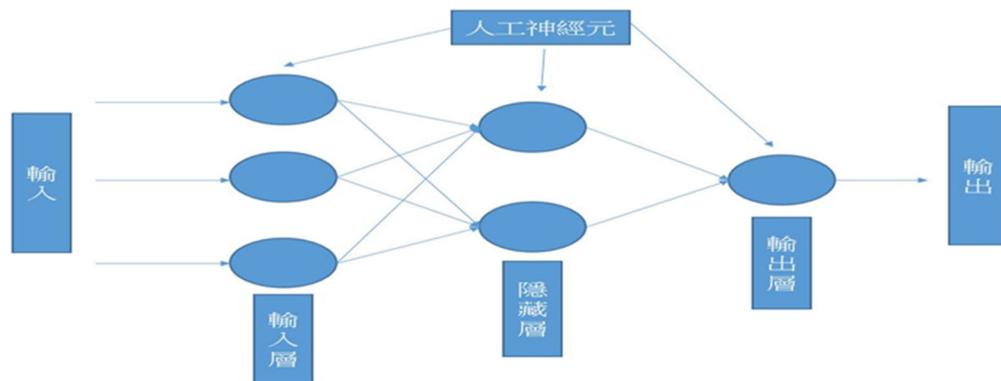


圖 2 類神經網路架構圖[18]

自 1986 年由 Rumelnar 等學者研究發表後，類神經網路便飽受注意，基本元件如下所敘述：

1. 輸入層：接收訊息。
2. 隱藏層：接收輸入層的訊息加以處理。
3. 輸出層：將處理好的訊息傳出。

倒傳遞類神經網路演算法分為兩個步驟，第一個階段為「學習過程」；第二個階段為「回想過程」。

1. 學習過程：網路依學習演算法，從範例中學習，經反覆的運算，以調整網路連結的加權值。
2. 回想過程：網路接受外來輸入，並依回想演算法，經反覆運算後，由輸出及神經元將結果送出。

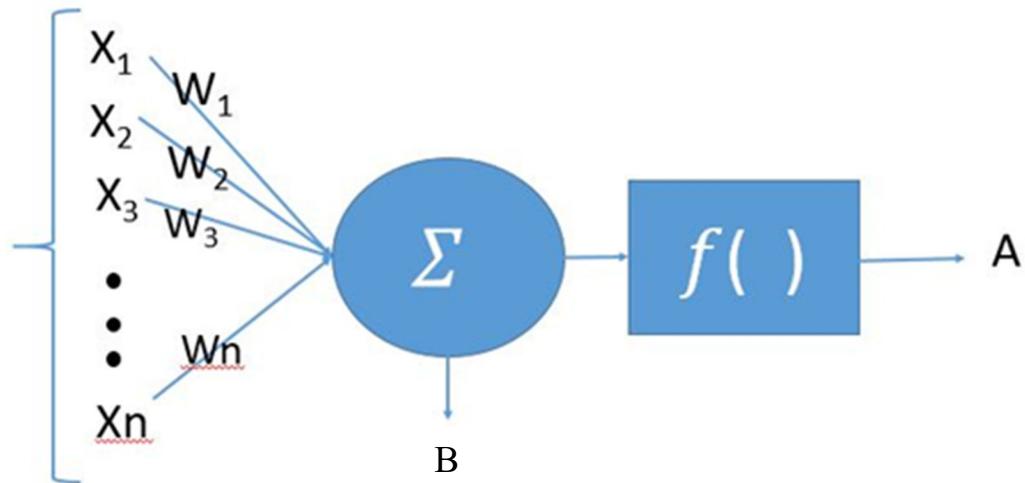


圖 3 類神經網路架構模型圖[18]

X：輸入訊號。

W：連接加權值。

B：閾值或偏權值。

$\Sigma$ ：集成函數，將 X 值和 W 值相乘後的總合。

$f()$ ：轉換函數，將 X 值、W 值和 B 值轉換成處理單元輸出。

A：輸出結果。

## 2.4 財務指標

財務指標分析[6]指總結和評價企業財務狀況與經營成果的分析指依據民國 92 年 12 月 30 日修訂的「公開發行公司年報應行記載事項準則」之第 19 條之 1 規定，公司所提供之財務概況應記載下列事項最近五年度簡明資產負債表及損益表，並註明會計師姓名及其查核意見。公司過去五年度的簡明資產負債表及損益表都會存留一份於台灣證券交易所，以供投資。

本研究主要參考整體經濟因素，經由過去學者的研究中，我們採用較常用到的 11 項財務指標，再透過決策樹篩選出更重要之財務指標，用於預測模型。

### 一、股東權益報酬率

每一段時間內，公司利用股東投資，創造利潤，股東權益報酬率越高，代表公司經營能力佳。是一種衡量公司獲利能力的財務指標。

公式：股東權益報酬率=稅後盈餘/股東權益總額

## 二、總資產報酬成長率

用來衡量公司是否充分利用公司資產，創造利潤。比率越高代表公司資產報酬運用的效率越好，具有良好的獲利能力。

公式：總資產報酬成長率=(稅後淨利+利息×[1-稅率])/平均總資產×100%

## 三、常續性 EPS (Earnings Per Share)

是指公開市場上，每股給投資者帶來的收益。是公司的獲利指標。

公式：常續性 EPS = 常續性利益 / 流通在外股數

## 四、營業利益率

企業創造一單位的營收所能產生的獲利。由於計算方式將過程中耗用的成本列入考量，因此在關注企業的獲利能力時，營業利益率是一項重要指標。

公式：營業利益率=(營業收入-銷貨成本-營業費用)/營業收入

## 五、銷貨毛利率

毛利率越高則表示企業的盈利能力越強，控制成本的能力越強。

公式：銷貨毛利率=(銷售收入-銷售成本)/銷售收入×100%

## 六、稅前淨利率

稅前淨利率高時，代表公司運用所有公司資源的獲利能力強。

公式：稅前淨利率=稅前淨利/營收×100%

## 七、流動比率

用來衡量公司短期內償還債務能力，又稱營運資金比率。過低的流動比例代表若發生問題時，可能會發生週轉不靈。

公式：流動比率=流動資產/流動負債

## 八、速動比率

用來衡量公司流動性的重要比率之一，又稱為酸性測試比率。

公式：速動比率=(流動資產 - 存貨)/ 流動負債

## 九、現金流量比率

反映了企業透過經營獲取足夠資金來償還債務和兌現承諾的能力。

公式：現金流量比率=營業活動的淨現金流量/流動負債

## 十、負債比率

當企業處於成長期，獲利持續向上，適當的舉債能夠提供企業充足的資金擴張，創造更多獲利來源。

公式：負債比率=負債/總資產

## 十一、長期資金適合率

是指固定資產的長期資金來源，反映了企業的償債能力。比率愈高表示資金來源愈穩固。

公式：長期資金適合率=(業主權益+長期負債)/(固定資產+長期投資)×100%

## 2.5 應用類神經網路之相關文獻

類神經網路是模擬人腦的神經系統所建構的模型，主要從過去資料樣本中學習，再經過許多的平行處理，進而達到知識的累積及呈現的目的。

財務分析和預測的領域方面大量使用到類神經網路這項技術，加上類神經網路可以接受模糊不清和不完整的輸入訊息，因此在提高財務獲利率或者是傳統分析方法上都有的顯著的成果，不過目前看來在預測未來的股價漲跌幅正確度還是有不少成長空間。

曾思博[7]提出「資金配置」理論，判斷出類神經網路輸出的信心值大小決定入資金的比重程度。信心值越大，投入資金的比重越大，反之，信心值不足夠的時候，只能投入部份資金，或者完全不交易。在實證中發現，資金配置模型結合類神經網路股價預測，將會使整體個投資報酬率提高，統整後依時間排序如表 2 所示。

表 2 類神經網路探討股票之相關文獻

時間	作者	方法	結論
1991	Bergerson & Wunsch[19]	倒傳遞類神經網路專家系統	投資報酬可達 6.6 倍。
1995	Chiang, Urban, Baldrige[18]	倒傳遞類神經網路、線性迴歸、非線性迴歸	預測績效：類神經網路模式優於線性迴歸模式，線性迴歸模式又優於非線性迴歸模式。

時間	作者	方法	結論
1996	故國瑜[8]	類神經網路	紡織與電子電機兩項產業的獲利績效相當，且都比整體市場來得好。
1998	張道行[2]	倒傳遞類神經網路	受到股價短線消息影響波動，所以使用短期技術指標成功率比以基本陳資料預測長期股價趨勢的成功率來得低。
1999	吳秉奇[6]	倒傳遞類神經網路	等待出場策略比一般出場策略較佳，利用較少次的交易數累積獲利。
1999	曾思博[7]	倒傳遞類神經網路	利用資金配置投入，每次投入相同資金有比較好的報酬率。
2004	黃綺年[9]	集群分析、自我組織映射網路、機率類神經網路、迴歸分析、倒傳遞類神經網路、改良式倒傳遞類神經網路	無論分類或預測效果類神經網路均優於統計方法。
2005	蔡正修[10]	迴歸分析、倒傳遞類神經網路、適應性網路模糊推論系統	預測效果以傳統統計方法最佳，其次為改良式倒傳遞類神經網路，而適應性網路模糊推論系統績效最差。
2011	蘇偉庭[11]	自我組織映射網路、類神經網路	自我組織映射網路預測結果較為準確，平均預測度 60%
2015	端木和奕[12]	類神經網路、倒傳遞類神經網路、模糊分群	沒有一種指標對每檔股票都最有關連性，只有對某檔股票適合的指標。良好的分群有助於類神經預測。
2017	簡壬申[13]	類神經網路	類神經模型擬定投資策略買賣及持平，準確率達 57%
2018	林憲民[14]	線性迴歸、類神經網路	受到外資操盤情況增加，導致技術指標較重要於財務指標

## 第參章 研究方法

本研究透過台灣經濟新報 TEJ 資料庫，收集5G概念股公司2014年到2019年共5年期間的季資料為樣本，經由過去諸多學者研究中採用11項財務指標，資料共計1344筆，先進行資料預處理，清除重複及無效(空值)的資料紀錄，以確保每間公司資料的完整性，清除異常值後，共計採納47家5G概念股公司的11項財務指標資料，共計1128筆，研究方式使用了SPSS Modeler決策樹用於篩選出更重要的變數，最終篩選出6項財務指標，再利用類神經網路建立模型，進行實驗，以期找出影響5G類股投資報酬率之關鍵因素。

### 3.1 研究流程

本研究流程分為六個步驟，以下說明。

- [1]研究動機與目的:針對 5G 概念股進行股價重要因素預測。
- [2]搜尋相關文獻:找出過去較常用的 11 項財務指標及相關變數用於參考。
- [3]決策數篩選:運用決策樹進一步做變數的重要性篩選。
- [4]建立預測模型:建立類神經網路模型，預測準確度。
- [5]實驗結果分析:用模型預測結果來分析影響 5G 概念股投報率之因素。
- [6]研究結論:最後將分析後的結果，整理為結論。

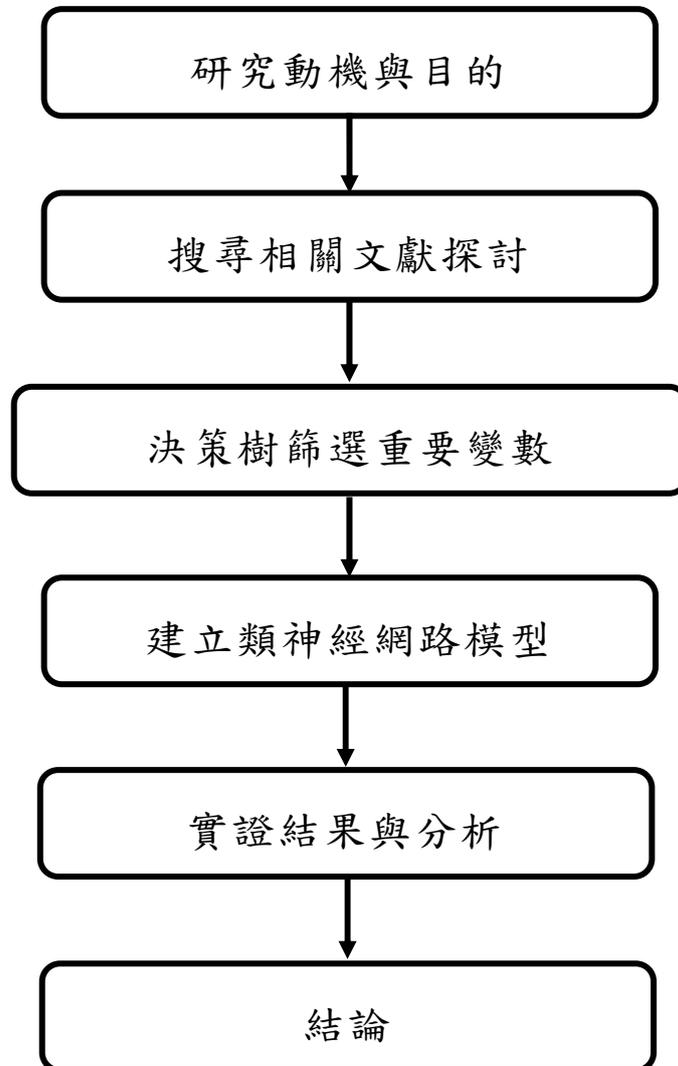


圖 4 研究流程圖

### 3.2 篩選重要變數

本研究主要參考整體經濟因素，經由過去許多學者的研究中，我們採用最常用到的 11 項財務指標，再透過決策樹篩選出更重要之財務指標，用於預測模型，如表 3 所示。

表 3 篩選重要變數

財務指標	
股東權益報酬率	總資產成長率
常續性 EPS	營業利益率
銷貨毛利率	稅前淨利率
流動比率	速動比率
現金流量比率	負債比率
長期資金適合率	

### 3.3 資料預處理

大部分資料是不完整、不一致的，無法直接進行資料探勘，或探勘結果差強人意。為了提高其質量，便使用了資料預處理技術，像資料清理、集成(整合)等，這些資料處理技術在資料探勘之前使用，將能大大提高資料探勘模式的質量，還能降低實際所需要的時間。

對於原始資料先稽核完整性和準確性。完整性稽核是檢查單位或個體是否有遺漏，所有的調查指標及項目是否齊全。準確性稽核主要包括二點，一是檢查資料資料是否真實地反映了實際情況，內容是否符合實際；二是檢查資料是否有錯誤，計算是否正確等。

最後做資料清理，主要是要達到格式標準化，清理填寫缺失或異常的值，重複資料的清除。最後資料整合(Data integration)，目的為將蒐集到的大量資料合併與對齊，達成資料一致性、排除資料重複等問題，能更準確、快速地進行資料探勘。

資料處理首先要先經過四項工作，分別是選取輸入資料、轉換資料、資料探勘、解釋結果，如圖 5 所示。

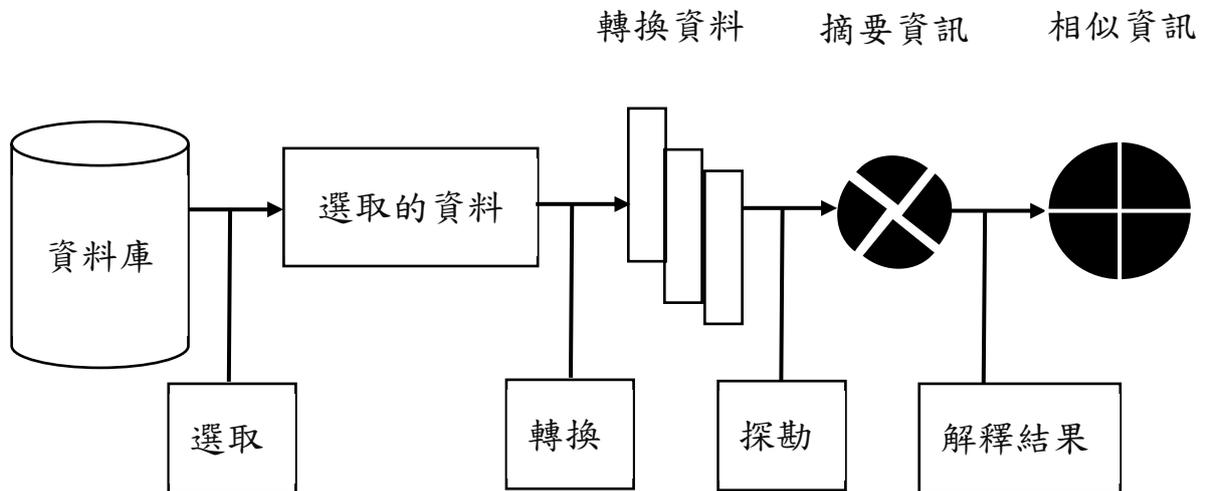


圖 5 資料探勘執行流程圖[15]

選取輸入資料：是指定要探勘和分析的資料，資料來源是 TEJ 資料庫中的公司資料，從多個資料表中，取得表格或記錄文字檔。

1. 轉換資料：為降低資料量，首先將收集的資料作整理、清除重複及無效（空值）的資料紀錄，確保公司資料的完整。
2. 資料探勘：方法使用分類、分群及關聯等，從轉換後的資料中發掘存在的多種特徵及資訊。
3. 解釋結果：以 Excel 檔案匯出結果。

### 3.4 決策樹

本研究採用 C5.0 決策樹演算法，主要能夠處理連續型變數與離散型變數作以解析，並將運算的結果產生出決策樹(decision tree)或者是規則集(rulesets)兩種模型，再經由最大資訊增益 (information gain value)的欄位來分析樣本。

以下為 C5.0 決策數演算法資訊增益計算公式說明如下：

在 C5.0 初始化時，在不考慮屬性下資訊亂度：

$$Info(K) = - \sum_{b=1}^a \left\{ \left[ freq \left( C_b, \frac{S}{|X|} \right) \right] \log_2 \left[ freq \left( C_b, \frac{S}{|X|} \right) \right] \right\} \quad (1)$$

考慮屬性的情況下資訊亂度：

$$Info_y(X) = - \sum_{b=1}^w \left[ \left( \frac{|X_b|}{|X|} \right) Info(X_b) \right] \quad (2)$$

公式(1)與公式(2)相減後的資訊度差量：

$$Gain(Y) = Info(B) - Info_y(B) \quad (3)$$

分割節點的熵值：

$$SplitInfo(Y) = - \sum_{b=1}^w \left[ \frac{|X_b|}{|X|} \log_2 \frac{|X_b|}{|X|} \right] \quad (4)$$

資訊增益比例：

$$GainRatio(Y) = \frac{Gain(Y)}{SplitInfo(Y)} \quad (5)$$

在公式 (1) 中  $|X|$  是為訓練資料在各類別之總個數；目標的屬性中含有的類別為  $C$ ；在類別下包含的資料筆為  $X_b$ ；在公式 (2) 中  $|X|$  是為在選定屬性下的訓練資料中各類別之總個數；輸入的屬性數量為  $W$ ； $Info_y(B)$  代表在某一情況選擇下一個候選屬性的熵值 (Entropy)；在公式 (3) 中  $Gain(Y)$  代表資訊增益,  $Info(B)$  為增加某一屬性之熵值。

本研究選用 SPSS 決策樹為篩選重要變數的方法，將 47 家 5G 概念股的財務指標設為輸入資料，先利用 K-means 將資料庫分類為二群，分別為營餘與非營餘公司，再將 11 個財務指標導入，再將資料分割為 67% 的訓練資料與 33% 的測試資料，篩選後得到更重要的 6 項財務指標，依重要程度排序為銷貨毛利率、速動比率、負債比率、總資產成長率、股東權益報酬率、現金流量比率，共計 1128 筆。

### 3.5 類神經網路

類神經網路是一種運算系統，範圍包含了軟硬體，用於模擬生物神經的能力，人工神經元可從外在環境中學習，並將學習的資料，加以分析後給予權重，再透過運算後神經元轉換為函數，最後轉換計算得到輸出值，用於套用至外在環境。

人工神經元的輸入值與輸出值關係式如下：

$$Y_j = f(\sum_i w_{ij}x_i - \theta_j)$$

$Y_j$  = 輸出變數，模擬神經元模型輸出訊號

$f$  = 模擬神經元的轉換變數，包含應限制函數、線性及非線性函數

$w_{ij}$  = 連結加權值，模擬神經元模型的連結強度

$x_i$  = 輸入變數，模擬神經元輸入訊號

$\theta_j$  = 偏權值，模擬神經元的閾值

本研究先使用 TEJ 資料庫取得 5G 相關公司類股的資料集，經由決策樹篩選重要變數後，再將資料集分割為 67%的訓練資料與 33%的測試資料，利用 K-means 將資料庫分類出營餘與非營餘公司，將篩選過後的財務指標輸入模型內，模型採用多層感知器 MLP，隱藏層設定為 1 層。

### 3.6 研究範圍

本研究從 TEJ 資料庫選擇較具指標性的台灣 5G 產業各公司股票，採用 2014 年到 2019 年期間季資料為研究資料集，共計 1344 筆，再進行資料預處理，去除異常資料後，共計選用 47 間公司，1128 筆資料，公司如表 4 所示。

表 4 股票公司名稱

公司名稱	
2314 台揚科技股份有限公司	2317 鴻海科技集團造股份有限公司
2330 台灣積體電路製造股份有限公司	2332 友訊科技股份有限公司
2345 智邦科技公司	2353 宏碁股份有限公司

公司名稱	
2356 英業達集團	2368 金像電子股份有限公司
2382 廣達電腦股份有限公司	2383 台光電子材料股份有限公司
2395 研華股份有限公司	2454 聯發科技股份有限公司
2455 全新光電科技股份有限公司	2485 兆赫電子股份有限公司
3016 嘉晶電子股份有限公司	3037 欣興電子股份有限公司
3045 台灣大哥大股份有限公司	3062 建漢科技股份有限公司
3105 穩懋半導體股份有限公司	3152 璟德電子工業股份有限公司
3163 波若威科技股份有限公司	3221 台灣嘉碩科技股份有限公司
3234 光環科技股份有限公司	3363 上詮光纖通信股份有限公司
3380 明泰科技股份有限公司	3419 謙裕實業股份有限公司
3450 聯鈞光電股份有限公司	3491 昇達科技股份有限公司
3596 智易科技股份有限公司	3682 亞太電信股份有限公司
3704 合勤科技股份有限公司	4904 遠傳電信股份有限公司
4906 正文科技股份有限公司	4958 臻鼎科技控股股份有限公司
4971 英特磊科技股份有限公司	4977 眾達科技股份有限公司
5274 信驊科技股份有限公司	5349 先豐通訊股份有限公司
5388 中磊電子股份有限公司	5439 高技企業股份有限公司
6152 百一電子股份有限公司	6213 聯茂電子股份有限公司
6269 台郡科技股份有限公司	6274 台燿科技股份有限公司
6285 啟碁科技股份有限公司	8046 南亞電路板股份有限公司
8086 宏捷科技股份有限公司	

### 3.7 研究環境

使用台灣經濟新報(TEJ)資料庫抓取公司資料，再用 SPSS 進行篩選及實驗，以上都是經過眾多考量比較後，再挑選出功能最完善、容易使用者操作的軟體，藉此來進行本次研究。

#### 3.7.1 台灣經濟新報 TEJ

台灣經濟新報(TEJ)[16]成立於1990年4月，是由一群精通資料收集、經濟分析與電腦應用的專業人員共同組成的團隊，專門提供證券金融市場基本分析所需的資訊服務。其主要業務是銷售國內外證金、金融產業與總體經濟方面數據資料，並提供經濟分析、模型設計與資料庫構建方面的諮詢服務。隨著資訊科技蓬勃發展與電腦普及的時代，藉由使用資料庫來進行理財投資分析，勢必成為大家使用的方式之一。因此許多國內外知名的機構，也都以台灣經濟新報作為資料供應商。台灣經濟新報 (TEJ) 系統圖，如圖6所示。

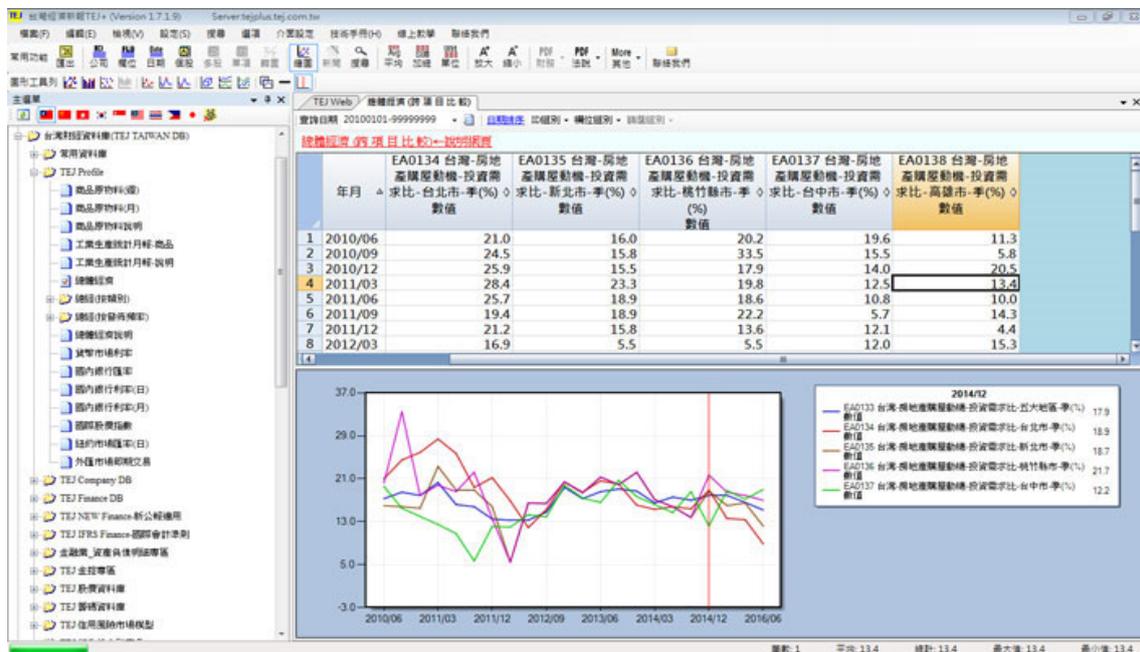


圖 6 台灣經濟新報(TEJ)系統圖

### 3.7.2 統計產品與服務解決方案 SPSS Modeler

SPSS 是 Statistical Product and Service Solutions 的簡稱，為 IBM 公司的一系列用於統計學分析運算、數據挖掘、預測分析和決策支持任務的軟體產品及相關服務的總稱，我們使用的部分是 SPSS Modeler 數據挖掘。利用此軟體進行決策數及類神經的研究與實驗，系統圖展示如圖 7 所示。

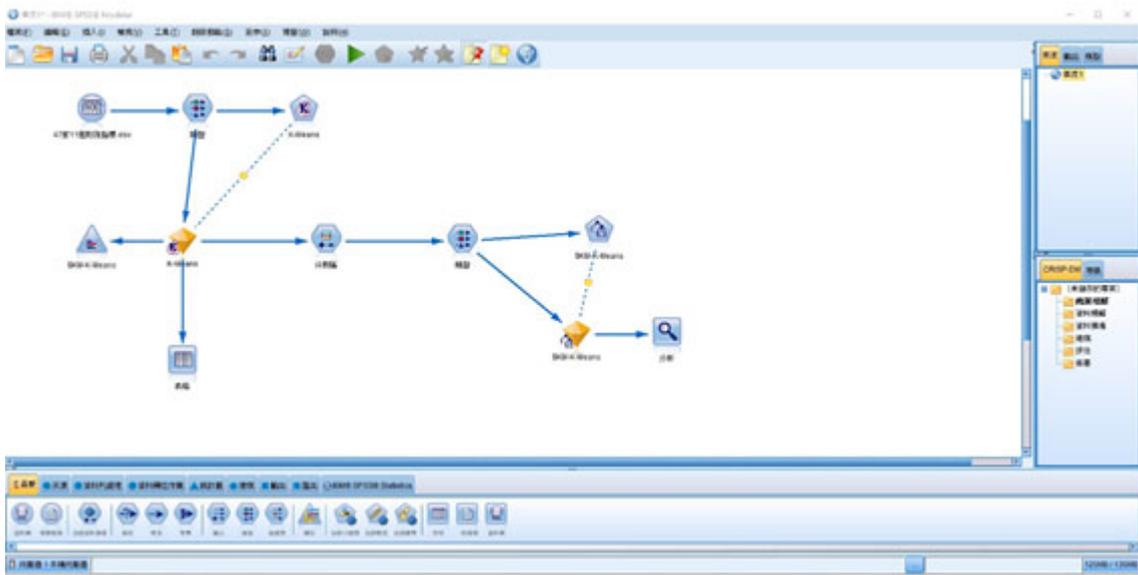


圖 7 SPSS Modeler 操作介面

# 第肆章 實驗結果與分析

在台灣經濟新報TEJ平台中，我們採用47家具有代表性的5G概念股公司股票之交易資料，加入了股東權益報酬率、總資產成長率、常續性EPS、營業利率、銷貨毛利率、稅前淨利率、流動比率、速動比率、現金流量比率、負債比率、長期資金適合率，總共11個財務指標變數來進行預處理，如圖8所示。

	D	E	F	G	H	I	J	K	L	M	N	O	P
1	所獲報酬率	股東權益總額	股東權益報酬率	總資產成長率	常續性EPS	營業利率	銷貨毛利率	稅前淨利率	流動比率	速動比率	現金流量比率	負債比率	長期資金適合率
2	6.491	2,276,391	-0.07	-21.55	2.11	-8.86	10.65	-9.78	115.15	67.33	-2.55	63.17	260.16
3	5711.079	832,100,225	0.02	1.21	8.37	2.37	6.04	2.89	145.41	115.44	10.18	59.41	266.57
4	5455.495	898,692,792	0.05	22.33	7.57	35.43	47.5	35.96	204.07	177.44	54.39	30.8	134.21
5	44.566	13,947,538	0.01	-0.05	0.77	1.2	26.16	1.81	182.08	104.68	-5.07	42.97	1,669.12
6	10,314	7,830,761	0.03	1.13	1.04	0.92	15.67	3.8	197.66	141.73	-3.72	44.69	590.48
7	144.817	57,650,516	0.00	-17.28	-4.45	0.17	8.31	0.19	133.15	101.82	-1.19	68.24	1,225.36
8	758.302	64,101,305	0.03	6.07	2.39	1.67	4.97	2.32	130.74	103.76	-7.69	68.27	233.86
9	-26.328	6,294,496	-0.04	0.13	-1.06	-3.5	7.51	-6.3	99.36	73.04	0.12	71.53	109.64
10	1,298,015	137,076,742	0.04	4.15	4.78	1.34	4.05	2.85	129.09	101.26	4.19	75.03	354.75
11	43,915	7,445,995	0.03	5.11	2.86	7.18	15.8	7	143.3	115.31	12.72	50.97	157.85
12	278,315	21,019,645	0.06	12.72	7.74	16.76	41.48	17.61	229.06	159.34	8.48	28.11	258.89
13	1,361,623	228,398,339	0.04	35.92	22.05	23.5	48.31	25	307.72	279.89	13.34	23.5	1,546.88
14	10,913	3,603,729	0.02	-7.09	1.63	20.18	31.95	22.68	528.14	436.35	25.78	9.23	199.24
15	52,822	7,181,752	0.02	5.6	2.43	4.27	18.17	11.71	388.3	291.19	2.87	22.47	486.15
16	375	1,059,428	0.03	0.71	4.06	0.76	12.98	10.62	209.41	173.69	6.62	28.23	152.53
17	27,477	50,012,804	0.00	-0.61	0.75	0.68	9.16	0.62	153.87	130.08	5.23	48.43	147.4
18	863,488	64,406,303	0.07	36.98	6.36	18.43	35.11	18.32	57.52	48.55	13.79	54.94	204.36
19	5,674	5,147,233	0.00	12.96	0.35	0.25	5.7	0.68	183.13	95.37	12.95	40.63	1,347.13
20	36,595	15,244,872	0.01	-14.08	1.56	12.21	30	15	285.35	230.67	24.69	28.36	152.39
21	17,965	2,650,656	0.04	8.43	7.02	37.08	51.59	42.93	940.81	901.98	81.28	7.72	329.33
22	-5,990	1,591,857	0.04	49.3	4.72	7.98	19.77	13.69	250.52	148.1	-4.48	36.76	315.11
23	8,371	1,553,799	0.03	8.68	1.46	8.23	20.91	11.37	248.13	154.94	9.42	27.07	248.98
24	9,831	1,552,178	0.03	9.04	1.19	9.5	20.97	10.86	224.95	160.35	16.41	29.25	204.57
25	212	1,260,070	0.01	25.53	1.06	0.57	25.54	4.69	480.99	400.49	6.59	25.13	445.72
26	8,597	9,895,427	0.00	5.41	1.88	1.26	13.89	1.05	202.01	110.38	-5.61	41.98	301.28
27	186	1,687,857	-0.01	-4.19	-1.27	-3.65	17.05	-3.64	181.45	109.01	3.57	46.12	281.12
28	55,782	2,723,795	0.07	10.94	7.55	14.95	21.4	16.43	184.56	149.7	26.6	43.61	171.12
29	9,516	1,448,081	0.03	7.41	3.28	17.67	36.71	20.55	323.41	250.29	3.06	16.88	206.25
30	62,627	6,734,830	0.03	3.13	4.16	5.75	16.92	6.82	222.23	134.14	-2.01	40.48	526.44
31	60,025	35,517,057	0.01	-2.79	0.53	7.88	34.9	9.05	251.56	226.48	33.49	10.25	238.81
32	33,418	9,988,878	0.01	11.66	1.02	0.72	27.35	2.18	230.19	145.66	4.63	35.62	509.3
33	658,261	77,401,352	0.04	21.45	4.1	17.94	41.04	16.65	105.26	76.11	35.73	36.8	204.48
34	18,768	9,044,795	0.01	-4.45	1.17	-1.76	11.74	1.89	142.76	120.71	1.26	49.88	232.25

圖 8 TEJ 匯出 Excel

#### 4.1 決策樹篩選財務指標之重要變數

(1) 匯入Excel資料集：採用47家具有代表5G公司的股票交易資料，加入股東權益報酬率、總資產成長率、常績性EPS、營業利益率、銷貨毛利率、稅前淨利率、流動比率、速動比率、現金流量比率、負債比率、長期資金適合率，總共11個財務指標變數來進行預處理，如圖9所示。

公司	年月	股東權益報酬率	總資產成長率	常績性EPS	營業利益率	銷貨毛利率	稅前淨利率	流動比率	速動比率	現金流量比率	負債比率	長期資金適合率
2314 台揚	2014-03-31	-0.072	-21.550	-2.110	-8.860	10.650	-9.780	115.110	67.330	-2.550	63.170	260.160
2317 鴻海	2014-03-31	0.024	1.210	8.370	2.370	6.040	2.890	145.410	115.440	10.180	59.410	266.570
2330 台積電	2014-03-31	0.053	22.330	7.570	35.430	47.500	35.960	204.070	177.440	54.390	30.800	134.210
2332 友訊	2014-03-31	0.006	-0.050	0.770	1.200	28.160	1.810	182.000	104.600	-5.070	42.570	1069.170
2345 智邦	2014-03-31	0.026	1.130	1.040	0.920	15.670	3.800	197.660	141.730	-3.720	44.600	560.480
2353 宏碁	2014-03-31	0.000	-17.280	-4.450	0.170	8.310	0.190	133.110	101.820	-1.190	68.240	1225.360
2356 英業達	2014-03-31	0.030	6.070	2.390	1.670	4.970	2.320	130.740	103.760	-7.090	68.240	233.860
2368 金像電	2014-03-31	-0.036	0.130	-1.060	-3.500	7.510	-5.300	99.360	73.040	0.120	71.500	109.640
2382 廣達	2014-03-31	0.035	4.150	4.780	1.340	4.050	2.850	129.000	101.260	4.190	75.000	354.750
2383 台光電	2014-03-31	0.035	5.110	2.860	7.180	15.800	7.000	143.300	115.310	12.720	50.910	157.750

圖 9 匯入 Excel 資料集

(2) 使用K-mean 集群分析：使用K-mean 集群分析，把所有公司分為兩群，來判斷公司是否有營餘，如圖10所示。

公司	年月	股東權益報酬率	總資產成長率	常績性EPS	營業利益率	銷貨毛利率	稅前淨利率	流動...	速動...	現金流量比率	負債...	長期資金適合率	\$KM-K-Means
1	2314 台揚	2014-03-31	-0.072	-21.550	-2.110	-8.860	10.650	-9.780	115.110	67.330	-2.550	63.170	260.160 叢集-1
2	2317 鴻海	2014-03-31	0.024	1.210	8.370	2.370	6.040	2.890	145.410	115.440	10.180	59.410	266.570 叢集-1
3	2330 台積電	2014-03-31	0.053	22.330	7.570	35.430	47.500	35.960	204.070	177.440	54.390	30.800	134.210 叢集-2
4	2332 友訊	2014-03-31	0.006	-0.050	0.770	1.200	28.160	1.810	182.000	104.600	-5.070	42.570	1069.170 叢集-1
5	2345 智邦	2014-03-31	0.026	1.130	1.040	0.920	15.670	3.800	197.660	141.730	-3.720	44.600	560.480 叢集-1
6	2353 宏碁	2014-03-31	0.000	-17.280	-4.450	0.170	8.310	0.190	133.110	101.820	-1.190	68.240	1225.360 叢集-1
7	2356 英業達	2014-03-31	0.030	6.070	2.390	1.670	4.970	2.320	130.740	103.760	-7.090	68.240	233.860 叢集-1
8	2368 金像電	2014-03-31	-0.036	0.130	-1.060	-3.500	7.510	-5.300	99.360	73.040	0.120	71.500	109.640 叢集-1
9	2382 廣達	2014-03-31	0.035	4.150	4.780	1.340	4.050	2.850	129.000	101.260	4.190	75.000	354.750 叢集-1
10	2383 台光電	2014-03-31	0.035	5.110	2.860	7.180	15.800	7.000	143.300	115.310	12.720	50.910	157.750 叢集-1
11	2395 研華	2014-03-31	0.055	12.720	7.740	16.760	41.480	17.610	229.000	159.300	8.480	28.100	258.890 叢集-2
12	2454 聯發科	2014-03-31	0.044	35.920	22.050	23.500	48.310	25.000	307.700	279.800	13.340	23.500	1546.880 叢集-2
13	2455 全新	2014-03-31	0.021	-7.090	1.630	20.180	31.950	22.680	528.100	436.300	25.780	9.230	199.240 叢集-2
14	2485 兆赫	2014-03-31	0.025	5.600	2.430	4.270	18.170	11.710	388.300	291.100	2.870	22.400	486.150 叢集-1
15	3016 嘉品	2014-03-31	0.019	0.710	-0.060	0.760	12.980	10.620	209.400	173.600	6.620	28.200	192.520 叢集-1
16	3037 欣興	2014-03-31	0.001	-0.610	0.750	0.680	9.160	0.620	153.800	130.000	5.230	48.400	147.400 叢集-1
17	3045 台灣大	2014-03-31	0.065	36.980	6.360	18.430	35.110	18.320	57.520	48.550	13.790	54.900	204.360 叢集-1
18	3062 建漢	2014-03-31	0.004	12.960	0.350	0.250	5.700	0.680	183.100	95.370	12.950	40.600	1347.130 叢集-1
19	3105 禮懋	2014-03-31	0.014	-14.080	1.560	12.210	30.000	15.000	285.300	230.600	24.690	28.300	152.390 叢集-2
20	3152 璟德	2014-03-31	0.037	8.430	7.020	37.080	51.590	42.930	940.800	901.900	81.560	7.720	329.330 叢集-2
21	3163 波若威	2014-03-31	0.043	49.300	4.720	7.980	19.770	13.690	250.500	148.100	-4.480	36.700	315.110 叢集-1

圖 10 K-mean 集群分析

(3) 設定分割區：設定訓練分割區大小為67%和測試分割區大小為33%，如圖11所示。



圖 11 設定分割區

(4) 設定欄位：先設定各個欄位測量的型別(例如連續、種類、旗標…)，再進行讀取每個欄位的值，最後再設定每個財務指標變數為輸入變數，如圖 12 所示。



圖 12 設定欄位

(5) 建立 C5.0 決策數模組：建立決策樹 C5.0 的模組，在設定規則，如圖 13 所示。

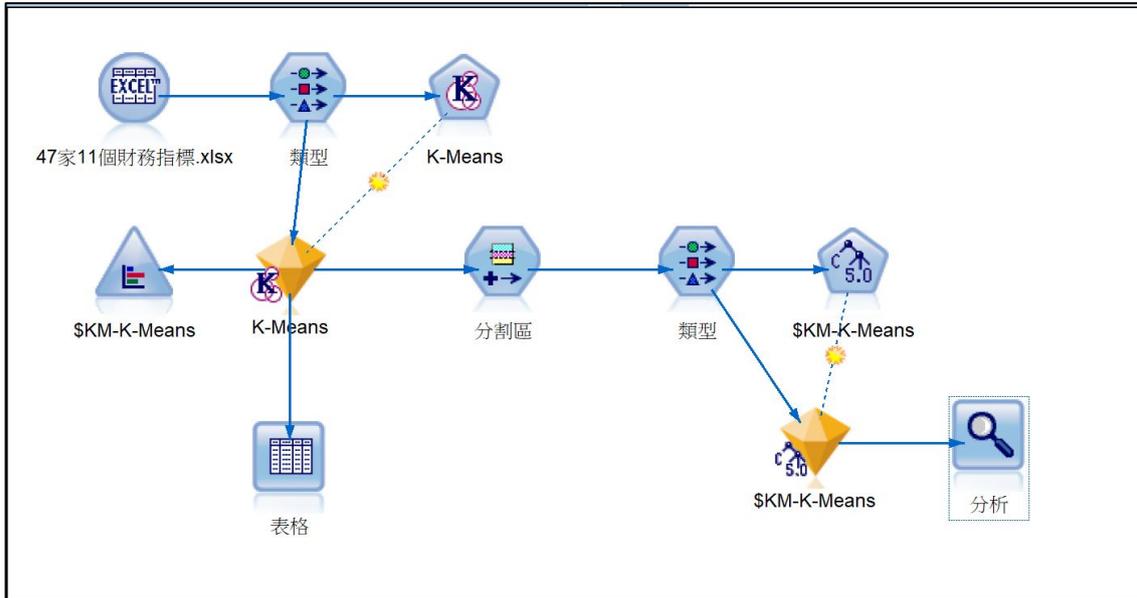


圖 13 建立 C5.0 決策數模組

(6) 實驗結果：使用決策樹 C5.0 來預測變數的重要性，最終得出 6 個重要的變數，以「銷貨毛利率」影響最大，速動比率、負債比率、總資產成長率、股東權益報酬率和現金流量比率次等影響，還有其它 5 個變數，因為影響並不大，在最終的實驗結果上並不會特別顯現出來，如圖 14 所示。

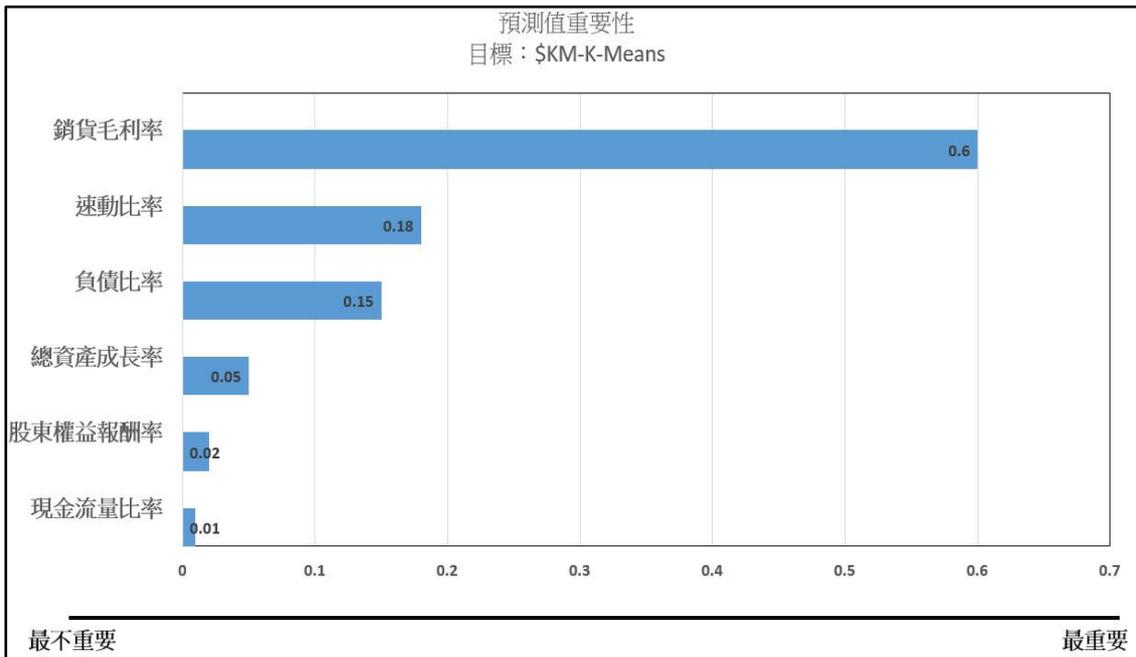


圖 14 決策樹實驗結果

## 4.2 類神經網路模型

以決策樹篩選後得出更具影響力的 6 項財務指標，刪除重要性較低的財務指標後，5G 類股資料集共計為 1128 筆資料，利用 K-means 將資料庫分類出營餘與非營餘公司，再將資料分割為 67% 訓練資料與 33% 測試資料，將決策樹篩選過後的 6 項關鍵因素輸入類神經網路模型內，類神經網路模型設定為（多層感知器 MLP），隱藏層設定為一層。類神經網路模型輸入變數如表 5

表 5 類神經網路模型輸入變數

銷貨毛利率
速動比率
負債比率
總資產成長率
股東權益報酬率
現金流量比率

本研究實驗顯示該模型精確度為 96%。類神經網路模型摘要如圖 15 所示。

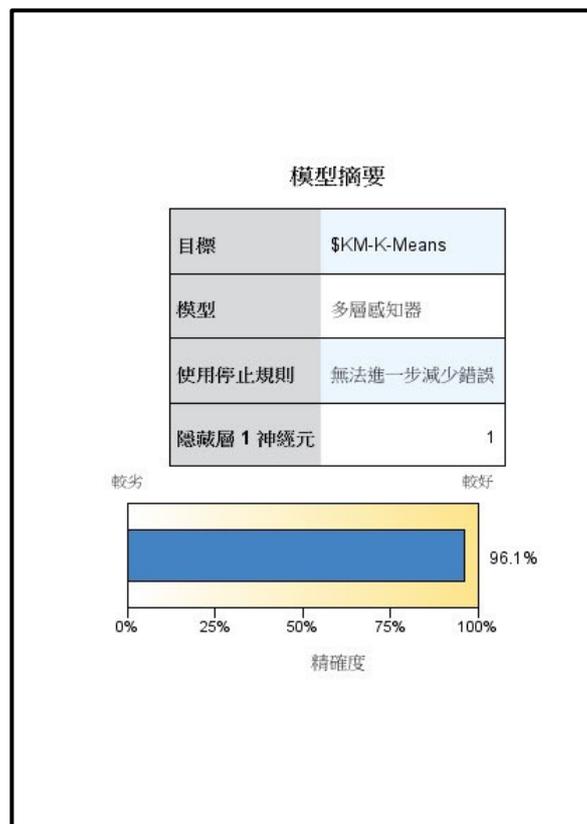


圖 15 類神經網路模型摘要

本研究將類神經網路實驗研究成果變數整理成表 6

表 6 類神經網路實驗成果

變數名稱	變數重要性
股東權益報酬率	0.33
現金流量比率	0.30
速動比率	0.17
銷貨毛利率	0.13
負債比率	0.05
總資產成長率	0.02

\* 變數趨近於 1 為重要；趨近於 0 為非重要

本研究選用 6 項財務指標，資料共 1128 筆，利用 K-means 將資料庫分類出營餘與非營餘公司，再將資料分割為 67%訓練資料與 33%測試資料，將影響投資報酬率之 6 項關鍵因素輸入模型，將模型設為多層感知器，隱藏層設為 1 層。經過類神經模型所得出的研究成果其準確度達 96%，故實驗證實使用 6 項財務指標，在類神經網路模型中可看出預測股價效果良好。

## 第五章 結論

本研究採用 11 項財務指標作為研究樣本，使用決策樹 C5.0 來預測變數的重要性，最終得出 6 個對營業利益較顯著的重要變數，以「銷貨毛利率」影響最大，速動比率、負債比率、總資產成長率、股東權益報酬率和現金流量比率為次等影響。

本研究以 C5.0 決策樹分析出影響股票投資報酬率之關鍵因素，再利用類神經網路進行實驗，找出最具影響力的變數，分析出這些影響投資報酬率之關鍵因素，希望能讓投資人更能掌握股價漲跌的趨勢，有利於投資人有效的投資。

藉由類神經網路模型進行實驗，結果其準確度為 96%，故實驗可知，透過類神經網路學習以及選取適合的基本指標，可對投資人提供一個良好的投資環境，在選擇高投資報酬率的股票外，同時也能降低不少風險，讓投資人多一種選股的方式可以參考。

投資人主要關注的顯然還是股票報酬率的高低，希望未來有更多的相關研究論文，就能讓投資人在選擇買賣股票時能夠有更好的選擇，也期望關於類神經的應用會越來越普及，讓一般大眾都能使用此方法來選擇投資策略。

## 參考文獻

- [1] 許銘源，利用資料探勘技術於連鎖超商之營業額分析，南台科技大學資訊管理研究所碩士學位論文，2013年。
- [2] 張道行，使用類神經網路以短期技術指標作次日股價趨勢預測之研究，國立交通大學資訊科學研究所論文研究，1998年。
- [3] 黃世勇，應用資料探勘技術於動物醫院病例系統之研究，世新大學資訊管理學系碩士學位論文，2012年。
- [4] 王珮紋，利用資料探勘技術建立現金與模式：決策術方法之應用，中正大學會計與資訊科技研究所碩士論文，2012年。
- [5] 黃鈺琳，應用自組織映射圖網路與K-Means於中國大陸股票型基金與QDII基金投資策略之研究，國立臺灣科技大學資訊管理系碩士論文，2010年。
- [6] 吳秉奇，類神經網路在台股指數期貨的指數預測應用，中央資管所碩士論文，1999年。
- [7] 曾思博，類神經網路於股價預測與資金配置之應用，中央資管所碩士論文，1999年。
- [8] 故國瑜，類神經網路產業盈餘預測及其投資策略之研究-以電子電機及紡織業為例，政治大學資訊管理研究所碩士論文，1996年。
- [9] 黃綺年，統計方法與類神經網路應用於國內開放式股票型基金投資績效分類及投資報酬率預測之研究，國立成功大學統計學研究所碩士論文，2004年。
- [10] 蔡正修，台灣上市電子類股價指數走勢預測之研究，國立成功大學統計學研究所碩士論文，2005年。
- [11] 蘇偉庭，以類神經網路分析財報預測台灣上市公司股價之變動，中國文化大學資訊管理學系碩士論文，2011年。
- [12] 端木和奕，多重彈性倒傳遞類神經模型以台灣4G概念股為例，中華大學資訊管理學系，2015年。
- [13] 簡壬申，類神經網路在股票預測之獲利可能性研究—以台灣50成分股為例，國立雲林科技大學財務金融系碩士論文，2017年。
- [14] 林憲民，運用資料探勘技術預測股價-以蘋果公司產品之供應商為例，嶺東科技大學資訊管理系碩士論文，2018年。
- [15] 黃彥文，資料探勘之應用 - 會員消費特徵之發掘，國立屏東科技大學資訊管理學系碩士論文，2001年。

[16]劉育成、黃子彥、鄭銘峰、楊欣玟、陳逸珊，應用自組織映射圖於科技類股票型基金投資策略之研究，嶺東科技大學，2017年。

[17]Duda， R. O.， Hart， P. E.， & Stork， D. G.(2001) - *Pattern classification*. New York: Wiley.

[18]W.C .Chiang， Urban， T.L.and Baldrige， G.W.(1995)， *A Neural Network Approach to Mutual Fund Net Aasset Value Forecasting*， Omega， Int.J. Mgmt. Sci.Vol.24， No.2， pp.205-210.

[19]K .Bergerson， and D.C.Wunsch (1991)， *A Commodity Trading Model Based on a Neural Network-expert System Hybrid*， IJCNN-91， I， pp.289-293。

109  
學年度

嶺東科技大學

資訊管理系

應用類神經網路探討影響CQ類股投資報酬率之關鍵因素