

東海大學資訊管理研究所
碩士學位論文

以資料探勘分析上市電子零組件公司之經營績效
**Using Data Mining to Analyze the Performance of
Listed Electronic Components Companies**

指導教授：姜自強 博士
研 究 生：李佳鴻 撰

中華民國 107 年 6 月

東海大學資訊管理學系碩士學位
考試委員審定書

資訊管理學系研究所 李佳鴻 君所提之論文

以資料探勘分析上市電子零組件公司之經營績效

經本考試委員會審查，符合碩士資格標準。

學位考試委員會 召集人：陳好言 (簽章)

委員：鄭永真

楊朝棟

姜爾強

呂峻益

中華民國 107 年 6 月 19 日

謝誌

研究生兩年的生活，首先要感謝的就是指導教授—姜自強博士，不論是在程式語言與撰寫論文的觀念，也感謝老師讓學生有自由發揮的空間，在學生遇到困難或疑惑時老師會細心且不遺餘力的教導，並指出正確的方向及相關意見參考，讓學生可以自主學習並且提供許多想法上的啟發。

在研究室裡，要感謝的是同一位指導教授的復榮、昕穎同學，不論是在報告的討論、論文的準備及準備口試的過程中的幫忙，也要感謝研究所的班代家廣、映瑋、彥齊、俊昭、岳穎同學，在平時的討論與想法的激盪上助益良多；另外要特別感謝 EMBA 的慧芬學姐及 Lily 的照顧，讓我在工讀期間學習到許多東西及未來生涯規劃的建議，獲益良多。

感謝我的家人—父親-李文棋和母親-黃麥，在背後的支持與鼓勵，讓我在人生的道路上有更多不同的抉擇方向，另外要感謝兩位姐姐—大姐-韻慈、二姐-佳柔，特別感謝二姐在撰寫論文時提供的建議與意見。

還有在以前在工作職場的朋友皓爺和柚子，在生活上的一些瑣事的聊天紓壓與陪伴，也非常感謝系上的助教藝音學姐、祐陞學長，解決我們平時的疑難和困惑，並提供許多學習的資源與資訊供我們運用。

最後，感謝幫助我求學道路上給予幫助與鼓勵的人，讓我能順利完成碩士學位，謝謝你們。

李佳鴻 謹誌於
東海大學資訊管學系
2018 年 06 月

論文名稱：以資料探勘分析上市電子零組件公司之經營績效

校所名稱：東海大學資訊管理學系研究所

畢業時間：民國 107 年 06 月

研究生：李佳鴻

指導教授：姜自強

論文摘要：

隨著金融市場的觀念與制度的不斷演進，一般民眾利用工商報章雜誌、證券商提供的相關財務資料、著名投資理財專家的建議等等，以此參考並且決定投資標的已成為日常的行為，在眾多的財務資料項目之中，普遍民眾對於影響公司獲利能力的關鍵財務項目卻是模稜兩可，甚至是顛倒黑白的情況。

因此，本研究利用台灣經濟新報資料庫 (TEJ) 之中的財報資料，希望藉由資料探勘的方法來找出適合投資的標的；本研究採用 2007 至 2016 年的財報資料為分析的期間，並以其中的經營能力作為投入、獲利能力作為產出，計算出資料包絡分析法(DEA)的值並以此作為主要目標參數，找出公司財務體質較佳的公司，最後再進行決策樹、分群分析等等資料探勘方法，再更進一步篩選，以期找到最佳的投資標的公司來獲利。

關鍵字：資料探勘(Data Mining)、資料包絡分析法(Data Envelopment Analysis)、決策樹(Decision tree)、分群分析(K-Means)、時間序列分析

Title of Thesis : Using Data Mining to Analyze the Performance of Listed Electronic Components Companies

Name of Institute : Tunghai University, Institute of Information Management

Graduation Time : (06/2018)

Student Name : Chia-Hung Li

Advisor Name : Tzu-Chiang Chiang

Abstract :

With the continuous evolution of the concept and system of the financial market, the general public takes advantage of the relevant financial information provided by the business newspaper and magazine, the securities dealers, the suggestions of the famous investment financial experts and so on, which can be used as a reference and decides that the investment standard has become a daily behavior. Among the numerous financial information items, the general public is in the shadow of the shadow. The key financial items that ring the company's profitability are ambiguous or even black and white.

Therefore, this study uses the financial data from the Taiwan economic new newspaper database (TEJ), and hopes to find the target of investment by means of data exploration. This study uses the financial data from 2007 to 2016 as the analysis period, and calculates the data with its operational capability as input and profitability as output. The value of envelope analysis (DEA) is used as the main target parameter to find out the company with better financial constitution. Finally, the data exploration methods such as decision tree, group analysis and so on are carried out, and further screening is made to find the best company to make a profit.

Keywords: Data Mining, Data Envelopment Analysis, Decision tree, K-Means, Time series analysis

目 錄

頁次

謝 誌	
目 錄	III
圖 目 錄	V
表 目 錄	VII
第一章 緒論	1
第一節 研究背景與動機	1
第二節 研究目的	3
第三節 研究流程	4
第四節 論文架構	5
第二章 文獻探討	6
第一節 資料探勘(Data Mining)	6
壹、資料探勘之定義	6
貳、決策樹(Decision tree)	9
參、群集分析	11
肆、時間序列分析	12
第二節 台灣經濟新報資料庫(TEJ)	13
壹、台灣經濟新報資料庫(TEJ)之介紹	13
貳、利用台灣經濟新報資料庫(TEJ)相關之研究	14
第三節 資料包絡分析法(DEA)	16
壹、資料包絡分析法(DEA)之定義	16
貳、資料包絡分析法(DEA)之 CCR 模式	16
參、資料包絡分析法(DEA)之 BCC 模式	17
肆、利用資料包絡分析法(DEA)相關之研究	18
第四節 相關分析(Correlation analysis)	20
壹、相關分析之定義	20
貳、相關分析相關之研究	20
第三章 研究方法	22
第一節 研究工具及軟體介紹	22
第二節 研究設計及研究流程	22
第三節 資料來源	22
第四節 分析變數的選取—投入及產出選項	24
壹、經營能力	24
貳、獲利能力	25
第四章 研究結果與分析	26
第一節 資料包絡分析法結果與說明	26

第二節	決策樹分析結果與說明	30
第三節	群集分析結果與說明	34
第四節	時間序列分析結果與說明	37
第五節	相關分析結果與說明	41
第五章	結論與研究限制	42



圖 目 錄

	頁次
圖 1-1 台灣股票市場各類股成交比重.....	2
圖 1-2 研究流程圖.....	4
圖 2-1 KDD 流程圖.....	7
圖 2-2 Farrell 模式效率分析圖.....	16
圖 3-1 研究設計及流程圖.....	22
圖 3-2 台灣上市電子零組件產業匯整圖.....	23
圖 4-1 2007~2016 年 DEA 排名前 10 位的廠商.....	27
圖 4-2 2007~2016 年 DEA 排名前 10 位的廠商.....	28
圖 4-3 2007~2016 年 DEA 排名前 10 位的廠商.....	28
圖 4-4 2007~2016 年 DEA 排名前 10 位的廠商.....	29
圖 4-5 2008 年決策樹分析圖.....	30
圖 4-6 2010 年決策樹分析圖.....	31
圖 4-7 2016 年決策樹分析圖.....	32
圖 4-8 產出與投入全要素之決策樹分析結果.....	33
圖 4-9 產出要素之決策樹分析結果.....	33
圖 4-10 2008 年最佳的分群數目.....	34
圖 4-11 2010 年最佳的分群數目.....	34
圖 4-12 2016 年最佳的分群數目.....	34
圖 4-13 2008 年分群分析圖.....	35
圖 4-14 2010 年分群分析圖.....	35
圖 4-15 2016 年分群分析圖.....	35
圖 4-16 2008 年分群分析圖.....	36
圖 4-17 2010 年分群分析圖.....	36
圖 4-18 川湖歷年 DEA 時間分析圖.....	37
圖 4-19 川湖 Holt 預測分析圖.....	37
圖 4-20 川湖 HoltWinters 預測未來四年分析圖.....	37
圖 4-21 新日興歷年 DEA 時間分析圖.....	38

圖 4-22	新日興 Holt 預測分析圖.....	38
圖 4-23	新日興 HoltWinters 預測未來四年分析圖	38
圖 4-24	達邁歷年 DEA 時間分析圖	39
圖 4-25	達邁 Holt 預測分析圖.....	39
圖 4-26	達邁 HoltWinters 預測未來四年分析圖	39
圖 4-27	鑫禾歷年 DEA 時間分析圖	40
圖 4-28	鑫禾 Holt 預測分析圖.....	40
圖 4-29	鑫禾 HoltWinters 預測未來四年分析圖	40



表 目 錄

	頁次
表 2-1 台灣經濟新報資料庫(TEJ)主要資料提供項目	13
表 3-1 經營能力之財務變數敘述性統計	24
表 3-2 獲利能力之財務變數敘述性統計	25
表 4-1 2007—2016 年 DEA 排名前 10 位的廠商	26



第一章 緒論

第一節 研究背景與動機

繼 2007 年金融危機之後，全球經濟景氣持續低迷，而在近年來國內經濟環境不景氣、定存的利息極低，各種的負面因素影響下，國人的痛苦指數只升不降，俗話說：「什麼都漲，只有員工的薪水不漲」，在這樣大環境影響之下，使得國人對於投資理財的觀念更加的重視，如何利用自身有限的資源，利用投資來達到個人資產價值極大化，變成人們關注的焦點。隨著金融環境改變與制度的改革，股票市場資訊也隨之更加資訊透明化，而其中股票投資是國人熱衷投資的標的之一，過去學者也指出，可以藉由公司的財務指標與相關資訊的分析，來投資獲利(Abarbanell and Bushee,1997)。

隨著金融市場的不斷演進，現今對上市上櫃公司的財務資訊取得也相當便利；然而上市上櫃公司之多，相關資訊資料龐大，如何採用所資料來進行分析是非常重要的。國內外已經有許多學者對於股票投資、分析及投資策略等等相關的研究，如基本面分析法、技術面分析法、訊息面分析法、籌碼面分析法...等等，該持有價值型的股票或是成長型的股票，短期持有，還是長期持有才能夠獲得利潤報酬。

本研究希望藉由資料探勘與資料包絡分析法(Data Envelopment Analysis ,DEA)的搭配，找出在上市股票中適合投資的公司，希望能夠藉由長期持有價值型的公司股票或是被低估的公司股票等。

其中，DEA 分析法可以同時衡量多項投入與產出因素並計算出相對效率的方法，其分析出的數據可供決策者改善經營績效(Chandra et al.,1978)，換言之，也有利於一般投資大眾在股票市場作為選擇股票的參考。

根據台灣證券交易所統計顯示，台灣股票市場各類股成交比重，在其中所有的股票類別產業當中，以電子股交易成交比重最大，整理如下圖所示：

台灣股票市場各類股成交比重，如下圖 1-1 所示

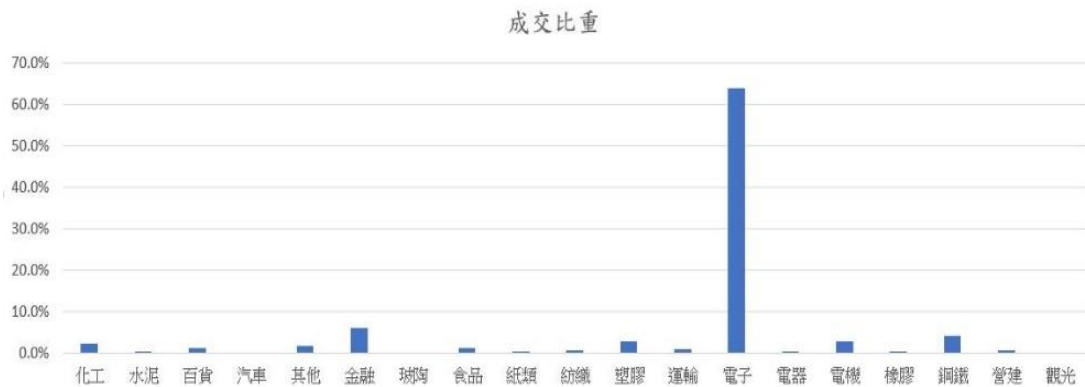


圖 1-1 台灣股票市場各類股成交比重

資料來源：台灣證券交易所

而在台灣上市電子類股產業，其中以電子零組件類別的廠商數量最多(林明璿,2017)，我們希望藉由股票市場成交比重最大類別及該類別之中廠商數量最多的產業類別，其影響力對於實際股票市場也尤為重要；另外，從 1986 年以來，台灣的股票市場中的上市公司數量、市值總額、成交股數等至今不斷增加，由此可見，國人對於股票市場投資的參與度日益提升，而股票市場之中，上市的門檻標準比上櫃更加嚴格，其成立時間也較為悠久、資本額龐大、獲利能力較佳等因素考量之下，因此將研究對象設定為上市電子零組件公司。

綜合上述，本研究嘗試利用台灣經濟新報資料庫(TEJ)2007 年至 2016 年台灣上市電子零組件公司的資料，來衡量經營績效分析，進一步探討與實際市場股價、報酬率之關聯，以期提供投資大眾相關資訊，並且作為選擇股票的參考。

第二節 研究目的

根據上述的研究背景與動機，本研究利用 TEJ 資料庫之中，研究範圍是從 2007 年至 2016 年總計 91 家台灣上市電子零組件公司的財務資料。藉由 DEA 分析法的計算，預期找出最佳的經營績效好的公司，並且以此分析結果為基礎，結合資料探勘中幾個常見的分析方法，在這 91 家上市電子零組件公司，做更進一步的統計分析。

本研究欲探討的目的整理如下：

1. 針對台灣 91 家上市電子零組件公司，利用資料包絡分析法，其分析出來的結果，提供公司經營者或高階主管瞭解並改善公司體質的參考，或是可以用來當作社會大眾股票選擇的一項指標。
2. 以資料包絡分析法其分析出來的結果，作為資料探勘分析方法的主要參數來進行後續的統計分析。
3. 利用資料探勘中幾個常見的分析方法，預測或求出適合社會大眾投資的公司。
4. 希望可以針對全球金融大事件的幾個年份，如 2008 年金融海嘯、2010 年歐債危機及 2016 年黑天鵝事件頻傳的一年，找出上市公司經營績效好的公司。
5. 希望可以藉由資料探勘分析方法，篩選出公司體質佳的上市公司，長期持有其股份並且獲利。

第三節 研究流程

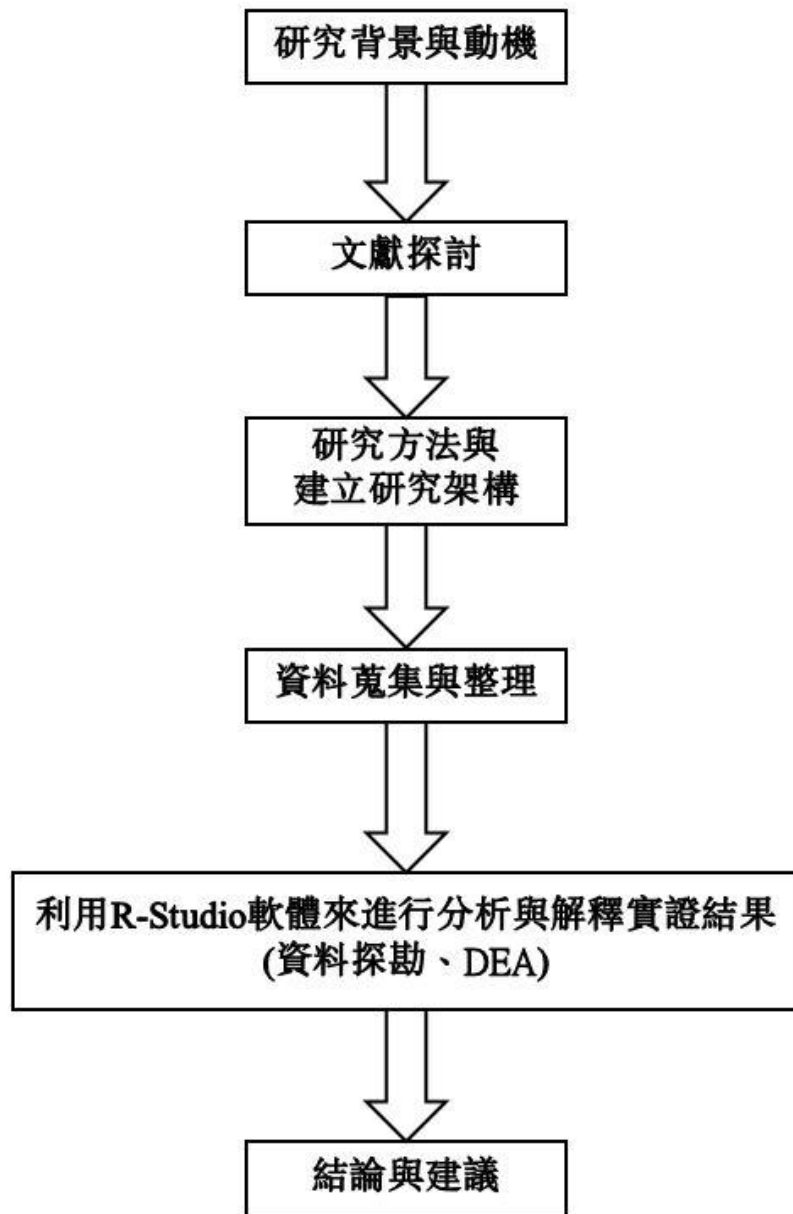


圖 1-2 研究流程圖

第四節 論文架構

本研究論文內容分為緒論、文獻探討、研究方法、研究結果與分析、結論及研究限制等總計五個章節，各章節內容簡要描述如下：

第一章：緒論

說明本研究之研究背景與動機、研究目的及研究流程。

第二章：文獻探討

介紹國內外與資料探勘、資料包絡分析法(DEA)、台灣經濟新報資料庫(TEJ)、決策樹分析、分群分析等相關研究期刊論文與書籍，並瞭解實務分析方法的應用與實證結果，作為本研究欲探討問題之依據。

第三章：研究方法

針對本研究採用的分析軟體工具、研究的設計與流程、資料探勘之中的統計分析方法、資料來源等作詳細的說明、解釋。

第四章：研究結果與分析

詳細描述、解釋各分析方法之輸出結果，並給予簡單扼要的表格或視覺化圖形的呈現。

第五章：結論及研究限制

根據研究分析結果，彙整並提出結論、建議及未來可以改善的方向，供後續相關研究之學者作為參考依據。

第二章 文獻探討

第一節 資料探勘(Data Mining)

壹、資料探勘之定義

在現今資訊爆炸的時代，資料探勘(Data Mining)是近年來非常熱門的詞彙，也是大眾熱衷的分析技術之一，國內外許多學者針對這個名詞有不同的看法與解釋，Hand et al.(2007)認為資料探勘一般都是對大型的數據資料集來進行分析來發現其未知的關係與規則，並且以一種比較新穎的方式來呈現數據結果，這對數據的擁有者來說是有用且可以清楚理解的。Frawley et al.(1991)認為，資料探勘是從資料庫挖掘出隱藏其中的訊息，所以也可以說是發現知識(Knowledge discovery)的過程，利用機器學習的方式，找出有用的規則、模式(Patterns)並結合該領域相關術語的綜合解釋，來判斷是否存在關聯性。在發現知識的過程當中，又以資料是否具有一定的規則、有意義的結果呈現及可靠度高的資料來源等要素對資料探勘的運用成功與否具有重大影響。在 Curt (1995)的研究中，資料探勘被定義為是資料轉換的過程，從一開始沒有規律的文字與數字的叢集資料轉變成資訊，進而轉變成知識，最後產生出有意義、有效益及具有相關的決策。Berry&Linoff (1997)認為資料探勘是從資料倉儲、資料庫等其他資料儲存空間之中，來尋找知識的過程。在 Fayyad et al.(1996)研究中指出，資料探勘是 KDD 過程中的一個步驟，其中包括了應用數據分析和發現演算方法，預期能夠在數據上產生特定的模式、模型等等。Liao & Chen (2004)認為資料探勘是建立不同類別的資料庫，從中擷取具有關聯性的資料庫來做更進一步的數據分析，預期能產出有用的成果，其有用的資訊甚至能夠以視覺化圖表的方式呈現，並協助管理者做出決策。

綜合上述各學者的說法，大多數對於資料探勘的解釋與分析的流程、方法都是相似的，普遍都有指出是藉由大型的數據資料集、資料庫或其他資訊儲存庫之中，利用現今既有的技術，如機器學習、統計分析、視覺化呈現、高度運算等方法，希望可以取得潛在有用、事前未知且有趣的成果，用現今時下資料科學家、資料分析師經常講的—「大數據分析」，但本研究的資料數據量並不符合這樣的說法，比較傾向擷取具有特徵或關聯性的資料，以便進行後續資料探勘的分析與研究。

本研究的資料探勘流程是採用 Fayyad et al.(1996) 提出的 KDD(Knowledge discovery in databases)方法，其流程與步驟如下圖 2-1 所示：

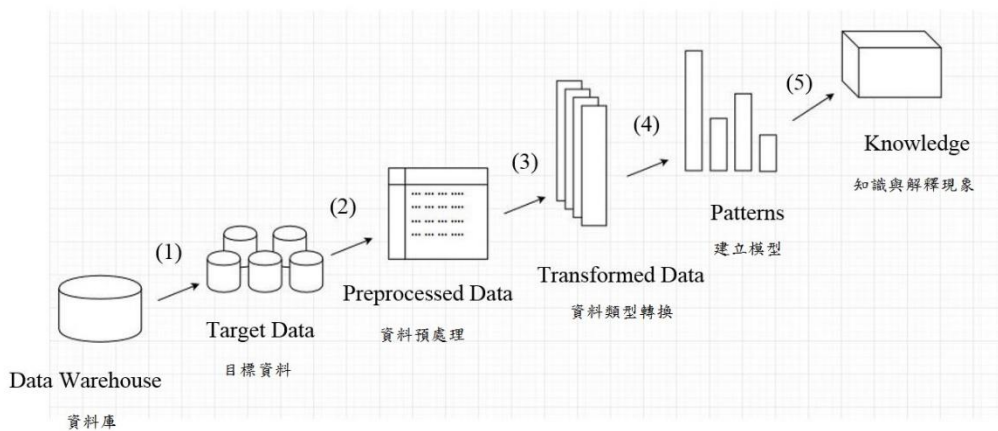


圖 2-1 KDD 流程圖

資料來源：Frawley,W.J.,Piatetsky-Shapiro,G.,and Matheus,C.J(1992).Knowledge discovery in database: An overview In Knowledge Discovery in Databases:1-27. Cambridge, MA:AAAI/MIT Press,Reprinted in AI Magazine,13(3)

本研究依照 KDD 方法的步驟有以下幾個重要的階段，分別為資料的選取、目標資料的選定、資料預處理、資料轉換、樣本或模型的評估以及最後的，成果或知識的呈現，各階段細部的內容簡要描述如下：

第一階段：資料選擇與抽樣

針對本研究欲研究的問題，從大型資料庫中，抽取其中部份資料來進行分析，找尋規則或建立模型等。

第二階段：資料預處理

資料庫中的資料，可能會有錯誤、遺失值、不完整(亦即空值)等情況，必須刪除或填補缺失空白的位址。

第三階段：資料轉換

此階段最重要的動作就是資料整合，資料的維度或格式必須保持一致性，以便後續的統計分析。

第四階段：資料探勘

運用資料探勘的技術，針對有趣的特徵或有用的資料等，來萃取資料並以視覺化的方式呈現結果，如決策樹分析、分群分析、時間序列分析等其他演算方法。

第五階段：資料分析解釋與評估

根據資料探勘的各種統計分析結果，彙整並且採用圖形或表格等方法，轉變這些資料探勘的成果，讓人可以簡單明瞭的看出結果，也可以提供管理者做決策使用；另外也可以針對本研究欲研究目的、問題，配合資料探勘的分析結果作解析並瞭解是否能達到預期的目標。



貳、決策樹(Decision tree)

一、決策樹之定義

決策樹是資料探勘方之中最常見且使用的方法，然而決策樹的演算法也不斷進化，這些演算法的基礎通常是貪婪演算法如 ID3、C4.5、C5.0 演算法等，但還是有使用不同理論及統計方法作為其演算法的基礎如 CHAID、CART 演算法等，其中本研究使用 R 語言中的 rpart 套件就是以 CART 演算法為基礎，以下為各演算法的公式與介紹：

(一) ID3

這個演算法是由 Quinlan (1986) 所提出，它是以熵 (Entropy) 為衡量的標準，以期取得最大增益資訊值 (Information gain) 來當作分類的準則，其公式如下：

$$\text{Entropy}(K) = \sum_{i=1}^m p_i * \log_2 p_i \quad (2-1)$$

K：資料集合 K 之中存在 m 個不同的類別

p_i ：類別 i 出現在資料集 K 的機率

$$\text{Gain}(K, L) = \text{Entropy}(K) - \sum_{j=1}^n \frac{|K_j|}{|K|} * \text{Entropy}(K_j) \quad (2-2)$$

假設類別 L 之中存在有 n 個不同的數值，而資料集合 K 藉由這些不同值而分割出 n 個不同的資料子集合。其中 $\text{Entropy}(K)$ 為資料集合 K 的資料雜亂程度，另外 $\text{Entropy}(K_j)$ 為資料子集合 K 的資料雜亂程度， $\text{Gain}(K, L)$ 意指類別 L 對於資料集合 K 產生的增益資訊值，Gain 的數值越大代表類別 L 的資料雜亂程度越小，是最佳用來分類的選擇。

(二) C4.5

C4.5 演算法可以說是 ID3 演算法所延伸改進的，它是利用增益比率 (Gain Ratio) 作為分類的準則，也改善了 ID3 演算法中無法處理連續型資料的問題。

(三) C5.0

C5.0 演算法可以說是 C4.5 演算法的商業版，一樣是依照增益比率 (Gain Ratio) 作為分類準則，採用拔靴 (Boosting) 的方法來提升其模型的準確率，使得其運算的速度變的比較快。

(四) CART&CHAID

CART 是採用 Gini 係數，而 CHAID 是採用卡方檢定法，這兩個演算法利用

的基礎劃分節點的準則不一樣，CART 在分割節點只會產生兩個分支，為二元法切割，而 CHAID 是依據卡方檢定分析出來的 P 值作為決策樹是否繼續生成，在資料處理的屬性方面，CHAID 只能處理類別型的資料型態，而 CART 則是改善 ID3 無法處理的部分，它能夠處理離散型與連續型的資料型態，適用的範圍更加廣闊。

二、決策樹相關之研究

施正宏(2004)探討希望藉由資料探勘中的決策樹分析方法，搭配台灣總體經濟相關指標和此研究之產業相關個股財報等資料，以期找出股價漲跌之特徵、預測未來台灣股票市場股價趨勢走向及整理出一些指標提供公司管理者或投資大眾作為投資選股之參考，此研究領域是選取佔台灣股票成交比重約七成的電子類股，並利用 IBM 的統計軟體 IM 作為決策樹的分析工具，並以 CART 為演算法來建立模型，此篇研究結果發現總體經濟相關指標與產業相關財務報表資料的結合提高了預測的準確率及投資報酬率，並發現與電子類股相關、重要的總體經濟指標為工業生產指數。

林志杰(2006)探討 2002~2005 年林口長庚醫院中風病人，希望藉由資料探勘的決策樹、分群、分類等分析方法，以期找出中風病患的特徵規則及中醫與西醫辨症之差異，此研究的資料特性它是採用 C4.5 的演算法來建立模型，而研究結果發現中醫與西醫各一套辨症的規則及中西醫的辨症標準結合可以提升確診的判斷率，為中風症狀提供另類的觀點、預測方法，造福未來醫學臨床的參考準則並提供更有效率的診療。

鄭忠樑(2002)探討以 1995~2001 年電子類股的財務報表作為研究標的，然而早期決策樹中的 CART 演算法大部份都是應用在醫療科學領域，此研究希望可以把此種方法應用到台灣股票市場股價分析、預測及建模，其研究結果發現，淨利率與稅前淨利率此兩項財務要素對公司的報酬率影響最為重要，其中利用 CART 建立的股價報酬率的預測模型具有一定的成效，並且提供兩個部分及不同投資組合規則的方案，給予大眾在股票投資上選擇之參考。

參、群集分析

一、群集分析之定義

群集分析 (Cluster analysis)，用一般淺顯易懂的話來形容，就好比是物以類聚，在一個大組織之中，群體與群體之間的差異，群體之內同質性，以這樣的概念來進行分群；其中它又分為階層式與非階層式(亦稱為分割式)兩種分析方法，其對應的演算法也不同，其適用時機也不一樣，前項是適用於分群數量未知、樣本數量少等情況，而後項是適用於分群數量已知、樣本數量多等情況，階層式分析方法之中，最常使用的是 Ward's Method (1963)，中文譯稱為華德法，也有最小變異數法等別稱，它是由平方差的增減量來定義兩個群集間的鄰近值；而非階層式分析方法之中，最早由 MacQueen(1967)所提出的 K-means 演算法，其分析的中心點通常為一群資料點的平均，並以它來計算與每一個資料點的誤差，而最小誤差的平方的群集更能代表這個分群群集的中心資料點，其中公式如下：

$$\text{Min. SSE} = \sum_{i=1}^k \sum_{x \in C_i} \text{dist}(c_i, x)^2 \quad (2-3)$$

第 i 群的中心點公式如下：

$$c_i = (1/m_i) * \sum_{x \in C_i} x \quad (2-4)$$

x：一項物件

C_i ：第 i 個群集

c_i ：群集 C_i 的群集中心

c：所有點的群集中心

m_i ：第 i 個群集中物件的數量

m：資料集中物件的數量

k：群集的數量

dist：兩個資料之間的歐幾里德距離

二、集群分析相關之研究

郭宜易(2017)是探討 PCHome 線上購物資料並利用資料探勘中的集群分析來了解記憶卡的歷史與未來的趨勢，並比較不同集群探勘方法間的差異—K-Means 和 Agglomerative，其研究結果顯示，在記憶卡品牌、價格與類別、效能及觀覽人數上的分群是效果顯著的，而且不同的分群分析方法，前項分析方法分四群與

後項分析方法分為三群其市場的區分較佳，利用這樣的方法也可以提供相關領域廠商對於關鍵商品屬性的開發與推廣的一項參考。

曾惠敏(2015)探討汽車百貨業個案研究並利用探勘方法中的 K-Means 分群與分類分析及 RFM 分析方法，希望能夠找出目標客群與商品的特徵，以提供廠商參考並建立不同的行銷策略，其研究結果指出，依據探勘的分群方法把顧客分為高與低收益客群，高收益客群特徵為中年且消費地區為中部及會員級別為 E1，由於顧客的資料屬性變數太少，兩個客群間的差異不大，導致分類分析結果不理想，找不出其中關鍵變數。

肆、時間序列分析

一、時間序列分析之定義

資料的產生都有其時間的先後順序，把數據資料以時間先後順序排序形成的數列，希望藉由過去的歷史或即時的資料，找出一定的規律或預測未來的趨勢，而不同的時間間隔如年、月、日或每一小時等等，所呈現出來的規則也會有所不同，時間序列分析應用的領域也是相當廣泛如金融經濟、數學與統計、天氣預測、理工科系領域相關工程學等等，本研究所使用的 R 語言的套件—forecast，其演算法的基礎為 ARIMA 模型，也是時間序列分析方法之一，它是利用差分處理非平穩序列轉為平穩序列，並採用其他時間序列分析的模型來進行估算、預測。

二、時間序列分析相關之研究

王貞治(2014)以 ARIMA、類神經網路等分析模型，比較其不同方法分析及預測，其資料來源為新台幣兌換美元的每日匯率資料，期間為 2009 年 1 月至 2014 年 2 月。研究結果顯示，傳統的計量模型所建立的預測結果比人工智慧的演算法來的佳，但後者的準確度比前者來的佳，而在人工智慧分析方法之中，又以果蠅演算法為最佳的準確率。

翁雅君(2015)以時間序列分析方法中的 ARIMA 分析模型作為分群的依據，並搭配統計檢定的分析，希望改善收費方式及預測模型與周遭環境的關係，其資料來源是台北市交通局提供的信義區 Youbike 的基本數據資料，期間為 2015 年 3 月到 5 月。研究結果發現，上班族與學生是資料來源中最穩定的對象，而分群的

分析發現，停車點的佔多數為公園與學校這兩個場域，但因為地區性的建物環境與型態的不同，停車點為捷運站或火車站的可能性也是不容忽視的。

高建國(102)是以 R 語言的時間序列分析套件來探討台灣個股的波動度並以此分析產出作為圖形判斷，其資料來源為 TEJ 財金資料庫中台灣 50 指數的 16 項主要的上市類股，期間為 2001 年 1 月至 2012 年 12 月。研究結果顯示，其資料都不需經過差分的處理，且圖形的各個類股皆有符合常態分配的曲線。

林碁域(2016)以時間序列分析方法來預測鴻海、佳格與台泥這三家的股價預測分析，比較三家的預測分析準確率且找出其股價趨勢，並搭配統計的檢定方法，找出合適的投資策略組合，其研究期間為 105 年 3 月為期三十天。研究結果指出，時間序列預測分析股價走勢不適用於長時間，以期間間隔短的時間軸為佳，而其中鴻海在股價的績效表現是較佳的，此研究也建議聽取巴菲特的建議，從股票市場的基本面去選股且長期持有才是最佳的投資方式，而不是去預測短期的股價。

第二節 台灣經濟新報資料庫(TEJ)

壹、台灣經濟新報資料庫(TEJ)之介紹

台灣經濟新報資料庫(TEJ)自 1990 年成立至今，是國內著名的金融財經資料庫，其資料庫提供資料服務的國家分別為；中國、香港、新加坡、馬來西亞、菲律賓、泰國、韓國、日本以及台灣共九個國家，並提供幾個主要的資料項目，如下表 2-1 所示：

表 2-1 台灣經濟新報資料庫(TEJ)主要資料提供項目

總體經濟資料	商品原物料資料	企業營運資訊
企業財務資料	證券金融市場	

其中，五項主要的資料項目簡介如下：

總體經濟資料：國民所得及經濟成長、人口及就業、貨幣供給及金融市場、景氣指標、工業生產、零售批發銷售及房市景氣指標、物價水平、財政統計、證券市場交易。

商品原物料資料：各不同領域之商品原物料資料，如塑化物料、百貨相關、資訊電子等 14 種類別，總計 1140 項產銷存及報價資料和 199 項國際重要原物料資料。

企業營運資訊：公司基本資料、資本形成模組、董監事及股東資訊等。

企業財務資料：跨域公司財務比較、連結新舊國際會計準則等總計 890 多項財務資料。

證券金融市場：基金市場、債券市場、股票市場、權證市場、期貨市場、選擇權市場。

貳、利用台灣經濟新報資料庫(TEJ)相關之研究

王尹柔(2014)主要研究目標是採用台灣經濟新報資料庫(TEJ)中的台灣指數選擇權，其研究目的有兩個，第一點是探討台指選擇權市場波動性偏態指標與報酬率是否存在關聯，第二點是探討台指選擇權市場波動性偏態指標其特性是否與投資者本身對投資的風險度以及其投資行為是否存在關聯，是傾向選擇權指標與報酬率、投資人交易行為及對投資的風險態度的相關研究，這篇研究所採用的資料及研究目的雖然與本研究不一樣，但也是採用台灣經濟新報資料庫(TEJ)中所提供的資料做為研究的主要資料來源與分析。

在另一篇探討公司經營績效、企業生命週期及企業社會責任報告中的環境績效指標的研究中提到，它是以台灣地區上市電子公司為研究對象，並且採用台灣經濟新報資料庫(TEJ)中與經營績效衡量相關的財務資料，如固定資產報酬率、營業毛利率、每股盈餘等，其研究成果發現，在不同的企業生命週期與其環境績效指標存在的相關不顯著，但經營績效對環境績效指標存在顯著正相關的，而其中部分資料正相關的程度在不同的企業生命週期中存在顯著差異(陳郁璿，2015)。上述的研究與本研究的對象都是屬於上市電子業領域的，而本研究並無採用企業社會責任報告書中的環境績效指標，但是對於公司經營績效這方面是一樣感興趣的，雖然與本研究的方向不同，但在針對公司經營績效的財務資料參數的選取上，提供許多不同的見解與想法。

陳昭伶(2018)主要探討公司治理、經營績效及財務的關聯性，此篇研究也是採用台灣經濟新報資料庫(TEJ)所提供的公司治理與經營績效相關的財務資料，利

用統計軟體 SAS 進行分析來解釋公司治理與財務變數之間的關係與影響，希望藉此了解公司潛在的危機，並提供公司經營者、高階管理人員或投資人等一項參考指標，以提前因應或避免公司或其利害關係人可能產生的損失。該研究與本研究的目的有相似的地方，都是希望藉由統計軟體的分析成果，以便提供企業經營者、投資人等可供參考的財務指標來因應。

陳柏宇(2017)針對台灣上市上櫃生物技術產業為研究對象，並以台灣經濟新報資料庫(TEJ)所提供財務報表資料為主要資料來源，其中包括經營能力、獲利能力、償債能力等相關財務指標，並採用統計軟體 R 語言來進行資料探勘的分析，以期探勘出有效益的成果，而此研究的目的有三個，第一點是分析台灣上市上櫃生物技術產業公司，其經營能力、獲利能力、償債能力指標是否存在關聯性或差異，第二點是針對經營能力指標其內部財務變數之間是否存在關聯性或差異，第三點是利用資料探勘中常見的分群分析，以期找出不同群組間的特徵，並達到資料減量。

在另一篇研究，謝彥馨(2017)也是使用與上述研究類似的流程與研究方法，也是以台灣經濟新報資料庫(TEJ)所提供財務報表資料為主要資料來源，但此研究的對象不同，它是以蘋果公司在台灣上市上櫃供應商作為研究對象，希望藉由統計軟體 R 語言其內建的多種套件來進行資料探勘的分析，以期挖掘出有用的資訊，提供公司經營者、投資人等作為另一種的財務指標參考，或是供給相關領域的學者作為後續的參考依據。

廖瑞成(2017)是探討台灣上市食品公司營運績效的分析，其資料來源是採用公開資訊觀測站和台灣經濟新報資料庫(TEJ)所提供財務報表資料，以經濟能力與獲利能力相關財務參數作為主要資料，並藉由 Entropy 權重法與直覺模糊 VIKOR 方法進行分析評估，以期分析成果能夠解釋上市食品公司營運績效優劣之原因及經營者改善公司體質的參考。

江誼庭(2017)是研究台灣製造業上市櫃公司現金流量對持有現金其敏感度的分析，此研究的樣本資料來源是由台灣經濟新報資料庫(TEJ)所提供，採用最小平方方法及四階動差法來進行分析評估，而兩種不同的分析方法的實證結果都是一致性的。綜合上述幾位學者的研究，利用台灣經濟新報資料庫(TEJ)提供的資料作為資料來源，並且在不同產業與領域的應用上，是相當多元且具有意義的，是顯而易見的。

第三節 資料包絡分析法(DEA)

壹、資料包絡分析法(DEA)之定義

在 1957 年 M.J Farrell 為解決生產面效率的分析，提出了以生產函數邊界 (Product Function Frontier, PPF) 的觀念，並且在評算估計後提出生產邊界是由最有效率的單位所組成、固定規模報酬為前題假設以及生產邊界凸向原點且每一點的斜率都不為正值，共三個基本假設。

Farrell (1957) 提出假設廠商使用投入要素 X_1 、 X_2 ，只生產單一產品 Y ，而每一單位產出由 X_1 、 X_2 的可能組合構成等量曲線定義為效率邊界；換句話說，在固定規模報酬假設之下，在效率邊界曲線上的每一點都具有相同的效率值，Farrell 模式效率分析圖如下：

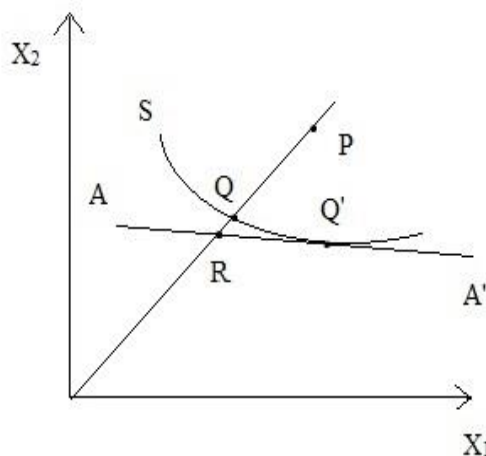


圖 2-2 Farrell 模式效率分析圖

資料來源：Farrell, M. J. (1957). The Measurement of Productive Efficiency, Journal of the Royal Statistical Society, Series A, General, Vol. 120, Part 3, pp. 253-281

P 點與 Q 點是相同的投入要素比率，組合效率與 Q 點相同，而 Q' 點是在有效率的組合中，最低的生產成本的組合；為改善技術效率，可透過模型的分析指出哪個投入要素未達最佳效率，並經由改變投入要素的比率改善組合效率。

貳、資料包絡分析法(DEA)之 CCR 模式

Charnes, Cooper & Rhodes (1978) 利用 Farrell 為理論基礎，並新增了固定規模之

規模效率遞增效率分析這項假設，改善只能處理單一產出效率分析問題，並建立多項投入、多項產出的效率分析模式，這項研究解決固定規模效率分析的線性規劃模式，稱為 CCR(Charnes、Cooper 及 Rhodes)模式。CCR 模式的效率分析方法可針對不同產業、組織或單位進行績效評估，其中每一個被評估的組織或單位都被看作一個決策單位(Decision Making Unit, DMU)，並且篩選適當的投入和產出要素來評估 DMU 的效率，用在以產出對於投入要素的比率建出線性組合，其比率值介於 0~1。這些比率代表每一個不同被評估的組織或單位的相對效率值，其數值越接近 1，則表示該組織單位有效率程度越高；DEA 之 CCR 模式的公式如下：

Max.

$$h_k = \frac{\sum_{r=1}^s U_r * Y_{rk}}{\sum_{i=1}^m V_i * X_{ik}} \quad (2-5)$$

Subject to :

$$\left(\frac{\sum_{r=1}^s U_r * Y_{rj}}{\sum_{i=1}^m V_i * X_{ij}} \right) \leq 1 \quad (2-6)$$

$i = 1, 2, 3, \dots, m$; $r = 1, 2, 3, \dots, s$; $j = 1, 2, 3, \dots, n$

Y_{rj} : 第 j 個 DMU 的第 r 個產出值 ; X_{ij} : 第 j 個 DMU 的第 i 個產出值

U_r : 第 j 個 DMU 的第 r 個產出要素的加權值

V_i : 第 j 個 DMU 的第 i 個產出要素的加權值

h_k : 第 k 個 DMU 的效率值

參、資料包絡分析法(DEA)之 BCC 模式

在 CCR 模式發展數年之後，由 Banker, Charnes & Cooper (1984) 提出改善的方法—BCC 模式，利用 Shephard (1953, 1970) 所提出的距離函數等方法，延伸了 CCR 模式的發展，並且可以適用於更多不同情境，並擴增了 CCR 模式之中固定規模之規模效率的假設，換句話說規模效率是可以固定、遞增或遞減，也比較符合現實社會實際發生的情形，而 BCC 模式可以計算出每一個不同被評估的組織或單位的單純技術效率值，其中可以區分為兩種—單純技術效率、規模效率。而規模效率的運算公式為：規模效率 = 總技術效率 / 單純技術效率。資料包絡分析法(DEA)之 BCC 模式(非線性規劃)的公式圖如下：

Max.

$$h_k = \sum_{r=1}^s U_r * Y_{rk} - U_o / \sum_{i=1}^m V_i * X_{ik} \quad (2-7)$$

Subject to :

$$(\sum_{r=1}^s U_r * Y_{rj} - U_o / \sum_{r=1}^s U_r * Y_{rj} - U_o) \leq 1 \quad (2-8)$$

$i = 1, 2, 3, \dots, m$; $r = 1, 2, 3, \dots, s$; $j = 1, 2, 3, \dots, n$

Y_{rj} : 第 j 個 DMU 的第 r 個產出值 ; X_{ij} : 第 j 個 DMU 的第 i 個產出值

U_r : 第 j 個 DMU 的第 r 個產出要素的加權值

V_i : 第 j 個 DMU 的第 i 個產出要素的加權值

h_k : 第 k 個 DMU 的效率值

U_o : 用來評估規模效率的型態

為了方便求解及運算，轉換上述 BCC 模式(非線性規劃)的公式成為線性規劃的公式，其公式如下：

Max.

$$h_k = \sum_{r=1}^s U_r * Y_{rk} - U_o \quad (2-9)$$

Subject to :

$$\sum_{i=1}^m V_i * X_{ik} = 1 \quad (2-10)$$

$$\sum_{r=1}^s U_{rj} * Y_{rj} - \sum_{r=1}^s U_r * Y_{rj} - U_o \leq 0 \quad (2-11)$$

肆、利用資料包絡分析法(DEA)相關之研究

陳禹志(2013)探討在金融海嘯的影響下，針對 22 個亞太地區的機場營運績效的研究，選用機場的地理位置、經濟成長率、機場規模及經營權的型態作為分析的參數，並利用資料包絡分析法(DEA)算出經營績效，其研究成果表示，機場的地理位置與規模對經營績效具有明顯的影響力，並且在最後在機場營運管理和營銷面提供一些建議與改善方向，以便經營者或研究相關領域的學者作為參考。

另一篇研究也是利用資料包絡分析法(DEA)計算出各國物流產業的營運績效，其投入與產出的選用為地理位置、經濟指標、所得水準、國家規模及物流費用等要素，希望藉由分析成果的比較找出值得仿效的國家，並探討此研究所提出的資通訊技術能力、基礎建設程度、創新能力對各國物流產業的營運績效是否具有影響力，綜合上述之研究發現，找出影響台灣物流產業的營運績效的要素，並

提出具體改善的方向與建議並讓台灣在全世界物流產業地位提昇，以期有朝一日能與歐美各先進國家並駕齊驅。(洪韻琇，2011)

洪海玲(2002)也是採用資料包絡分析法(DEA)針對台灣國內製造業做營運績效的研究，但製造業依其特性的不同，又可細分為民生工業、資訊電子工業、化學工業、金屬機械工業等，其研究分析所選用的投入與產出項目，分別為企業管理的生產管理、財務管理、行銷管理及人力資源管理指標和七項生產力的總要素，希望藉由分析找出影響各產業營運績效的原因，讓各產業的資源能做更有效的分配，以期提升國內製造業的生產力績效，其分析的成果也可以供相關領域學者研究或經濟管理者作為參考。

陳政任(2002)探討台灣資訊電子產業的經營績效，其研究的方法除了利用資料包絡分析法(DEA)之外，也結合了非財務指標的問卷調查所產生的項目，做為評估該產業的投入與產出要素，根據研究結果顯示，該產業在企業內部的管理面是有待加強改進的，另外也提出企業文化對於產業的經營績效的改善是具有影響力的。從研究的分析也可以看出經營績效好的公司，讓廠商、經營者仿效、瞭解改善方向並提升公司在該產業的地位，其研究成果也讓資訊電子產業在經營績效與管理面作為參考。這篇研究領域與本研究很有相關，除了一樣是採用資料包絡分析法(DEA)之外，後續的統計分析方法也有相似的地方，雖然與本研究的目的不同，作為利用相似資料領域與研究方法，提供本研究很好的參考與思考方向。

盧思珩(2006)探討 2000—2004 年之間上市製造業公司，希望藉由資料包絡分析法(DEA)的分析結果—經營績效，瞭解它是否能夠拿來作為未來盈餘的預測，並且與股價是否有相關，而另一方面經營績效的效率值，對資本市場研究之中的資訊內含及相關評價的觀點是否具有影響力。此研究的成果顯示，經營績效的效率值與未來盈餘及股價的關係，是呈現顯著且正相關的，除了具有預測未來股價、盈餘的功能之外，DEA 的效率值具備提供輔助預測的效果，而且對於資本市場的兩項研究也存在正面的影響力。此篇研究的研究成果提供許多在研究目的構思上的想法且具有很大的參考價值。

詹國華(2006)利用資料包絡分析法(DEA)並配合橫斷面與縱剖面的分析，來探討台灣 14 家金融控股公司 2003—2005 年的財務績效，以期提出對各金融控股公司財務績效改善的方向與該類產業的整體評估分析。其研究成果指出財務績效的評估，與獲利能力存在正向相關，並發現在各金融控股公司分析比較之下，經營

主體為保險業的公司，其財務績效的效率值為最佳，此篇研究的結論及過去許多研究也證實資料包絡分析法(DEA)，是可以評估金融控股公司的財務績效；而這樣的訊息也可以供經營者瞭解公司現況或社會大眾作為投資參考的指標。

第四節 相關分析(Correlation analysis)

壹、相關分析之定義

關聯分析 (Association analysis)，也是大家耳熟能詳的啤酒尿布理論，關聯規則探勘演算法之中 apriori 是最先被提出的，此分析意指假設在一個資料集之中，尋找支持度 \geq 最小支持度並且信賴度 \geq 最小信賴度的所有可能的組合規則，最小支持度、信賴度都是其相對應的門檻值。關聯分析涵蓋的範圍相對廣，如各變數之間的差別、影響及預測等關係，而相關分析也包括其中，它是探討兩個變數項目之間的關聯程度，分析出來的相關係數解讀可分為大小及方向，相關係數的大小意指兩變數間相關的程度是強還是弱，相關係數的方向為正相關還是負相關。本研究利用 R 語言運算相關分析是利用皮爾森相關係數(Pearson's correlation coefficient)來定義，它也適用本研究的資料型態—連續型變數。

貳、相關分析相關之研究

康勝修(2006)探討雲林科大圖書館的資料庫做資料探勘，希望可以找出借閱者與館藏查詢之間是否存在一定的規則、特徵值如男女性或不同院所、科系的學生在借閱的書籍等是否存在不一樣的地方，另一方面也希望勘查出一些有趣的資訊，可以供館方改善服務流程及查詢系統功能的提升，其研究結果藉由關聯分析探勘出的規則並結合相關分析做進一步的評估，確實協助館方達到他們想完成的目標，並且把館藏的分層重新排序、針對個系所更有主題、相關性的查詢，讓借閱者快速的查詢到他們想要的書籍。

呂旻娟(2013)探討影響台灣兩大家航空公司獲利與否的重要因素—燃料油價格與美元匯率與財務績效之間的關係，及比較兩家財務表現差異，並且利用相關與迴歸分析作為此研究的分析方法，此研究結果指出燃料油價格的影響並不顯著，因為兩家公司對於油價上漲時都有其因應的對策，但是匯率對於兩家公司的財務績效影響是相當明顯的，尤其是對於一些與獲利相關的財務變數，而從兩家

的財務分析與比較，不論是近期的表現或是長期的預測，都一致認為長榮航空的表現是比較優秀的。



第三章 研究方法

第一節 研究工具及軟體介紹

本研究採用 R-Studio Version 3.5.0 作為挖掘工具的原因，因為其本質是一個開源的統計軟體，它提供了一個可以快速建立視覺化的環境背景之外，也有許多統計分析相關的套件可以使用，現今作為許多資料科學、數據分析的軟體使用，是一個可以將資料分析、預測或模型建立等並結合視覺化呈現的研究工具。

第二節 研究設計及研究流程

本研究設計之流程圖，如下圖：

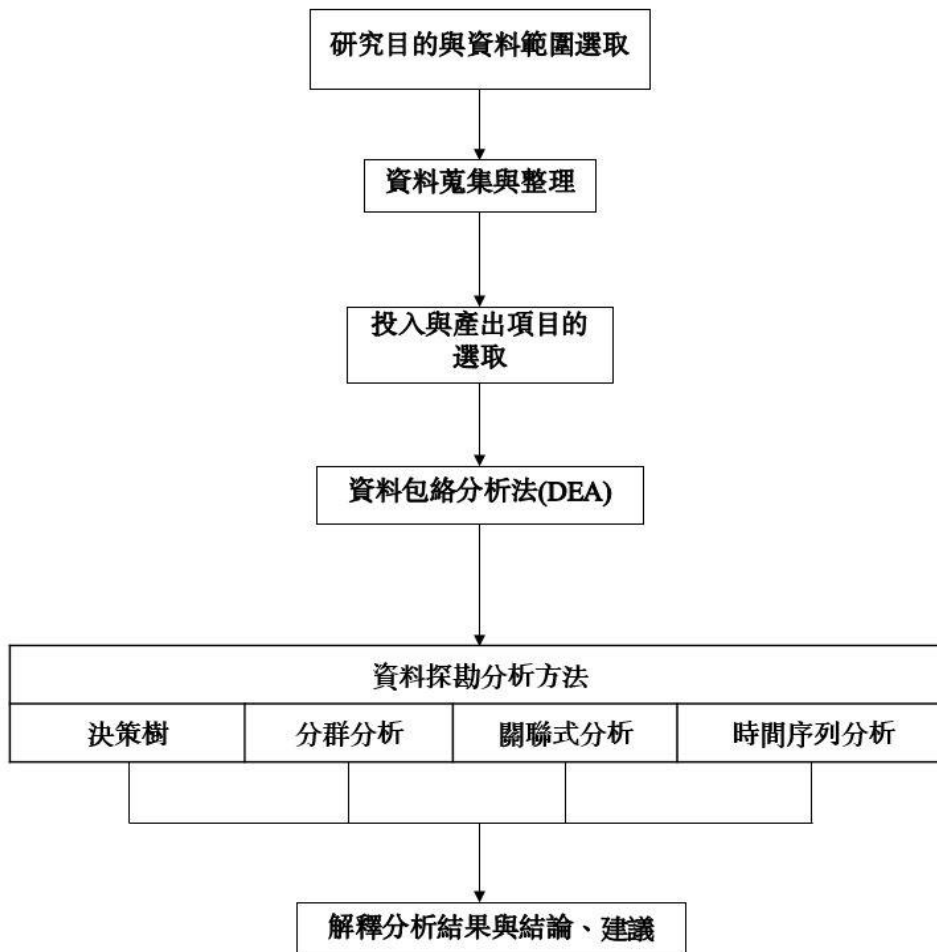


圖 3-1 研究設計及流程圖

第三節 資料來源

本研究的主要資料來源為台灣經濟新報資料庫(TEJ)，分別採用下述上市公司

2007—2016 年的財務報表資料，其中選取出與經營能力與獲利能力相關之財務資料作為資料包絡分析法(DEA)的評估、計算。研究範圍為 91 家台灣上市電子零組件產業，整理如下圖：

公司名稱	股票代號	公司名稱	股票代號	公司名稱	股票代號	公司名稱	股票代號	公司名稱	股票代號
首利	1471	信錦	1582	川湖	2059	台達電	2308	華通	2313
楠梓電	2316	國巨	2327	廣宇	2328	敬鵬	2355	燿華	2367
金像電	2368	智寶	2375	台光電	2383	群光	2385	正歲	2392
毅嘉	2402	環科	2413	錫新	2415	新巨	2420	建準	2421
興勤	2428	錫旺科	2429	聯昌	2431	太空梭	2440	新利虹	2443
奇力新	2456	飛宏	2457	建通	2460	良得電	2462	志聖	2467
立隆電	2472	鉅祥	2476	大毅	2478	百容	2483	希華	2484
華新科	2492	揚博	2493	健和興	3003	今皓	3011	全漢	3015
鴻名	3021	信邦	3023	禾伸堂	3026	偉訓	3032	欣興	3037
晶技	3042	健鼎	3044	立德	3058	日電貿	3090	晟鈦	3229
勝德	3296	聯德	3308	同泰	3321	泰碩	3338	新日興	3376
台端	3432	維熹	3501	嘉澤	3533	聯穎	3550	力銘	3593
宏致	3605	谷崧	3607	達邁	3645	健策	3653	新至陞	3679
銘鈺	4545	致伸	4915	鑫禾	4999	瀚宇博	5469	競國	6108
鎰勝	6115	金橋	6133	柏承	6141	嘉聯益	6153	鈞寶	6155
捷泰	6165	精成科	6191	佳必琪	6197	詮欣	6205	聯茂	6213
聚鼎	6224	定穎	6251	台郡	6269	康舒	6282	群電	6412
鈺邦	6449	台虹	8039	南電	8046	瀚荃	8103	志超	8213
菱光	8249								

圖 3-2 台灣上市電子零組件產業匯整圖

第四節 分析變數的選取—投入及產出選項

本研究以 DEA 分析在電子零組件公司的財務指標，並採用台灣經濟新報資料庫(TEJ)提供財務報表的資料，分為經營能力與獲利能力來分析，選取有相關財務比率，依據完整性、可衡量的、可解構性、不重覆性、最小化等五種特性，進行所選財務比率的刪減。

壹、經營能力

1. 流動比率：用來評估短期債務清償能力常用的比率，也是衡量企業短期風險的指標。
2. 速動比率：意思是指，速動資產對於流動負債的比率，也是衡量企業流動資產中可以立即變現用於償還流動負債的能力。
3. 總資產週轉次數：意指公司在一定期間內之收入，可以替換幾次總資產，也反映公司運用總資產來創造營收的能力。
4. 應收帳款週轉率：意指在一段時期內應收帳款轉換為現金資產的次數，應收帳款週轉率越高則收賬時間越短，而應收帳款的收回越快。
5. 固定資產週轉次數：意指公司在一定期間內之收入，可以替換幾次固定資產，也反映公司運用固定資產來創造營收的能力。

表 3-1 經營能力之財務變數敘述性統計

敘述性統計			
財務變數名稱	最大值	平均值	最小值
流動比率	5331.18	207.165	13.32
速動比率	5232.91	170.721	9.45
總資產週轉次數	4.29	0.798	0
應收帳款週轉次	41.33	4.0478	0.03
固定資產週轉次數	18132.85	105.7	0.01

資料來源：本研究整理

貳、獲利能力

1. 稅後淨利率：意指稅後淨利對於銷貨收入淨額的比率，並且可以看出公司在扣除稅額後的獲利狀況。
2. 營業毛利率：意指營業毛利額對於營業淨收入的比率，常用於企業獲利能力的分析。
3. 營業利益率：考量了在取得收入的過程中，所耗用的一切成本，常用於反映企業獲利能力的指標。
4. 每股盈餘(元)：意指公司盈餘對於加權平均股數的比率，而比率越高越好，表示公司的獲利能力越好。

表 3-2 獲利能力之財務變數敘述性統計

敘述性統計			
財務變數名稱	最大值	平均值	最小值
稅後淨利率	53874.9	34.403	-10831.92
營業毛利率	68.44	17.015	-64.4
營業利益率	2429.64	2.101	-6152.21
每股盈餘(元)	54.98	2.395	-10.1

資料來源：本研究整理

第四章 研究結果與分析

第一節 資料包絡分析法結果與說明

本研究是以 DEA 的分析結果作為其他分析的目標變數，其中本研究是以 DEA 之中的 CCR 模式進行分析進而求得 2007 到 2016 年 91 家公司共 310 筆數值，而下表為 2007 年至 2016 年，91 家電子零組件公司財務報表經由 DEA 分析之結果，十年期間由大到小排序，排名前 10 名的廠商，整理如下表：

表 4-1 2007—2016 年 DEA 排名前 10 位的廠商

年份 廠商	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
1582 信錦		1	1	0.915	0.564	1	0.678	0.702		
2059 川湖	0.658	0.589	0.739	0.666	0.778	0.674	0.756	0.893	0.998	1
2316 楠梓電									0.624	
2327 國巨							0.533	0.874	0.998	0.785
2375 智寶		0.533								0.81
2385 群光										0.64
2456 奇力新	0.557			0.608						
2467 志聖		0.504	0.542			0.739	0.65	0.611		
2478 大毅			0.511	0.685	0.63					
2484 希華	0.572									
2493 揚博		0.599			0.511					
3044 健鼎			0.629	0.875	1			0.597	0.701	1
3376 新日興	1	0.903	1	0.751	0.683	1	1	0.946	1	0.804
3533 嘉澤				0.561						
3605 宏致	1					0.494				
3645 達邁				0.85	0.989	0.867	0.866	0.955	1	0.96
3653 健策			0.55	0.572						
3679 新至陞	1									1
4545 銘鈺							0.539			

4915 致伸	1									
4999 鑫禾		0.491	0.772		0.519	0.567	0.585	0.678	0.852	1
6141 柏承	0.734									
6155 鈞寶	0.788	0.78	0.77	0.702	0.658	0.744	0.635		0.566	
6165 捷泰		0.554								
6213 聯茂									0.533	
6224 聚鼎	0.538	0.514	0.497		0.526	0.556	0.783	0.69	0.684	0.7
6269 台郡								0.529		
8103 瀚荃						0.565				

資料來源：本研究整理

依據表 4-1，予以 2007—2016 年 DEA 排名前 10 位的廠商視覺化呈現的圖形，如下圖 4-1~圖 4-4：

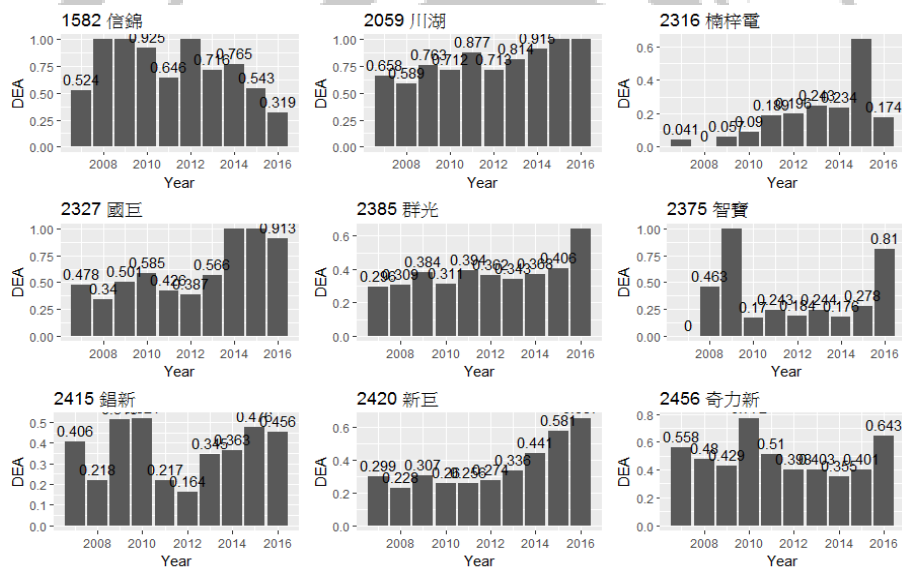


圖 4-1 2007~2016 年 DEA 排名前 10 位的廠商

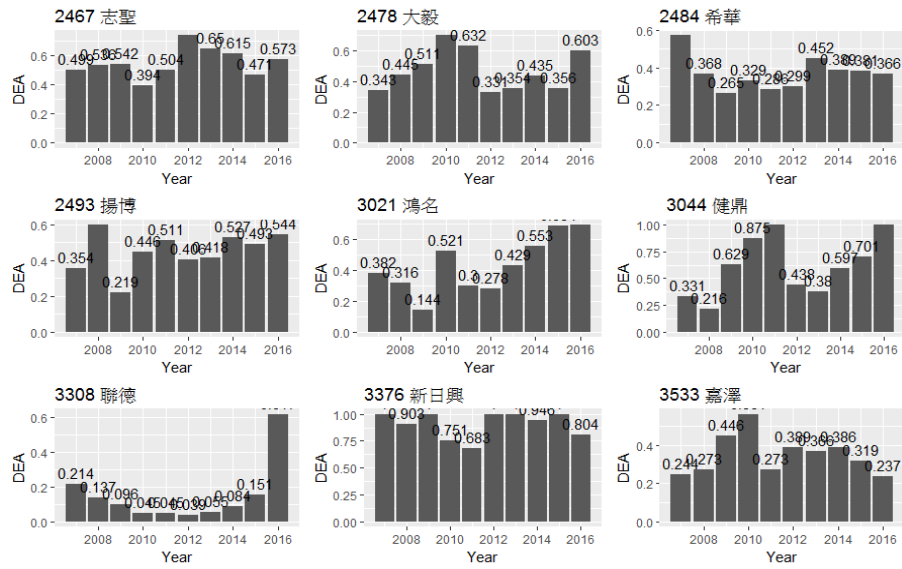


圖 4-2 2007~2016 年 DEA 排名前 10 位的廠商

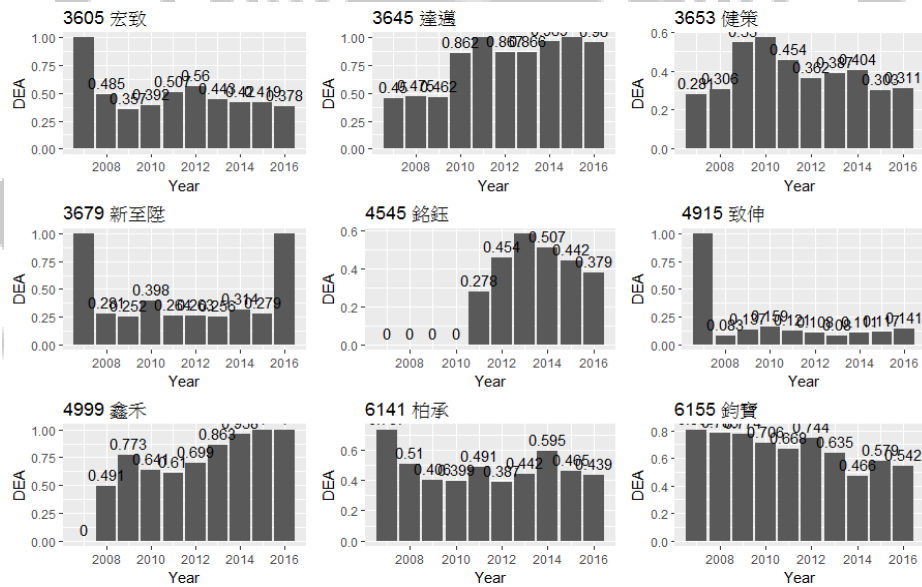


圖 4-3 2007~2016 年 DEA 排名前 10 位的廠商

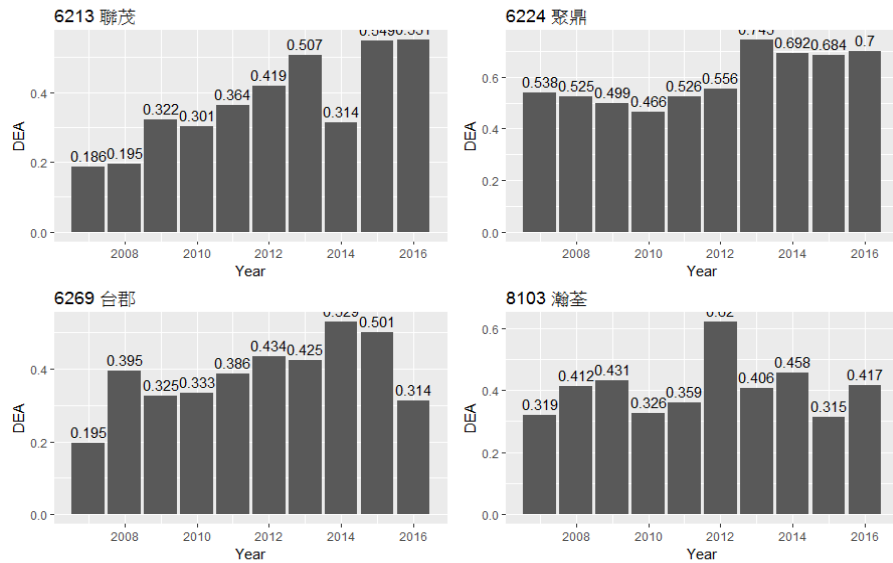


圖 4-4 2007~2016 年 DEA 排名前 10 位的廠商

由上述的四張圖表中，我們可以發現到其 DEA 數值逐年慢慢上升的廠商有川湖、國巨、群光、新巨、揚博、鴻名、達邁、鑫禾、聯茂、聚鼎總計 10 家，它們的經營效率有往好的方向持續發展、改善，但是也有經營效率 > 0.5 且長期維持的廠商，包含信錦、川湖、志聖、新日興、達邁、馨禾、鈞寶、聚鼎總計 8 家，其經濟績效表現也是不容忽視，也是可以作為投資標的選項。

第二節 決策樹分析結果與說明

本研究將針對發生金融大事件的 2008、2010、2016 年進行單年度的決策樹分析，其中各單年度的分析項目為全部投入與產出要素，其分析結果如下圖：

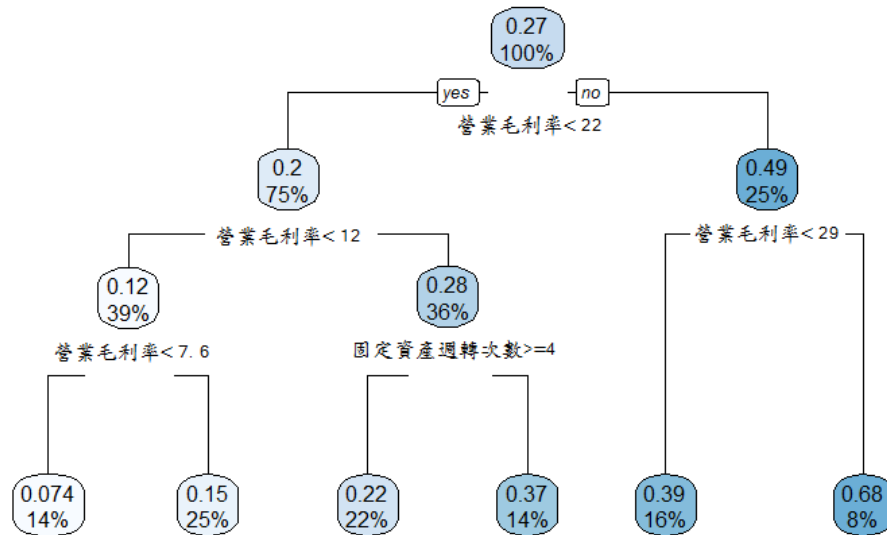


圖 4-5 2008 年決策樹分析圖

依照 2008 年單年度進行決策樹分析的結果，營業毛利率 >22 ，營業毛利率 >29 的數值為最佳的選擇，而分析出來的上市公司分別為，信錦、川湖、鈞寶、新日興、聚鼎、志聖、宏致，其中針對前五名的廠商其 DEA 數值皆大於 0.5，分別為信錦、川湖、鈞寶、新日興、聚鼎，在華爾街金融海嘯影響最嚴重的一年，其經營績效仍高於平均值，是值得讓人關注、投資的廠商。

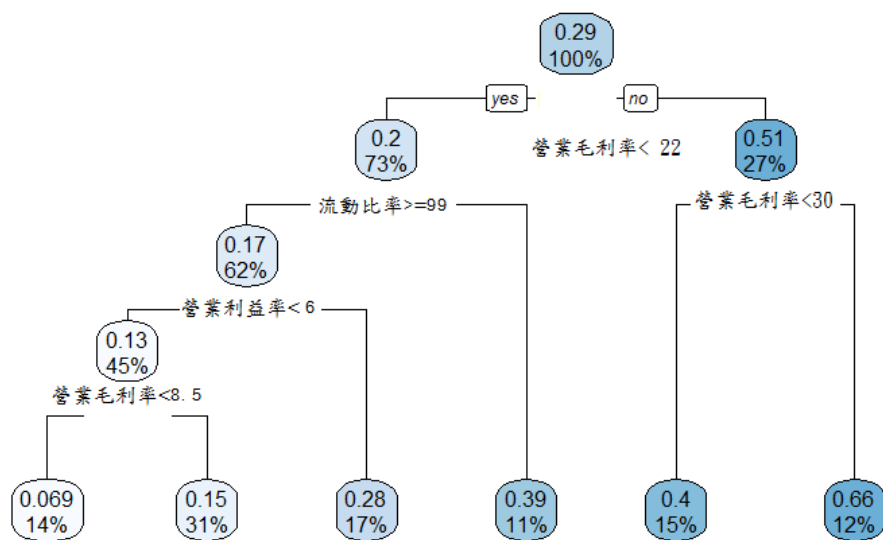


圖 4-6 2010 年決策樹分析圖

依照 2010 年單年度進行決策樹分析的結果，營業毛利率 >22 ，營業毛利率 >30 的數值為最佳的選擇，而分析出來的上市公司分別為，信錦、川湖、大毅、鴻名、新日興、達邁、健策、鑫禾、鈞寶、聚鼎，其中針對前五名的廠商其 DEA 數值皆大於 0.5，分別為信錦、大毅、新日興、達邁、鈞寶，在 2010 年為歐債危機影響最嚴重的一年，其經營績效仍高於平均值，而在這前五名的廠商之中信錦、新日興這兩家廠商在上次金融危機的時候，經營績效的表現也是相當好，值得我們持續關注。

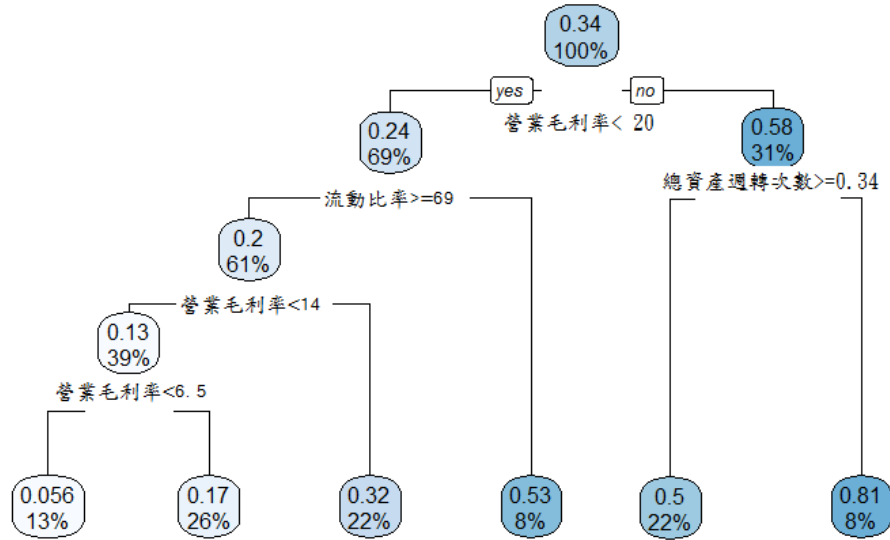


圖 4-7 2016 年決策樹分析圖

依照 2016 年單年度進行決策樹分析的結果，營業毛利率 >20 ，總資產週轉次數 >0.34 的數值為最佳的選擇，而分析出來的上市公司分別為，台光電、新巨、興勤、奇力新、志聖、百容、希華、揚博、健和興、信邦、晶技、新日興、宏致、達邁、健策、銘鈺、柏承、詮欣、聚鼎、瀚荃，其中針對前五名的廠商其 DEA 數值皆大於 0.5，分別為新巨、奇力新、新日興、達邁、聚鼎，在 2016 年黑天鵝事件頻傳的一年，經營績效仍然表現非常好，其中可以發現在這三個影響金融市場的大事件當中，新日興表現依然穩健，其後續金融事件發生時是可以考慮作為投資標的的公司之一。

本研究進行決策樹分析將 2007~2016 年資料分為全部經營能力和獲利能力要素和只單一看獲利能力要素，而全要素投入決策樹分析的結果，如下圖 4-8：

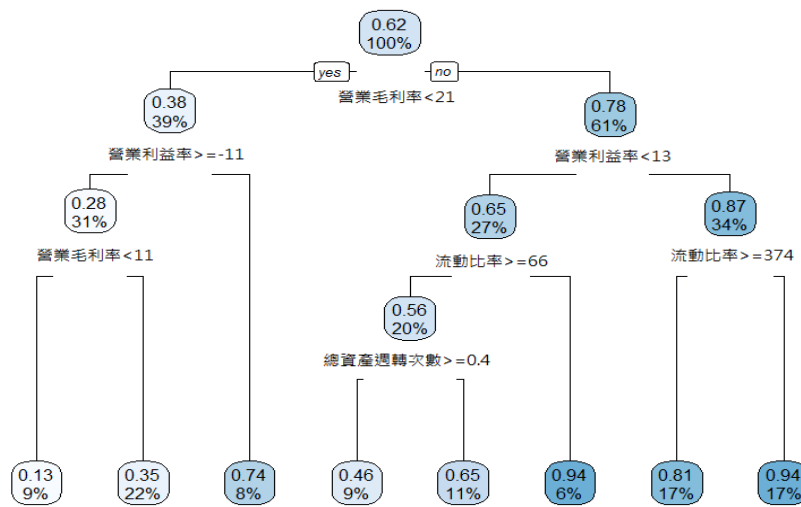


圖 4-8 產出與投入全要素之決策樹分析結果

依照決策樹分析結果，營業毛利率 >21 ，營業利益率 >13 ，流動比率 ≥ 374 的數值為最佳的選擇，而分析出來的上市公司分別為，川湖、新日興、鑫禾，依照財務報表的分析，這三家的 DEA 數值與股價逐年上升，符合本研究採用決策樹分析找出體質及未來前景較佳的上市公司。而單一分析獲利能力要素決策樹分析的結果，如下圖 4-9：

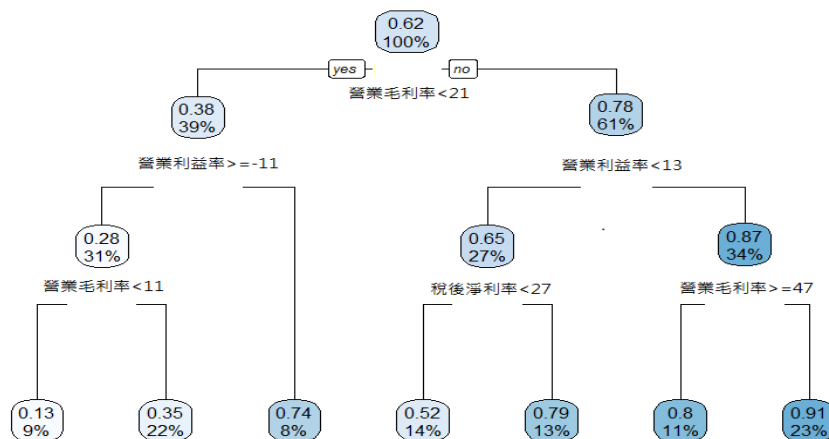


圖 4-9 產出要素之決策樹分析結果

依照決策樹分析結果，營業毛利率 >21 ，營業利益率 >13 ，營業毛利率 ≥ 47 的數值為最佳的選擇，而分析出來的上市公司分別為，川湖、新日興、鑫禾，和

上一個分析出來的公司一致，依照財務報表的分析，這三家的 DEA 數值與股價逐年上升，符合本研究採用決策樹分析找出體質及未來前景較佳的上市公司。

第三節 群集分析結果與說明

我們利用 R 語言的套件—factoextra 來找出最佳的分群數量，針對的時間點為發生金融大事件的 2008、2010、2016 年，其分析的結果如下圖：

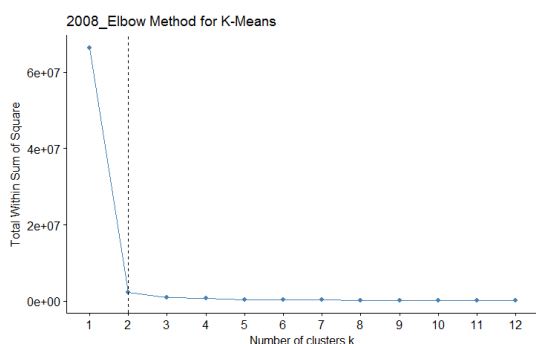


圖 4-10 2008 年最佳的分群數目

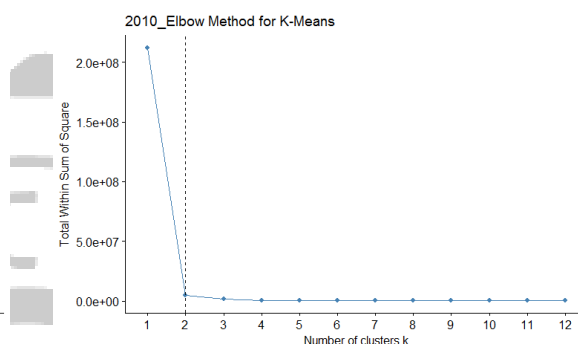


圖 4-11 2010 年最佳的分群數目

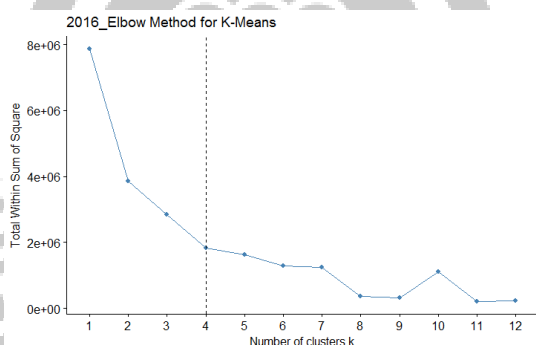


圖 4-12 2016 年最佳的分群數目

根據此套件應用到 K-Means 的最佳分群數量各別為，2008、2010 年為 2 群，2016 年為 4 群，並依此分析結果的分群數目來作為分群分析的參考，而 K-Means 分群分析的結果如下圖：

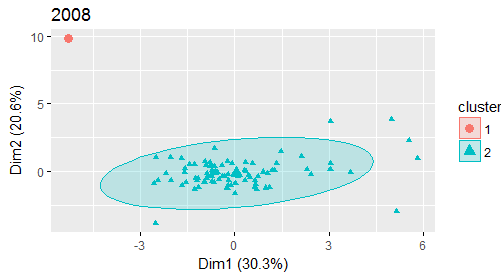


圖 4-13 2008 年分群分析圖

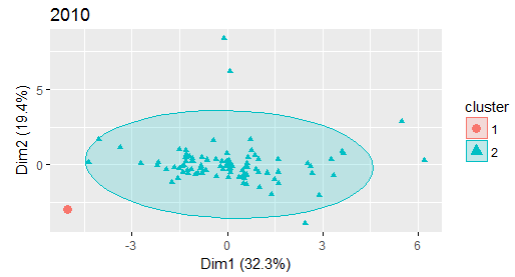


圖 4-14 2010 年分群分析圖

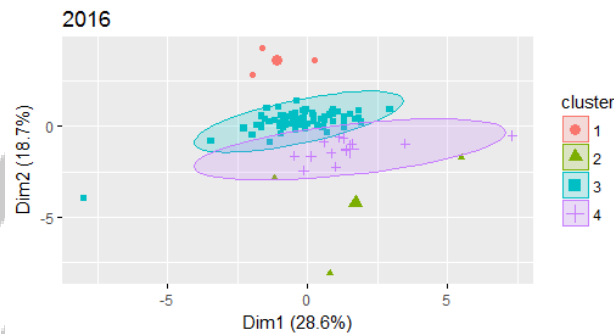


圖 4-15 2016 年分群分析圖

依照分群分析結果分析，如圖 4-13、圖 4-14 所示，2008 年與 2010 年的第一個分群的廠商皆是精成科，其 DEA 分析值分別為 0.059、0.032，而其餘的廠商都是落點在第二個分群，其分析結果與決策樹分析判斷出來的結果不同，也無法確切去判斷投入、產出要素及 DEA 數值的關係，所以再後續分群分析將針對 2008 年與 2010 年再多一個分群分析，來觀察是否分群分析的結果與財務要素及 DEA 數值的關係能夠更加明確；2016 年的分群分析如圖 4-15 所示，第一群的廠商為鋁新、廣宇、致伸，第二群的廠商為川湖、捷泰、南電，第四群的廠商為百容、希華、揚博、健和興、晟鈦、台端、維熹、健策、鑫禾、金橋、聚鼎、台郡、台虹，而其他沒提到的廠商皆分為第三群，其中在 2016 年 DEA 數值排名前 10 名的廠商分別為川湖、國巨、智寶、健鼎、新日興、達邁、新至陞、鑫禾、聚鼎，10 家當中落點在第三群的廠商數量是最多的，其中與分群相關性最高的要素為每股盈餘，其次是稅後淨利率及營業毛利率，這顯示 K-Means 分群分析的結果與獲利能力的要素有很大的相關性。

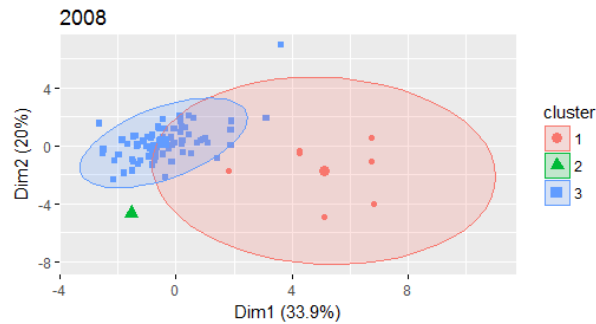


圖 4-16 2008 年分群分析圖

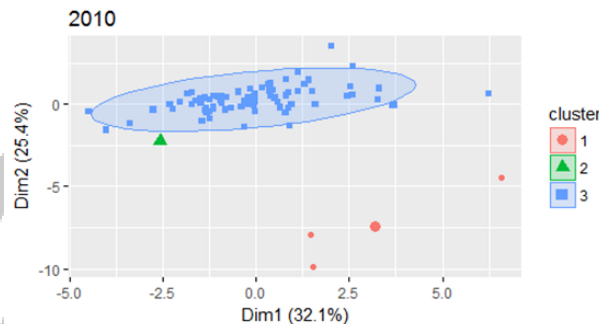


圖 4-17 2010 年分群分析圖

增加 2008 年與 2010 年的分群數目，而 2008 年的分群分析如圖 4-16 所示，第一群的廠商為川湖、揚博、今皓、新日興、維熹、鑫禾、南電，第二群的廠商為精成科，剩下的廠商為第三群，其中在 2008 年 DEA 數值排名前 10 名的廠商分別為信錦、川湖、志聖、揚博、新日興、鑫禾、柏承、鈞寶、捷泰、聚鼎，10 家當中落點在第三群的的廠商數量是最多的，其中與分群相關性最高的要素為營業毛利率，其次是稅後淨利率，由此顯示 K-Means 分群分析的結果與獲利能力的要素有很大的相關性；2010 年的分群分析如圖 4-17 所示，第一群的廠商為川湖、智寶、捷泰，第二群的廠商一樣為精成科，剩下的廠商為第三群，其中在 2010 年 DEA 數值排名前 10 名的廠商分別為信錦、川湖、國巨、奇力新、大毅、健鼎、新日興、達邁、鑫禾、鈞寶，10 家當中落點在第三群的的廠商數量是最多的，其中與分群相關性最高的要素為稅後淨利率，其次是營業毛利率及營業利益率，由此顯示 K-Means 分群分析的結果與獲利能力的要素有很大的相關性。

第四節 時間序列分析結果與說明

綜合 2008 年、2010 年、2016 年決策樹分析結果較佳的前五名的廠商以及 2007~2016 年決策樹分析結果較佳的廠商，我們整理出來的廠商分別為川湖、新日興、達邁、鑫禾，進一步做時間序列分析，其分析結果如下圖所示：

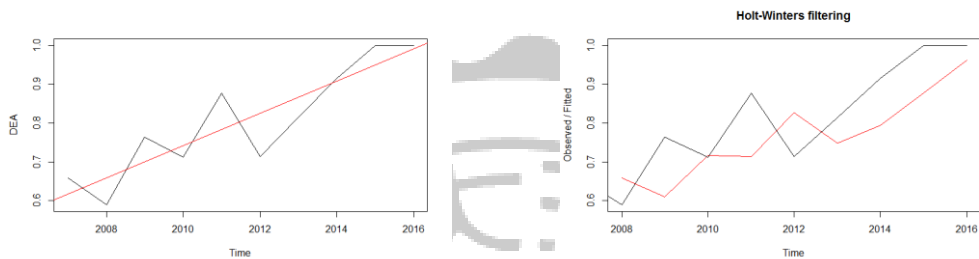


圖 4-18 川湖歷年 DEA 時間分析圖

圖 4-19 川湖 Holt 預測分析圖

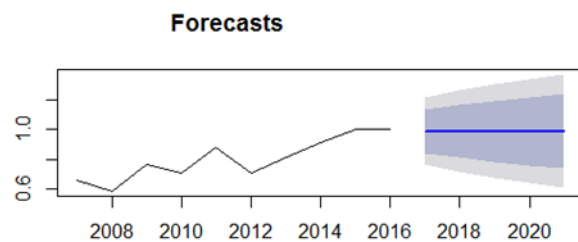


圖 4-20 川湖 HoltWinters 預測未來四年分析圖

廠商川湖的時間序列分析結果，圖 4-18 中的紅色線為 DEA 時間分析圖的趨勢線，隨著時間的演進，有正向成長的趨勢，而圖 4-19 為依據 DEA 的數值作 Holt 預測且為同個時間軸，對照其實際 DEA 的數值，與紅色線條的成長方向是相似的，最後在圖 4-20 在針對川湖歷年的 DEA 的數值，做未來四年的預測，而計算出來的預測數值皆為 0.988，以長期的趨勢來看，川湖的經營績效是趨於正向成長的，是個值得期待與觀察的投資目標。

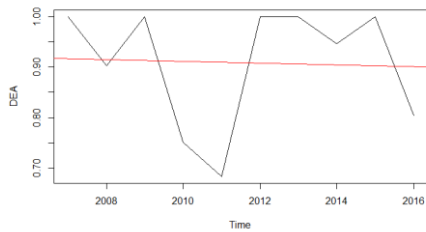


圖 4-21 新日興歷年 DEA 時間分析圖

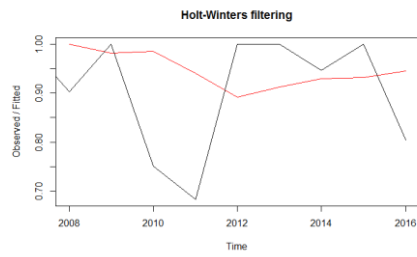


圖 4-22 新日興 Holt 預測分析圖

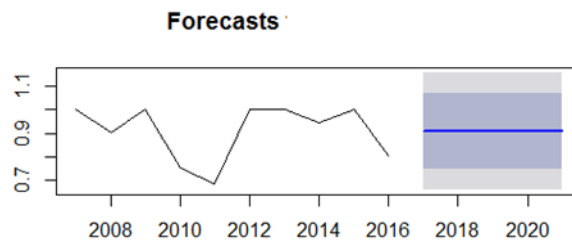


圖 4-23 新日興 HoltWinters 預測未來四年分析圖

廠商新日興的時間序列分析結果，圖 4-21 中的紅色線為 DEA 時間分析圖的趨勢線，隨著時間的演進，是呈現水平成長的趨勢，而圖 4-22 為依據 DEA 的數值作 Holt 預測且為同個時間軸，對照其實際 DEA 的數值，與紅色線條的成長方向是有些不同的地方，短期間來看，在 2010 與 2011 的歐債金融事件影響下，是受到很大的影響，由此也可以看出其客戶群是落在歐美地區的可能性相當大，最後在圖 4-23 在針對新日興歷年的 DEA 的數值，做未來四年的預測，而計算出來的預測數值皆為 0.909，以長期的趨勢來看，新日興的經營績效是長期來看是趨於穩定發展的，也是個值得期待與觀察的投資目標。

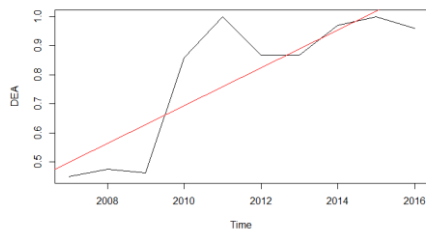


圖 4-24 達邁歷年 DEA 時間分析圖

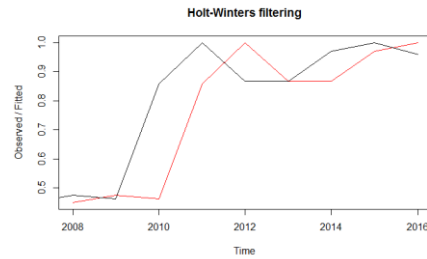


圖 4-25 達邁 Holt 預測分析圖

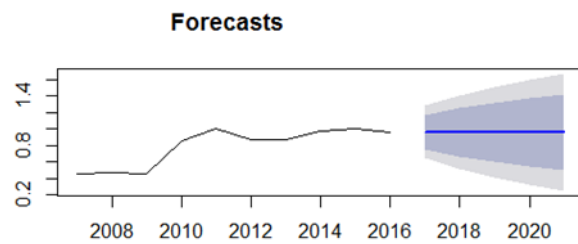


圖 4-26 達邁 HoltWinters 預測未來四年分析圖

廠商達邁的時間序列分析結果，圖 4-24 中的紅色線為 DEA 時間分析圖的趨勢線，隨著時間的演進，有正向成長的趨勢，而圖 4-25 為依據 DEA 的數值作 Holt 預測且為同個時間軸，對照其實際 DEA 的數值，與紅色線條的成長方向是相似的，最後在圖 4-26 在針對達邁歷年的 DEA 的數值，做未來四年的預測，而計算出來的預測數值皆為 0.96，以長期的趨勢來看，達邁的經營績效是趨於正向成長的，是個值得期待與觀察的投資目標。

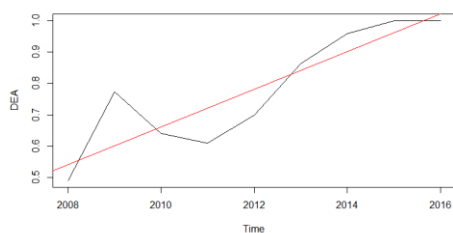


圖 4-27 鑫禾歷年 DEA 時間分析圖

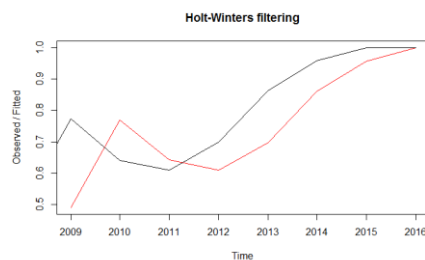


圖 4-28 鑫禾 Holt 預測分析圖

Forecasts

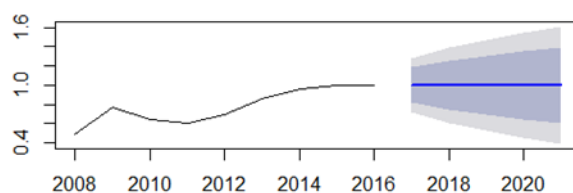


圖 4-29 鑫禾 HoltWinters 預測未來四年分析圖

廠商鑫禾的時間序列分析結果，圖 4-27 中的紅色線為 DEA 時間分析圖的趨勢線，隨著時間的演進，有正向成長的趨勢，而圖 4-28 為依據 DEA 的數值作 Holt 預測且為同個時間軸，對照其實際 DEA 的數值，與紅色線條的成長方向是相似的，最後在圖 4-29 在針對鑫禾歷年的 DEA 的數值，做未來四年的預測，而計算出來的預測數值皆為 0.999，以長期的趨勢來看，鑫禾的經營績效是趨於正向成長的，是個值得期待與觀察的投資目標。

第五節 相關分析結果與說明

本研究希望透過全部的投入與產出的要素和 DEA 分析出來的結果，來進一步探討它們是否具有一定的相關性，一般而言，相關係數大於 0.8 為極高度相關，介於 0.8 與 0.6 之間為高度相關，介於 0.6 與 0.4 之間為普通相關，介於 0.4 與 0.2 之間為低度相關，相關係數小於 0.2 為極低度相關，其分析的結果如下表：

表 4-2 投入及產出變數與 DEA 數值的 Pearson's 相關

變數	T 值	P 值	相關係數
流動比率	2.041	0.042 (*)	0.069
速動比率	1.492	0.136	0.05
總資產週轉次數	-13.941	$2.2e^{-16}$ (***)	-0.426
應收帳款週轉次數	0.833	0.405	0.028
固定資產週轉次數	-2.864	0.004 (**)	-0.096
稅後淨利率	0.211	0.833	0.007
營業毛利率	33.496	$2.2e^{-16}$ (***)	0.749
營業利益率	2.232	0.026 (*)	0.075
每股盈餘(元)	15.869	$2.2e^{-16}$ (***)	0.472

註：* $p < 0.05$ 、** $p < 0.01$ 、*** $p < 0.001$

由表 4-2 的分析結果發現，營業毛利率與 DEA 是呈現高度且正向的相關，總資產週轉次數、每股盈餘(元) 與 DEA 是呈現普通相關，而前項是負向的相關，後項則是正向的相關，其餘變數與 DEA 的相關性並不顯著，由此可以看出獲利能力的要素影響 DEA 的分析結果是有很大的相關性。

第五章 結論與研究限制

本研究利用了幾個資料探勘的方法論與統計檢定的分析，其主要的分析變數是在 DEA 的分析數值，它也是影響後續探勘方法的依據，針對這個重要的變數，在選擇投入與產出的要素上，試驗許多不同的組合，以目前的投入與產出的組合所計算出的 DEA 數值，是比較符合實際廠商的財務報表的資訊的，經由決策樹分析的結果，可以看出營業毛利率與營業利益率是影響節點判斷的重要參數，而在集群分析方面，可以清楚看到稅後淨利率、營業毛利率與營業利益率，這三個是影響分群分割時的參考依據，由此也可以看出，在決策樹與集群分析所依據的參數其相關性是存在的，從相關分析的結果，獲利能力的要素對於 DEA 的分析結果是有很大的影響的。本研究的貢獻是在於針對 DEA 分析值做時間序列的未來預測，有別於以往的短時間、週期的時間序列預測，以期找出合適的投資標的。

綜合上述統整：

1. 影響公司營運績效的要素主要是獲利能力當中的營業毛利率以及營業毛利率。
2. 獲利能力對於公司營運績效影響是顯著的。
3. 經由決策樹分析所整理出來的廠商，在時間序列的分析的結果表現，以長期持有的角度觀察，都是值得一再關注或投資的參考目標。

研究限制與建議：

在本研究所採用的電子零組件產業領域，有少部份的公司上市時間較晚而容易被忽略的產業，因為分析期間短暫無法看出其價值，所以無法全面代表電子零組件產業領域。

未來改善方向：

1. 以年為時間單位的預測可能造成預測結果較不準確，可以改善使用季或者是月，讓預測更精準。
2. 電子類股的其他類別廠商也可以拿來做分析、比較，以期可以找出模範的企業、廠商來仿效，提升台灣電子類股產業的地位。
3. 建議搭配其他財務分析指標與 DEA 分析，讓資料探勘的分析或預測能夠更加準確。

參考文獻

中文文獻

1. 王尹柔(民 103)。台灣加權股價指數報酬與波動性之特性。22-36。
2. 王貞治(民 103)。匯率預測模型績效之研究-時間序列及人工智慧模型之應用。24-70。
3. 江誼庭(民 106)。臺灣製造業上市櫃公司現金流量對持有現金的敏感度研究。16-33。
4. 呂旻娟(民 102)。國際原油價格與美元匯率變動 對國籍航空公司財務績效影響之研究。41-66。
5. 林基域(民 105)。股票價格趨勢預測之研究。5-57。
6. 林志杰(民 95)。應用資料探勘技術建立中西醫腦中風診斷模型之研究。16-56。
7. 洪韻琇 (民 100)。以資料包絡法分析台灣與各國物流效率。32-45。
8. 洪海玲(民 91)。以資料包絡分析法作製造業之營運效率分析。27-50。
9. 施正宏(民 93)。結合總體經濟指標及個股財報資料以預測個股漲跌-以台灣電子類股為例。23-90。
10. 高建國民 102)。台灣股市波動度探討。5-42。
11. 翁雅君(民 104)。透過時間序列方法分析 YouBike 使用狀況：以信義區為例。6-35。
12. 康勝修(民 95)。應用資料探勘技術於圖書館借閱紀錄分析與館藏查詢服務。27-54。
13. 黃麗蒨(民 89)。臺灣地區地下水品質之統計研究。5-31。
14. 曾惠敏(民 104)。顧客價值分析與行銷策略之研究—以汽車百貨業為例。11-54。
15. 詹國華(民 95)。臺灣金融控股公司財務績效評估之研究—資料包絡法之應用。34-89。
16. 廖瑞成(民 106)。應用 Entropy 權重法與 VIKOR 於臺灣上市食品公司財務績效評估之研究。32-29。
17. 郭宜易(民 106)。應用資料採礦技術解讀記憶卡產品的發展應用。6-60。
18. 陳政任(民 91)。以資料包絡法探討台灣資訊電子業之核心能力與經營績效 4-6、32-49。

19. 陳郁璿(民 104)。以企業生命週期探討環境績效與公司經營績效關聯性之研究-以台灣上市電子公司為例。22-49。
20. 陳昭伶(民 107)。公司治理、經營績效與財務結構關聯性之研究。20-31。
21. 陳柏宇(民 106)。以資料探勘技術探討企業經營績效之研究-以台灣生技產業為例。19-33。
22. 陳禹志(民 102)。金融海嘯後亞太地區機場營運效率之研究：運用資料包絡分析。40-67。
23. 盧思珩(民 95)。資料包絡分析法效率值具有預測盈餘的資訊價值與資訊內涵之實證研究。28-59。
24. 謝彥馨(民 106)。以資料探勘技術探討蘋果在台供應商中各公司之經營績效。7-29。
25. 鄭忠樑(民 91)。運用分類樹於股價報酬率預測之研究。27-61。

英文文獻

1. Abarbanell, J. S. , & B. J. Bushee ,(1997). Fundamental Analysis, Future Earnings and Stock Prices. *Journal of Accounting Research*. pp 1-24.
2. Avijan,Dutta, Gautam, Bandopadhyay, Suchismita, & Sengupta,(2012). Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression. *International Journal of Business and Information*, Vol. 7, No. 1. pp 105-136.
3. Banker, R. D., Charnes, A. and Cooper, W. W.(1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis,*Management Science*, Vol. 30, No. 9, pp. 1078-1092.
4. Berry,M.J.and Linoff,G.(1997). *Data Mining Techniques for Marketing*. Wiley Press.
5. Curt,H.(1995).The Deville's in The Detail:Techniques,Tool,and Applications for Data mining and knowledge Discovery-Part 1. *Intelligent Software Strategies*,6(9):3
6. Fayyad,Usama M,G. Piatetsky-Shapiro and P. Symth. (1996).From data mining to knowledge discovery in databases,*AI Magazine*,pp.37-54,1996.
7. Frawley,W.J.,Piatetsky-Shapiro,G.,and atheus,C.J.(1992).Knowledge discovery in database: An overview In *Knowledge Discovery in Databases*:1-27. Cambridge, MA:AAAI/MIT Press,Reprinted in *AI Magazine*,13(3)Hand, D.J., Mannila, H. and Smyth, P. (2001).*Principles of Data Mining*, MIT Press, Cambridge, MA.
8. J. B. MacQueen.(1967). Some Methods for classification and Analysis of

- Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1:281-297
9. Liao,S.H.,&Chen,Y.J.(2004).Mining customer knowledge for electronic catalog marketing. *Expert Systems with Applications*,27(2),521–532
 10. Quinlan, J. R. (1986). Induction of Decision Trees. *Mach. Learn.*, 1(1), 81-106.
 11. Shephard, R. W. (1953). *Cost and Production Functions*, Princeton University Press, Princeton, NJ.
 12. Shephard, R. W. (1970). *The Theory of Cost and Production Functions*, Princeton University Press, Princeton, N.J.
 13. Sherman, D. (1982). *Data Envelopment Analysis and Other Approaches for Measuring the Efficiency of Hospital Operations*, DBA Thesis, Harvard University Graduate School of Business. Also available from University Microfilms International, Ann Arbor, Mich.
 14. Ward,J.H. (1963). Hierarchical Grouping to Optimize an Objective Function *J.Am.Statist. Assoc.* 58, 236-244

