

## 資料探勘於台灣股市之擇股分析

作者一：程守雄  
服務單位與職稱  
建國科技大學  
資訊管理系  
助理教授  
e-mail：

shcheng@ctu.edu.tw

作者二：陳亦伶  
服務單位與職稱  
建國科技大學  
資訊管理系  
學生  
e-mail：

c94806013@mail.ctu.edu.tw

### ■ 摘要

股票買賣是現今相當普遍的一種投資活動，投資人買賣股票皆希望能夠降低投資風險，提高報酬率。本研究運用資料探勘技術，從台灣股市上市、櫃公司的財報歷史資料中挖掘出個股財務指標與股價報酬率之漲跌關聯性，提供一般大眾做為投資之參考。

本文所提出之研究模式為選擇一般產業類股作為研究對象，以過去 3 年時間連續性季報表資料作為研究區間，對研究樣本進行關聯法則、群集分析、決策樹分類法等分析，再利用導出之規則對本年度進行擇股投資，並將所得之平均報酬率與大盤報酬率兩項數據進行實證分析比較。結果顯示本研究所得之平均報酬率明顯優於大盤報酬率。

**關鍵詞：**資料探勘、財務指標、關聯規則、群集、決策樹

### ■ Abstract

Stock trade is a popular investing activity and during this activity, investors expect to gain profit with lower risk. This research is aimed on the discover relationship between financial data indicators of public companies and return on investment on stock price using data mining technology.

The research has proposed a research method using general industry as research population. Data is gathering from past three years. Use the methods of relation rules, cluster, and decision tree, and it can provide suggestions for stock selection of current years. Further, use this suggestions to select some profitable stocks of this year. The outcome evidences the higher return on investment in proposed model than general market average.

**Keywords:** Data mining, Financial norm, Relation rules, Cluster, Decision Tree

### 1. 前言

現今投資理財已成為人們生活中的一

環，股市是眾多投資方法中最能獲取高額報酬的方法，但同時它也是眾多投資方法中最具風險的一種，為了降低風險，證期會的公開股市觀測站定期與不定期揭露各上市公司財務報表，其中包括五大項指標，財務結構、償債能力、經營能力、獲利能力、和現金流量等 20 項財務比率供投資人參考。

在過去的文獻中，具筆者所知，已有多位學者利用統計方法探討過各個財務比率與股價的關連性，例如，陳文勇[1]針對電子產業中哪些財務比率會影響每股盈餘，並應用 Logit 與 Tobit 來建立區別模型，並預測公司未來每股盈餘是高或低、成長或衰退。申立銘[2]的研究是從基本面出發，使投資者瞭解股價指數與總體經濟指標間關連性之後，再透過財務指標來評估個別航運公司的經營績效，作為投資人購入後長期持有航運類股票的參考。蘇祐萱[3]的研究是以藉由貝氏網路架構呈現財務指標與每股盈餘的因果關係，並且經由模式推論結果解釋機率的不確定性；其研究依據獲利能力、經營能力、財務結構、現金流量、償債能力選取 20 項財務指標變數。施正宏[4]的研究是同時考量外在的總體經濟榮枯總體經濟指標與內在的公司營運狀況個股財報資料，使能更準確的預測個別公司的股價走勢。利怡玫[5]的研究是利用因素分析法，以財務報表和相關資訊來評估台灣上市公司之經營績效，並探討其累積股票報酬率及累積異常報酬之關係。鄭文英等[6]以不同危機歷程定義經營財務階段，探討上市上櫃公司所處經營財務階段及其後續財務階段可能移轉情形、預期存活等資訊，偵測出可區分不同經營財務階段的財務指標，作為投資者擇股之參考。研究結果顯示，股東權益佔資產比率、速動比率及稅前純益佔實收資本比率的值若愈高，表示公司經營狀況愈佳。詹場等[7]以學理上可能影響股價的 13

總體經濟因素建構台灣證券市場指數之預測模型，運用逐步迴歸法與多變量之因素分析法建構，以台灣證券管理委員會規範上市公司揭露的 22 種財務比率為樣本。林江亮等[8]除了考慮各項財務比率，另加入關係人交易資訊後，探討相關議題。

資料探勘[13, 14]是近年來被廣泛應用於各領域來協助分析大量資料的技術，主要的訴求是希望透過資料探勘從大量的歷史資料中挖掘其中富含價值的知識。目前，應用資料探勘的技術去探討財務比率與股價的關連性的研究還比較少，楊宗彥[9]利用上市公司的財務比率來預測股票投資報酬，再以資料挖掘中的類神經網路與決策樹技術，以了解對於報酬預測準確度的高低。吳月明[10]利用上市公司的財務比率，運用人工智慧中的倒傳遞類神經網路及灰色理論中的灰預測模型，來對股票投資報酬做預測並分析兩種工具對其投資報酬預測準確度的高低。葉怡成等[11]透過資料探勘技術中的關聯規則，對台股股價漲跌進行分析，以確認關聯規則預測股價漲跌的可行性。研究發現股價漲跌形成的關聯規則會顯現在產業關聯裡，例如：化工、光學鏡頭、光碟等產業會形成關聯群組；而預測股價漲跌的關聯規則可預測到股價的補跌現象。然而，從廣大的個別投資者的角度來看，發展一套可以簡單表示又容易看懂的投資規則，以便於在複雜且眾多的股票中挑選適當的優質股票，是一個重要且有迫切需要的研究題目。

本研究使用營業費用率、現金流量比率、流動比率、速動比率、營業成本、營業利益、應收帳款週轉率(次)、應付帳款付現天數(天)、淨值報酬率(稅後)、淨值週轉率、每股盈餘、營業毛利率、淨值成長率、總資產報酬率等 14 項財務比率，利用資料探勘的技術對研究樣本進行關聯法則、群集分析、決策樹分類法等分析，歸納出簡單又容易理解的投資規則，大幅的簡化投資規則的複雜度。

## 2. 研究方法

### 2.1 資料來源

本研究樣本以民國 94 年第一季至 97 年第三季間，以上市、上櫃公司的財務指標作為條件屬性，股價報酬率為決策屬性，資料來源為台灣經濟新報資料庫(TEJ)所查得，14 項財務指標定義如下：

(1) 營業費用率%：營業費用/銷貨收入淨

額

- (2) 現金流量比率%：營業活動淨現金流量/流動負債
- (3) 流動比率%：流動資產/流動負債，用以衡量短期償債能力。流動資產(現金+有價證券+應受款項+存貨+預付費用)，流動負債代表一年之內必須動用流動資產償付的短期負債。流動比率越高代表短期償債能力越好，通常高於 100%。
- (4) 速動比率%：速動資產/流動負債=(現金+有價證券+應受款項)除以流動負債=(流動資產-存貨-預付費用)除以流動負債，衡量極短期清償負債的能力。越高則代表短期償債能力越好。
- (5) 營業成本：企業在一段期間內因為經常性的營業活動而銷售商品或提供勞務等等所應該要負擔的成本，其內容主要包括：銷貨成本、勞務成本等。銷貨成本又可分為：自製產品銷貨成本與外購產品的銷貨成本兩大類，通常對製造業而言是前者居多，而其他行業則大多為後者。由成本會計上的定義，營業成本是由產品在整個製造過程中所發生的成本，在製造業也稱為製造成本、產品成本。製造成本由直接材料+直接人力+製造費用(包含間接材料、間接人工及與產品製造或廠房營運有關的其他間接成本)所組成。
- (6) 營業利益：營業利益/實收資本
- (7) 應收帳款週轉率(次)：賒銷淨額(或銷售淨額)/平均應收帳款。衡量應收帳款收現的速度及授信政策是否太緊或太鬆。應收款項週轉率越高代表收款的效率越好。
- (8) 應付帳款付現天數(天)：平均應付帳款/營業成本\*天
- (9) 淨值報酬率(稅後)%：股東權益報酬率就是該年度股東權益的成長速度。其中稅後淨利部分係指扣除掉特別股股利後之盈餘，而股東權益係指普通股權益總額。由股東權益成長率可以看出，該公司的經營階層的目標是否與股東目標一致，是以股東權益為主要考量。股東權益報酬率係由於企業保留其盈餘所獲得，因此亦顯示一個企業如果不仰賴對外舉債也能促使其企

業成長的能力。其計算方式為：股東權益報酬率=(稅後淨利-股利發放數)/股東權益

- (10) 淨值週轉率：營業收入淨額/平均淨值
- (11) 每股盈餘：(稅後純益-特別股股利)/發行在外普通股股數
- (12) 營業毛利率：營業毛利/營業收入，通常用來比較同一產業公司產品競爭力的強弱，顯示出公司產品的訂價能力、製造成本的控制能力及市場佔有率，也可以用來比較不同產業間的產業趨勢變化。
- (13) 淨值成長率：淨值會隨著市場價格的波動使資產增加或減少，而其上下增減的幅度，則稱為淨值成長率。
- (14) 總資產報酬成長率：代表在某一段時間內(通常為一年)，公司利用總資產為股東所創造的利潤較上期所成長的比率。

## 2.2 關聯規則

關聯規則是由 Agrawal[15]於1993年首先提出的，關聯規則的形式如左所示” $X \rightarrow Y[\text{Support}, \text{Confidence}]$ ”，其中  $X \subset I, Y \subset I, X \cap Y = \phi$ ，語意為「IF  $X$  Then  $Y$ 」， $X$  表示為因， $Y$  表示是果，在此關聯規則有包含兩各重要的因素：

1. 支持度(Support;  $s$ )：指  $X$  此資料項目在資料庫  $D$  中所佔的比例，且  $s$  介於零到壹之間( $0 < s \leq 1$ )，如公式(1)。而  $s(X \rightarrow Y)$  形式為  $P(X \cup Y)$ ，表示為同時發生  $X$  與  $Y$  的兩個交易項目的機率。

$$s = \frac{X}{D}, (0 < s \leq 1) \quad (1)$$

2. 信賴度(Confidence;  $c$ )：表示發生某事件的情況下，發生另一事件的機率，亦可視為關聯的強度。如  $c(X \rightarrow Y)$  形式為  $P(Y | X)$ ，表示為在  $X$  資料項目下，發生  $Y$  資料項目之機率，且  $c$  介於零到壹之間( $0 < c \leq 1$ )，如公式(2)。

$$c(X \rightarrow Y) = \frac{X \cap Y}{X}, (0 < c \leq 1) \quad (2)$$

根據每個財務指標是否漲跌與股價是否漲跌，計算出每個財務指標與股價的關係，由各財務指標與股價呈現同向成長的次數，即可計算出各財務指標與股價的信賴度。

## 2.3 集群分析

本研究採用 k-means 集群法，k-means 集群的方式是以亂數隨機選擇各群之中心，也就是假設將資料分為  $k$  群，就以亂數隨機選擇  $k$  個數來做為  $k$  群的中心，再以每一數與群中心之距離，移動至與期較接近之群組，再計算其新的中心值，一直重覆至群內之元素不再變動為止。

## 2.4 決策樹分類模型

ID3為最早使用的決策樹演算法之一，ID3的主要核心是以遞迴的方式將訓練資料作切割。在每一次產生節點時，某些輸入的訓練子集將取出測試，以資訊獲取量來當作測試，在選取過後，將以具有最大資訊獲取量的值當作分支的節點，接下來依照其遞迴的動作選取下一個分支節點，直到每一個訓練資料都屬於一個分類之中 或是符合某個滿足條件。C4.5是ID3的延伸方法，改善了ID3產生過多子集合，而每個子集合僅包含少數資料的問題，並且具備處理連續數值型屬性、雜訊的處理，另外也兼具修剪樹的能力。C4.5 在決策樹的每個節點上使用資訊獲取量來選擇測試屬性，選擇具有最高資訊獲取量(或最大熵壓縮)的屬性作為當前節點的測試屬性。

本研究採用C5.0來產生決策樹並且預測結果類別。C5.0是延續C4.5的演算法架構，不同的是C5.0提供了在許多應用上較受歡迎的規則集，將分類的條件一一以規則形式表達，

增加了閱讀分類規則的可讀性。C5.0 可用來處理數值性或是名目性欄位的資料，分析結果可用容易理解的決策樹或是若-則的關係呈現。C5.0 與 C4.5 不同之處在於 C5.0 可以處理更多種資料型態，如日期、時間、時間戳記、序列性的離散型資料...等資料型態。將 C5.0 決策樹分類演算方法說明如下：

(1) 模型參數

C5.0 不限制只能做二元分割，這是與其它決策樹分類法不同的地方。先根據資訊獲取量來決定分割，建構一棵完整的決策樹，再針對每一個內部節點依定義的錯誤預估率來作決策樹修剪的動作，選擇資訊增大量最大的屬性為分割屬性。

(2) 分支演算法

根據資訊獲取量來決定分割的預測變數，假設資料集 S 包含二個類別 P 與 N，而其中 P 類別有 p 筆資料，N 類別有 n 筆資料，所含的資訊量表

$$I(p,n) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n} \quad (3)$$

若以預測變數 A 得出的分類數有 v 筆，其中包含了 P 類別  $p_i$  個與 N 類別  $n_i$  個，那麼以預測變數 A 的熵為

$$E(A) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} I(p_i, n_i) \quad (4)$$

，在此以 A 為分割節點可得到的資訊增加量為  $Gain(A) = I(p,n) - E(A)$ ，而 C5.0 選擇資訊增加量最大的屬性為分割節點。

(3) 決策樹停止成長規則

- A. 節點內只剩下一種值。
- B. 節點內的預測變數的數目已經與所有目標變數的數目皆相同時。
- C. 樹的分割層數達到預設的最大值。
- D. 母節點內的資料個數已小於預設可分割的值。
- E. 子節點內的資料個數已小於預設的值。

F. 修剪決策樹

採用錯誤預估率作判斷條件，由底部往上測試每樹葉節點，若被分類於某一子樹的所有訓練資料個數為 T，其中有 E 個訓練資料分類錯誤，而錯誤預估率就是由 E/T 評估，當有新的資料被測試時，則利用 E/T 判斷結果為不正確的機率有多高。若以一個樹葉節點代替後所得到的錯誤預估率比原先來得低則將修剪為單一樹葉節點，否則就保留原來的結構。

3. 實證分析

3.1 資料前處理

將台灣上市公司中的 600 家股票及上櫃公司中的 393 家公司的原始財務與股價資料，如表 3.1，先進行資料前處理，將各個財務比率與股價是否漲跌轉變成 T 與 F，處理後的資料如表 3.2。

表 3.1 原始資料

證券代碼	公司	營業費用率	有息負債利率	現金流量比率	流動比率	...	本學以來報酬率
1101	台泥	3.73	0.38	9.39	158	...	-12.4403
1102	亞泥	3.56	0.6	23.85	172.35	...	-7.1428
1103	嘉泥	7.77	0.58	1.48	117.69	...	-12.9035
1104	環泥	13.9	0.48	-2.28	161.58	...	-2.703
1108	幸福	5.7	0.8	17.14	104.43	...	-18.1054
1109	信大	9.88	0	58.76	789.91	...	-7.7571
1110	東泥	4.75	0.49	0.81	389.7	...	-13.8889
1201	味全	29.46	0.68	-5.48	100.65	...	-16.0666
1203	味王	27.23	0.49	-0.93	73.38	...	-13.2476
1210	大成	5.77	0.48	38.47	143.1	...	0.1197
1213	大欣	16.76	4.91	-9.74	319.7	...	-3.4188
1215	卜蜂	5.3	0.71	52.15	173.72	...	-15.4928
1216	統一	17.69	0.43	3.77	89.61	...	-12.0483
1217	愛之味	30.24	0.84	2.29	70.48	...	-10.6919

表 3.2 資料處理後的資料

證券代碼	公司	營業費用率	有息負債利率	現金流量比率	流動比率	...	本學以來報酬率
1101	台泥	T	T	T	T	...	F
1102	亞泥	T	T	T	T	...	F
1103	嘉泥	T	T	T	T	...	F
1104	環泥	T	T	F	T	...	F
1108	幸福	T	T	T	T	...	F
1109	信大	T	F	T	T	...	F
1110	東泥	T	T	T	T	...	F
1201	味全	T	T	F	T	...	F
1203	味王	T	T	F	T	...	F
1210	大成	T	T	T	T	...	T
1213	大欣	T	T	F	T	...	F
1215	卜蜂	T	T	T	T	...	F
1216	統一	T	T	T	T	...	F
1217	愛之味	T	T	T	T	...	F

3.2 關聯規則

利用關聯規則將各個財務比率與股價漲跌加以分析，算出各個財務比率與股價漲跌的信賴度。上市公司結果如表 3.3、3.4、3.5，上櫃公司結果如表 3.6、3.7、3.8。

表 3.3 上市公司各財務比率之信賴度(1)

期間	營業費用率	現金流量比率	流動比率	速動比率	營業成本
9401~9403	0.4533	0.5500	0.4533	0.4533	0.4533
9404~9406	0.4633	0.5100	0.4633	0.4633	0.4633
9407~9409	0.3650	0.4550	0.3667	0.3667	0.3667
9410~9412	0.7050	0.6483	0.7083	0.7083	0.7083
9501~9503	0.3850	0.5333	0.3850	0.3850	0.3850
9504~9506	0.6183	0.5483	0.6183	0.6183	0.6183
9507~9509	0.5367	0.5517	0.5367	0.5367	0.5367
9510~9512	0.9200	0.6950	0.9200	0.9200	0.9200
9601~9603	0.5450	0.5533	0.5450	0.5450	0.5450
9604~9606	0.8100	0.6683	0.8100	0.8100	0.8083
9607~9609	0.6150	0.5883	0.6150	0.6150	0.6150
9610~9612	0.0833	0.3167	0.0817	0.0817	0.0817
平均信賴度	0.5417	0.5515	0.5419	0.5419	0.5418

表 3.4 上市公司各財務比率之信賴度(2)

期間	營業利益	應收帳款週轉率次	應付帳款付現天數	淨值報酬率稅後	淨值週轉率次
9401~9403	0.5450	0.4550	0.6200	0.5683	0.4583
9404~9406	0.5033	0.4650	0.4633	0.5000	0.4650
9407~9409	0.5133	0.3667	0.3667	0.5117	0.3683
9410~9412	0.6533	0.7050	0.7100	0.6733	0.7067
9501~9503	0.4983	0.3850	0.3867	0.4700	0.3867
9504~9506	0.6150	0.6167	0.6183	0.6083	0.6167
9507~9509	0.6150	0.5350	0.5333	0.6250	0.5383
9510~9512	0.7183	0.9183	0.9167	0.7283	0.9183
9601~9603	0.5867	0.5433	0.5433	0.6000	0.5433
9604~9606	0.7350	0.8083	0.8083	0.7583	0.8133
9607~9609	0.6383	0.6167	0.6117	0.6350	0.6183
9610~9612	0.2350	0.0833	0.0850	0.2867	0.0833
平均信賴度	0.5714	0.5415	0.5553	0.5804	0.5431

表 3.5 上市公司各財務比率之信賴度(3)

期間	營業費用率	現金流量比率	流動比率	速動比率	營業成本
9401~9403	0.5725	0.5445	0.5725	0.5725	0.5725
9404~9406	0.4555	0.5394	0.4555	0.4555	0.4555
9407~9409	0.4198	0.5115	0.4198	0.4198	0.4198
9410~9412	0.6794	0.6158	0.6819	0.6819	0.6819
9501~9503	0.4275	0.4860	0.4275	0.4275	0.4275
9504~9506	0.5191	0.5089	0.5216	0.5216	0.5216
9507~9509	0.5242	0.5496	0.5242	0.5242	0.5242
9510~9512	0.8957	0.6845	0.8982	0.8982	0.8982
9601~9603	0.6310	0.5649	0.6310	0.6310	0.6310
9604~9606	0.7888	0.5903	0.7888	0.7888	0.7888
9607~9609	0.4758	0.4987	0.4758	0.4758	0.4758
9610~9612	0.0865	0.3384	0.0865	0.0865	0.0865
信賴平均	0.5397	0.5360	0.5403	0.5403	0.5403

表 3.6 上櫃公司各財務比率之信賴度(1)

期間	營業利益	應收帳款週轉率次	應付帳款付現天數	淨值報酬率稅後	淨值週轉率次
9401~9403	0.6209	0.5725	0.5725	0.6107	0.5725
9404~9406	0.5344	0.4555	0.4555	0.5522	0.4555
9407~9409	0.5547	0.4198	0.4198	0.5064	0.4198
9410~9412	0.6641	0.6819	0.6819	0.6845	0.6819
9501~9503	0.4936	0.4275	0.4275	0.5064	0.4275
9504~9506	0.5649	0.5216	0.5191	0.5700	0.5191
9507~9509	0.6387	0.5242	0.5267	0.6310	0.5242
9510~9512	0.7023	0.8982	0.8982	0.6743	0.8982
9601~9603	0.6209	0.6310	0.6310	0.6209	0.6285
9604~9606	0.6794	0.7888	0.7888	0.6947	0.7913
9607~9609	0.5344	0.4758	0.4758	0.5598	0.4758
9610~9612	0.3104	0.0865	0.0891	0.3486	0.0865
信賴平均	0.5765	0.5403	0.5405	0.5799	0.5401

表 3.7 上櫃公司各財務比率之信賴度(2)

期間	每股盈餘元	營業毛利率	淨值成長率	總資產報酬成長率
9401~9403	0.6132	0.6005	0.6183	0.5140
9404~9406	0.5496	0.4936	0.5471	0.6616
9407~9409	0.5089	0.4402	0.5954	0.6692
9410~9412	0.6819	0.6870	0.6005	0.6590
9501~9503	0.5115	0.4453	0.5344	0.5827
9504~9506	0.5700	0.5318	0.5165	0.5674
9507~9509	0.6310	0.5369	0.5954	0.6539
9510~9512	0.6692	0.8830	0.7430	0.5573
9601~9603	0.6158	0.6463	0.6285	0.5700
9604~9606	0.6947	0.7583	0.6768	0.5980
9607~9609	0.5573	0.4936	0.5496	0.5878
9610~9612	0.3486	0.1349	0.2748	0.4911
信賴平均	0.5793	0.5543	0.5734	0.5927

表 3.8 上櫃公司各財務比率之信賴度(3)

期間	每股盈餘元	營業毛利率	淨值成長率	總資產報酬成長率
9401~9403	0.6132	0.6005	0.6183	0.5140
9404~9406	0.5496	0.4936	0.5471	0.6616
9407~9409	0.5089	0.4402	0.5954	0.6692
9410~9412	0.6819	0.6870	0.6005	0.6590
9501~9503	0.5115	0.4453	0.5344	0.5827
9504~9506	0.5700	0.5318	0.5165	0.5674
9507~9509	0.6310	0.5369	0.5954	0.6539
9510~9512	0.6692	0.8830	0.7430	0.5573
9601~9603	0.6158	0.6463	0.6285	0.5700
9604~9606	0.6947	0.7583	0.6768	0.5980
9607~9609	0.5573	0.4936	0.5496	0.5878
9610~9612	0.3486	0.1349	0.2748	0.4911
信賴平均	0.5793	0.5543	0.5734	0.5927

### 3.3 集群分析

K-means 演算法，是透過平均信賴度的數據，用數據相近方式將資料分割成數個集群。由表 3.3、3.4、3.5 中的上市公司各財務比率之信賴度進行分群，本研究取 k=4 即分成 4 群做分析，分群結果如表 3.9。

表 3.9 上市公司經集群分析結果

序號	會計科目	平均信賴度	群組
1	營業費用率	0.5419	1
2	現金流量比率	0.5515	4
3	流動比率	0.5419	1
4	速動比率	0.5419	1
5	營業成本	0.5418	1
6	營業利益	0.5714	3
7	應收帳款週轉率次	0.5415	1
8	應付帳款付現天數	0.5553	4
9	淨值報酬率稅後	0.5804	2
10	淨值週轉率次	0.5431	1
11	每股盈餘元	0.5815	2
12	營業毛利率	0.5538	4
13	淨值成長率	0.5808	2
14	總資產報酬成長率	0.5878	2

由表 3.6、3.7、3.8 中的上櫃公司各財務比率之信賴度進行分群，分群結果如表 4.10。

表 3.10 上櫃公司經集群分析結果

序號 <sup>a</sup>	會計科目 <sup>b</sup>	信賴度平均 <sup>c</sup>	群組 <sup>d</sup>
1 <sup>e</sup>	營業費用率 <sup>e</sup>	0.5397 <sup>e</sup>	1 <sup>e</sup>
2 <sup>e</sup>	現金流量比率 <sup>e</sup>	0.5360 <sup>e</sup>	1 <sup>e</sup>
3 <sup>e</sup>	流動比率 <sup>e</sup>	0.5403 <sup>e</sup>	1 <sup>e</sup>
4 <sup>e</sup>	速動比率 <sup>e</sup>	0.5403 <sup>e</sup>	1 <sup>e</sup>
5 <sup>e</sup>	營業成本 <sup>e</sup>	0.5403 <sup>e</sup>	1 <sup>e</sup>
6 <sup>e</sup>	營業利益 <sup>e</sup>	0.5765 <sup>e</sup>	3 <sup>e</sup>
7 <sup>e</sup>	應收帳款週轉率 <sup>e</sup>	0.5403 <sup>e</sup>	1 <sup>e</sup>
8 <sup>e</sup>	應付帳款付現天數 <sup>e</sup>	0.5405 <sup>e</sup>	1 <sup>e</sup>
9 <sup>e</sup>	淨值報酬率稅後 <sup>e</sup>	0.5799 <sup>e</sup>	3 <sup>e</sup>
10 <sup>e</sup>	淨值週轉率 <sup>e</sup>	0.5401 <sup>e</sup>	1 <sup>e</sup>
11 <sup>e</sup>	每股盈餘元 <sup>e</sup>	0.5793 <sup>e</sup>	3 <sup>e</sup>
12 <sup>e</sup>	營業毛利率 <sup>e</sup>	0.5543 <sup>e</sup>	4 <sup>e</sup>
13 <sup>e</sup>	淨值成長率 <sup>e</sup>	0.5734 <sup>e</sup>	3 <sup>e</sup>
14 <sup>e</sup>	總資產報酬成長率 <sup>e</sup>	0.5927 <sup>e</sup>	2 <sup>e</sup>

3.4 決策樹

C5.0 規則歸納演算法，是透過解釋欄位跟輸出欄位間的關係，用遞迴方式將資料分割成子集後建立出決策樹並導出決策樹規則，嘗試解釋資料中不同部分跟輸出欄位或結果的關係。

由表 3.9，將財務比率與股價漲跌進行決策樹分析。規納出的規則如圖 3.1，說明如下：

- (1) 如果 總資產報酬成長率  $\leq 0.07$  且 總資產報酬成長率  $\leq -0.43$ ，則 股價下跌。
- (2) 如果 總資產報酬成長率  $\leq 0.07$  且 總資產報酬成長率  $> -0.43$  且 每股盈餘元  $> 0.9$  且 總資產報酬成長率  $> -0.38$ ，則 股價下跌。
- (3) 如果 總資產報酬成長率  $\leq 0.07$  且 總資產報酬成長率  $> -0.43$  且 每股盈餘元  $\leq 0.9$ ，則 股價上漲。
- (4) 如果 總資產報酬成長率  $\leq 0.07$  且 總資產報酬成長率  $> -0.43$  且 每股盈餘元  $> 0.9$  且 總資產報酬成長率  $\leq -0.38$ ，則 股價上漲。
- (5) 如果 總資產報酬成長率  $> 0.07$ ，則 股價上漲。

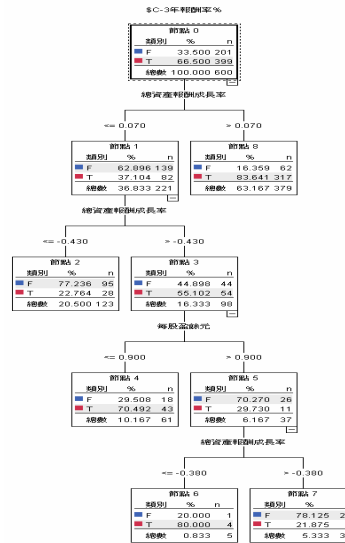


圖 3.1 上市公司決策樹產生之樹狀圖

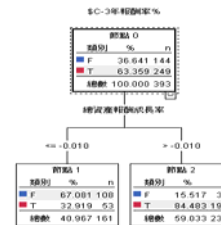


圖 3.2 上櫃公司決策樹產生之樹狀圖

由表 3.9，將財務比率與股價漲跌進行決策樹分析。規納出的規則如圖 3.1，說明如下：

- (1) 如果 總資產報酬成長率  $\leq -0.01$ ，則 股價下跌。
- (2) 如果 總資產報酬成長率  $> -0.01$ ，則 股價上漲。

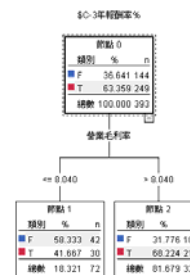


圖 3.3 上櫃公司決策樹產生之樹狀圖

圖 3.3 樹狀圖產生規則如下：

- (1) 如果 營業毛利率  $\leq 8.04$ ，則 股價下跌。
- (2) 如果 營業毛利率  $> 8.04$ ，則 股價上漲。

3.5 進行模擬投資

由圖 4.1 產生的規則對 97 年前三季上市公司的季財報資料進行擇股，符合上漲規則的公司數中，第一季有 227 家，其平均季報酬率為 22.68%；第二季有 56 家，其平均季報酬率為 19.74%；第三季有 12 家，其平均季報酬率為 11.65%。

由圖 3.2、3.3 產生的規則對 97 年前三季上櫃公司季財報資料進行擇股，符合上漲規則的公司數中，第一季有 73 家，其平均季報酬率為 14.2%；第二季有 37 家，其平均季報酬率為 15.43%；則第三季有 13 家，其平均季報酬率為 5.66%。

**表 3.11 上市、上櫃季報酬率與大盤報酬率比較表**

	上市平均季報酬率	上櫃平均季報酬率	大盤報酬率(不含金融)
97 年第一季	22.68%	14.2%	-0.02%
97 年第二季	19.74%	15.43%	-0.11%
97 年第三季	11.65%	5.66%	-0.21%

由表 3.11 得知，上市、上櫃平均季報酬率均遠高於大盤報酬率。

#### 4. 結論

本研究主要的貢獻在於利用關聯法則、群集分析、決策樹分類模型等資料探勘技術，從公司財務報表提供的資料，來尋找影響公司股價漲跌的規則，進而尋找優良的公司。由以上分析可以得知：

- (1) 實證結果得出之平均報酬率，都比大盤報酬率來得高。
- (2) 影響上市公司股價漲跌的主要決策財務比率為總資產報酬成長率、每股盈餘兩項。
- (3) 影響上櫃公司股價漲跌的主要決策財務比率為總資產報酬成長率、營業毛利率兩項。
- (4) 凡是財務報表亮麗的公司，自然可得到投資者的認同，股價自然上升，所以投資股票，研究公司財務報表，是首要任務，亦是必要工作。

#### 參考文獻

[1] 陳文勇，”台灣電子產業股價與經營績效-獲利能力之關聯研究”，大葉大學事業經營研究所碩士論文，2000 年。  
 [2] 申立銘，”台灣總體經濟指標與財務指標在選股策略上的應用—以上市、上櫃航

運類股為例”，國立海洋大學航運管理系碩士論文，2000 年。

[3] 蘇祐萱，”貝氏網路於輔助盈餘預估分析之研究”，元智大學資訊管理研究所碩士論文，2000 年。  
 [4] 施正宏，”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”，中原大學資訊管理研究所碩士論文，2003 年。  
 [5] 利怡玟，”台灣上市公司經營績效與股票報酬率之關聯性”，朝陽科技大學會計所碩士論文，2005 年。  
 [6] 鄭文英、李勝榮、葉憲弘，「台灣上市上櫃電子公司經營財務階段判別模式之建立」，風險管理學報第 8 卷第 1 期，2006 年。  
 [7] 詹場、李坤育、凌嘉駿、龔信銘，「台灣證券市場股價預測模型之建構—以總體經濟因素與公司財務指標為基礎」，國立臺北商業技術學院學報第 4 期，2003 年  
 [8] 林江亮、官月緞，「盈餘與股價報酬預測—關係人交易資訊之運用」，中華管理評論第 6 卷第 1 期，2003 年。  
 [9] 楊宗彥，”運用類神經網路與決策樹技術預測股票報酬率”，逢甲大學企業管理所碩士論文，2003 年。  
 [10] 吳月明，”股票報酬率預測模式績效之研究-倒傳遞類神經網路與灰預測之應用”，朝陽科技大學財務金融系碩士論文，2003 年。  
 [11] 葉怡成、林文盟，「上市公司間股價漲跌之關聯與預測—關聯探勘之研究」，商管科技季刊第 8 卷第 1 期，2007 年。  
 [12] 台灣經濟新報資料庫(TEJ)  
 [13] Han, J., Kamber, M. Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufmann Publishers. (2001)。  
 [14] Berry, J.M.A and Linoff, G., 「Data Mining Techniques: for Marketing, Sales, and Customer Support」, New York, John Wiley & Sons Inc., 1997.  
 [15] R. Agrawal, & R. Srikant, “Fast algorithms for mining association rules,” Proceedings of 1994 International Conference on Very Large Data Bases. Pp.487-499, 1994.

本研究感謝建國科技大學校內專題計畫補助，計畫編號：CTU97-RP-IM-007-053