

# 應用資料探勘於軟體業目標客戶的預測

鄭滄祥

南台科技大學企業管理系教授

cts@mail.stut.edu.tw

沈瑞祥

南台科技大學企業管理系研究生

stanley0229@gmail.com

## 摘要

軟體業對於目標客戶的預測上，礙於資料上蒐集不易的情況，對於目標客戶篩選，大都採用人為預估的方式，較難有既定的參考模型來預測客戶的再購意願，而一般零售業在對於目標客戶的預測上，較多採用顧客評價模式（RFM）的方法來預測目標客戶，本研究期望能在軟體業上應用零售業對目標客戶的預測模式，來進行目標客戶的預測，假若能依照零售業的方法，利用資料庫中的客戶購買記錄，來做為預估未來市場的決策依據，不僅可以避免掉因人為疏失所造成的客戶流失，還可將此應用擴大其使用範圍，來推測可能的銷售行為，進而開拓新客源的擴展。

本研究採回溯性研究法，個案收集期間為 2000 年至 2009 年，過濾掉一些有缺失遺漏的資料，共取得 281 筆銷售個案，此個案數便是本研究的探討重點。本文嘗試利用分類分析的方法，以零售業常用的顧客評價模式（RFM）為基礎，利用資料倉儲分析上常用的 C4.5 預測模式，搭配上 Weka 軟體的運算功能在所蒐集之 R、F、M 變數以及新增加的「銷售區域」變數中進行預測效能的評估，進而以決策樹建構出最具效能的預測模型，根據分析結果，所列的變數中，若以離散化後的 R、F 變數搭配上「銷售區域」變數是可以建構出一個可以應用在軟體業上之目標客群預測模型。

而透過本研究的分析結果，顯示決策樹模式是可以有效分類出可能再購的目標客群，以協助業務人員作為未來評估目標客群的重要參考依據。

**關鍵字：**創新管理與應用、目標客戶預測、分類分析、決策樹。

## 壹、緒論

在企業e化佈局的當下，若要能夠成為長期合作夥伴，首要完成項目便是讓客戶轉為忠誠客戶，為了做到這一件事，在銷售通路上的經營便顯得格外重要，然而，資訊軟體的使用者大部分為企業，顯少是一般消費者，相對在銷售上的複雜度也就變得更為棘手。

孫文秀(民82)曾提到軟體行銷通路開拓困難的原因有：(一)業務人員的銷售意願不高、(二)經銷商的進貨意願不高、(三)使用者購買的方便性等幾大項，根據一般銷售從業人員的經驗，以往在挑選最適客戶時，大多取決於個人經驗來過濾可能的潛在客戶，較無法採用傳統上常用的行銷研究方法，來取得各種客戶消費資訊，如問卷調查。

因此，利用資料庫取得顧客行為面的資訊，應是很好的替代方案。將購買紀錄存在資料庫中，這些資料是客戶過去決策結果的累積，其中亦隱含客戶對於產品的態度、偏好及行為訊息。倘若企業擁有完整的消費者交易紀錄，便可利用資料庫中這些代表消費者過去決策結果的訊息，來預測他們未來對各項行銷活動的可能反應；不僅如此，尚可擴大其應用範圍，用來推測資料庫外，具有類似行為模式消費者的可能回應，進行新客源開發的工作。

礙於軟體業的特性，少有文獻記載有關軟體業目標客戶的預測模式建構分析。根據文獻查證的結果，多數文獻採用 RFM 顧客價值評估方法進行顧客再購意願的探討，而 C4.5 決策樹建構法則是最常被用在建構分類預

測模式上。因此，本研究運用 RFM 顧客價值評估方法及 C4.5 決策樹演算法，再輔以蒐集的客戶軟體採購歷史記錄，建構出可用以預測顧客軟體再購行為的預測模式。

## 貳、 文獻探討

### (一). 顧客關係管理

近幾年來，為了強化客戶與企業之間的關係，許多企業不得不從客戶關係的角度開始思考，如何可以強化彼此之間的相依程度，所以客戶關係管理(Customer Relationship Management, CRM)是一個建造出與顧客的長期配合的經營策略關係，並且轉化成一套可以賺錢的系統(Ling & Yen, 2001)，因此良好的客戶關係管理不但可以幫助企業吸引到新的客群更讓現有客戶對公司持續有所貢獻。有鑑於此企業必須應用一些方法，有效地提高客戶關係，其中就包括了客戶關係管理，客戶價值分析等。此外，企業還必須透過建立良好的客戶關係管理強化市場營銷和銷售的成效。(Kalakota & Robinson, 1999)解釋說，在這樣的市場銷售策略上，客戶關係管理是整合功能相關領域，它通常利用資訊科技透過服務和技術來滿足客戶需求，以幫助企業提升與客戶的關係，進而有系統地，提高客戶忠誠度和增加整體經營利潤(Kalakota & Robinson, 1999)。而這種關係尤為明顯在服務部門(Ennew & Binks, 1996)。諸多學者對於客戶關係管理有不同的定義，本研究整理如表一。

表一：客戶關係管理定義

發表年份	學者	客戶關係定義
2001	Ling & Yen	客戶關係管理(CRM)包括一套流程，透過客戶關係管理(CRM)系統的建立來支持業務策略上長期的盈利與特定客戶的關係
2001	Swift	由企業的角度來理解和影響有意義的客戶行為，進而增加客戶，保留客戶，提昇客戶忠誠度和增加客戶盈利
2001	Parvatiyar Sheth	是一種如何保留和選擇合作客戶的策略，為公司和客戶創造卓越的價值這其中涉及到整合整個市場營銷面的問題，客戶服務機制和供應鏈中個職能的扮演，將能確保該公司在提供客戶價值上得以實現更高的效率和效益
2003	Kincaid	透過客戶關係管理(CRM)所取得的策略資訊，可協助公司管理者強化與客戶的關係在市場，銷售，服務和支援層面
2005	Ngai	成功的客戶關係管理(CRM)策略均建立在客戶數據和資訊科技工具基礎上，資訊共享和技術大大提昇了市場營銷機會，並改變了公司及其客戶之間的管理方式

### (二). 顧客價值評估分析

顧客價值評估模式的分析模型最早是由 Hughes 所提出 (Hughes, 1994 年)，這個模型主要區分三個客戶的重要屬性，分別為客戶的最近購買日、購買頻率、購買的貨幣價值。而若是將該三大變數應用預測顧客的消費行為，則 R (最近購買日) 是最強，其次是 F (購買頻率)，普遍來說，R&F 即解釋了 80% 客戶行為，M (購買的貨幣價值) 最弱，但一般而言，M 若應用在高單價或高服務品質的產品上則有很強的預測能力，可是若應用在低單價的產品上，則預測能力效能較差。故依照銷售預測能力來排列，會稱之為 R、F、M 模型。而本文中選定研究探討所選定的軟體業正好符合 M 變數中的高單價、高服務品質的特性，詳細定義的顧客價值評估模型描述如下：

最近購買日(R)：R 代表最近購買日，這是指最近消費行為的發生和目前時間的差距。若這個差距越小，則代表該客戶再次消費的機率增高。就軟體業而言，本研究將根據現實狀況，以年為單位計算客戶的最近購買日。

購買頻率(F)：F 代表購買頻率，這個變數記錄了顧客在某特定期間內購買產品的數量。本研究仍以年為單位，記錄客戶每年採購 E 化軟體的套數。

購買產品的價值(M)：M 代表購買產品的總價值，亦指在某特定期間顧客的總消費金額。

根據文獻 (Wu & Lin,2005)，研究表示 R 和 F 值越小而 M 該值越大，越有可能是滿足客戶要購買的產品或服務的企業了。所以顧客價值評估模式的方法是非常有效的將客戶屬性加以細分 (Newell, 1997)。

理論上，RFM 變數中，每個變數的權重會依產業特性不同而分別定義 (Marcus, C., 1998)。一般而言 RFM 變數權重會依照產業的特性不同而有不同的定義而為求較為精準的資訊來加以分析，國內外諸多學者透過 RFM 的顧客評價模式來達到預測市場的目的，尤其是在電信業、金融業、保險業、零售業與製造業等。

透過 RFM 模型不但可以應用在記錄消費者行為的資料庫，對於企業用戶的資料庫也很有預測效果，例如美國聯邦快遞公司 (Federal Express)，即利用 RFM 模型來達到區隔企業客戶，獲取更高利益。(Arthur Hughes, 1994)。如 Sung and Sang (1998) 將其應用在觀光旅館業者上，目的是要為了增加商店的銷售量，提昇顧客滿意度，其作法主要是將收集來的客戶交易資料以 RFM 的規則加以轉換後再透過資料探勘中 C4.5 的分類分析技術來剖析每個群組之間的分類法則，除將客戶區分為多個群組外並針對每個群組制定不同的行銷策略，進而來增加免稅商店的營業額；梁珍瑜 (2003) 則是以超級市場的產品銷售紀錄與顧客會員資料為主要研究範圍，而其顧客價值評估模式則是結合上對專家的訪談來決定，基於以上兩項基礎下進行顧客分群分類研究，其過程中再透過權重的建構，依據顧客價值評估模式所呈現的積分結果將顧客區分為四大群組；楊清潭 (2003) 依據顧客價值評估模式屬性，鎖定醫療器材業以及健康檢查服務業兩大族群的銷售顧客採用百分比方式將顧客區分成 10 群。除以上所列之外，如把 RFM 模型應用於汽車維修業 (劉世琪, 2003)、電子業 (林柏甫, 2003)、流通業及製造業 (游濬遠, 2003) 等之研究，均是採用 RFM 的預測分析模型，來設計出符合客戶差異化的行銷策略，進而透過釐清不同顧客需求來增加對顧客的熟悉度，提高客戶的忠誠度與滿意度，進而建立起與顧客的良好關係，讓客戶成為公司的長期合作夥伴。

### (三). 分類分析

資料探勘乃是運用各種不同的資料分析方法、儘可能地從既有的大量資料中發掘出事前未知、有趣並可用以提升企業行動力的知識 (Berry, 1997)。根據資料分析方式以及所得之分析結果的不同，資料探勘技術可被區分為分類分析 (Classification Analysis)、群集分析 (Clustering Analysis)、聯結法則分析 (Association Rule Analysis)、次序相關分析 (Sequential Pattern Analysis) 及鏈結分析 (Link Analysis) 等五類 (張勳騰, 1999；邱義堂, 2001；彭文正, 2001)。

分類分析著重於從已知類別的物件集合中，依據物件具有的屬性 (也就是可能影響物件類別的變數) 建構出一個可用以描述物件屬性與類別間相關性的分類模式 (如決策樹或決策法則)，用以協助對未經分類的資料進行類別歸屬的預測工作。用來尋找與建構分類模式的已分類資料可以是來自現有的企業內部歷史性資料，或由企業蒐集可分類之物件基本資料 (Profile)。

分類分析是一種根據資料屬性決定物件所屬類別的過程，用以產生「若則」的物件歸屬法則。倘若顧客忠誠度 (設分為高忠誠度顧客與低忠誠度兩類) 的推估可以利用顧客基本資料中的居住區域 (市區、市郊或鄉鎮)、年紀 (小於 21、21 至 30 或大於 30)、婚姻狀況 (未婚或已婚) 以及性別等屬性加以評定，企業就可收集一組分屬不同忠誠度類別的顧客基本資料以分類分析建構顧客的居住區域、年紀、婚姻狀況及性別等屬性與所屬忠誠度類別間的忠誠度類別預測模式，用以預測其他顧客可能所屬之忠誠度類別。可產生決策樹或決策法則的 ID3、C4.5、CN2 與 AQ15 以及倒傳遞類神經網路 (Backpropagation Neural Network) (Wei & Shaw, 2003; Quinlan, 1986; Han & Tang, 1995; Ha & Park, 1998; Clark & Niblett, 1989) 都是常見分類分析技術。由於決策樹分類模式是以樹狀資料結構為基礎的分類分析方法，其所產生的分類模式具有簡單明瞭以及便於人類瞭解與運用的特性，故決策樹的建構技術是最常被使用的分類分析技術。決策樹的建構是利用監督式的學習法從訓練範例集合中，以適當屬性挑選函數，從訓練範例的屬性中挑選出可用以建構決策樹根節點 (root node) 及內部節點 (internal nodes) 的屬性，用以建構決策樹並對訓練範例進行區分的處理。基本的決策樹建構演算法如下：

1. 從設定決策樹的根節點(Root Note)開始，設定根節點為目前節點  $C$ ，此時所有訓練範例(或稱之為訓練物件)將都屬於  $C$  的物件集合。若  $C$  中所有的物件都屬於同一類別(Class)，則將此類別設定為  $C$  之決策結果，然後停止，否則繼續執行步驟 2。
2. 對  $C$  中所有的物件，針對所有尚未出現在從根節點到目前節點路徑的所有屬性  $A_i$ (稱之為候選屬性)，利用屬性挑選函數挑選出一個具有最佳類別區分力的屬性  $A_c$  作為節點  $C$  的分類屬性。
3. 在節點  $C$ ，根據所挑選出的屬性  $A_c$  的  $m$  個屬性值(假設所選定的分類屬性有  $m$  個屬性值)，在節點  $C$  下建立子節點  $C_1$ 、 $C_2$ 、...、 $C_m$ ，並根據分類屬性值將  $C$  中的所有物件分派至適當的子節點中。
4. 將每個子節點  $C_i$  當成目前節點  $C$ ，並分別由步驟 1 繼續執行決策樹的建構。

在現有的決策樹方法中，ID3 與 C4.5 最常被採用兩種決策樹技術。C4.5 決策樹學習法是 Quinlan 於 1993 年提出，用以改良他在 1986 年所提出的 ID3 決策樹學習系統(Quinlan, 1986; Quinlan, 1993)。ID3 與 C4.5 兩項演算法的差異在於：ID3 決策樹僅能處理離散型態的類別性資料，而無法自動地處理包含連續性數值的資料。C4.5 則可以在建構決策樹過程中，彈性地對連續性數值的資料進行離散化的處理，不需以特定的標準事先將數值資料進行切割，並以適當的符號予以代替。因 C4.5 比 ID3 在處理數值資料上具有彈性，應用的範圍也更廣。因此本研究將採用 C4.5 技術建構目標客戶的預測模式。

## 參、 實證評估

### (一). 資料的蒐集與整理

本文採取回溯性研究的方式，自台灣某知名資訊軟體服務公司進行資料收集。本研究共蒐集得該公司自西元 2000 年至 2009 年 10 月的軟體銷售的歷史記錄，共可取得約 500 筆案例。本研究將鎖定於該公司中一項「中高階主管的決策支援軟體」的再購預測研究。經進行資料前處理後，並刪除具有資料遺漏問題的案例，共可獲得 281 筆可供研究的完整案例。

審視所收集的 281 筆資料，分析客戶於 2009 年的軟體採購情形。其中共有 197 位客戶曾採購「中高階主管的決策支援軟體」，將被歸類為 Y 類別，其餘未採購該軟體的 84 筆資料則被歸類為 N 類別。所收集的 281 筆資料的 R、F、M 變數的敘述性統計特徵將如表二所述。

表二：軟體購買案例的 RFM 變數統計特徵

評價變數	變數說明	內容值	統計特徵	說明
R	最近購買日	整數值	值域：1~8 平均數：1.25	若要預測顧客於 2009 年的再購可能，而該顧客最近採購軟體的年度為 2007 年，則 R 將被編碼為 2
F	購買軟體總數	整數值	值域：2~11 平均數：4.86	2000-2009 年間購買的軟體總數
M	總購買金額	連續值	值域：28萬~4354萬 平均值：約1045萬 標準差：約719萬	2000-2009 年間購買軟體的總金額

### (二). 模式評估程序與準則

本文將採用十摺交叉驗證法(10-fold Cross Validation)，針對建構出的分類模式進行效能評估。所謂的十摺交叉驗證法係將資料集合，隨機分成十份大小相仿的子資料集，取其中的九份子集作為建立模式的訓練資料，而以另一份子集作為驗證模式效能的測試資料。如此重複進行 10 次(10 摺)的實證實驗，讓每一份子集都能輪流作為測試資料，再將十次實驗所測得的模式效能進行平均後，用以代表實際分類模式之分類效能。十摺交叉驗證法的作法步驟如下：

1. 將樣本隨機分成十份
2. 自十份樣本中取其中的九份資料作為建立模式的訓練資料
3. 剩下的一份資料作為驗證模式效能的測試資料
4. 反覆進行 10 次 (10 摺) 預測，每一份資料均會輪為測試資料，藉由此分類運算方式，所分別得出的十組分類的正確率，再加以平均後所得到的值即為該預測模式的正確率。

本文把在 2009 年再購公司軟體的客戶歸類為目標顧客(Y 類別)，而將未於 2009 年再購的客戶歸類為非目標顧客(N 類別)，並以召回率(Recall)、準確率(Precision)、F 平均分數(F-measure)作為目標顧客預測模式的效能評估準則。若利用表三的分類結果矩陣(Kohavi and Provost, 1998)進行說明，A 代表模式準確預測出目標顧客的案例個數；B 代表將目標顧客誤認非目標顧客的案例個數；C 代表將非目標顧客誤認目標顧客的案例個數；D 則是模式準確預測出非目標顧客的案例個數。Recall 則表達出預測模式可以正確預測出目標顧客的比例，Precision 則表達出模式對目標顧客的正確預測比例，F-measure 則衡量模式在 Recall 及 Precision 兩項效標上的平衡水準。根據表三，Recall、Precision、以及 F-measure 的計算方式分別為：

$$\text{Recall} = \frac{A}{(A+B)} \quad \text{Precision} = \frac{A}{(A+C)} \quad \text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

表三：分類結果矩陣

單位：個數	預測為目標客戶(Y)	預測為非目標客戶(N)
實際為目標客戶(Y)	A	B
實際為非目標客戶(N)	C	D

### (三)． 預測模式的建構

本研究利用 Weka 探勘軟體中的 C4.5 模組，以表二所列出的 R、F、M 變數進行模式的建立，同時利用十摺交互驗證的程序，評估建構出之預測模式的效能。對建構出模式的評估結果將如表四所示：

表四：利用 R、F、M 所建構出預測模式的效能

Recall	Precision	F-measure	Class
99.5%	70.3%	0.824	Y
1.2%	50.0%	0.023	N

由表四的評估結果可知，利用 R、F、M 所建構出的決策樹模式，幾乎對 N 類案例不具識別效能，幾乎將所有案例皆預測為 Y 類別。這樣的模式雖然具有良好的辨識正確率，但卻不具任何預測價值。

### (四)． 對 M 變數的處理

由於利用 R、F、M 變數建構出的決策樹模式不具有有效的預測效能，因此本研究將對 R、F、M 變數進行處理，期望藉以提昇預測模型的效能。依照過往的銷售經驗，M 變數為首要關鍵影響因子，故本文優先自購買金額 (M) 開始進行離散化修正。

因為企業購買軟體的金額頗高，且總金額的分布範圍較廣，因此本文將購買金額(M)變數以 500 萬為一級距，進行離散化。將 M 變數內容以 500 萬至 4500 萬離散化為如表五的 9 個級距，並產生一個新的類別變數 M1。

由於 M 變數及 M1 變數均描述客戶於 2000 年至 2009 年間軟體的總購買金額，因此本文也針對本研究所探討的進行 10 年間的購買金額分析，參照 M1 變數的離散化方式，以 50 萬為劃分級距，共可將「中高階主管的決策支援軟體」的購買費用自 50 萬到 550 萬分成 11 個級距，並將此針對「中高階主管的決策支援軟體」的產品定義為「特定產品購買金額」(M2) 變數。M2 變數的內容編碼規則將如表六所示。



表五：M1 變數編碼原則

編碼	變數值範圍
1	低於5百萬(含)
2	5百萬至1千萬(含)
3	1千萬至1千5百萬(含)
4	1千5百萬至2千萬(含)
5	2千萬至2千5百萬(含)
6	2千5百萬至3千萬(含)
7	3千萬至3千5百萬(含)
8	3千5百萬至4千萬(含)
9	4千萬至4千5百萬(含)

表六：M2 變數編碼原則

編碼	變數值範圍
1	低於50萬(含)
2	50萬至100萬(含)
3	100萬至150萬(含)
4	150萬至200萬(含)
5	200萬至250萬(含)
6	250萬至300萬(含)
7	300萬至350萬(含)
8	350萬至400萬(含)
9	400萬至450萬(含)
10	450萬至500萬(含)
11	500萬至550萬(含)

根據評估的結果，將 M 變數以 M1 變數或 M2 變數取代都無法有效地提升預測模式的效能，因此 M 類型的變數對於軟體目標顧客的預測似乎沒有太大的影響力。然而變數 M1 與 M2 的解讀將比變數 M 更為容易，因此後續的實驗將保留這兩個變數的使用。

#### (五). 對 R 變數的處理

與資深業務人員詳談後，礙於資料蒐集的困難，已經無法在 M 變數再進行細分的資訊萃取處理，因而轉向對變數 R 與變數 F 變數的再處理，試圖找出可提升預測模式效能的影響變數。與專家訪談後的結果，發現目前所收集的資料僅能讓本研究從變數 R 中再萃取出「與客戶關係的維繫程度」的資訊，因此本研究將再增加一個變數 R1，呈現顧客「近三年」是否與公司有交易的關係。利用 R、R1、F、M1 與 M2 等變數所建構的預測模式效能如表七所示。由表七的評估結果可知，新增加的 R1 變數可以有效地提高預測模式在 N 類別的預測效能。

表七：利用 R、R1、F、M1 與 M2 所建構出預測模式的效能

Recall	Precision	F-measure	Class
88.8%	70.3%	0.785	Y
11.9%	31.3%	0.172	N

#### (六). 增加「銷售區域」的描述變數

除了上述與 R、F、M 相關的變數外，專家也提醒台灣的軟體銷售狀況在南、北區域將所不同。經

查閱銷售記錄後，發現銷售紀錄中記載的銷售區域分為北部與中南部地區。因此本研究將加入一個描述軟體「銷售地區」的變數 Area，其內容值則區分為北區與中南區。加入變數 Area 後，所建構出的預測模式效能如表八所示。比較表七與表八的結果，可發現加入變數 Area 後，不僅可讓預測模式在 N 類別的 F-measure 由 0.172 提升至 0.293，還讓預測模式 Y 類別的 F-measure 由 0.785 提升至 0.802。因此，變數 Area 的確可以有效地提升預測模式的效能。

表八：加入 Area 變數所得的預測模式效能

Recall	Precision	F-measure	Class
89.3%	72.7%	0.802	Y
21.4%	46.2%	0.293	N

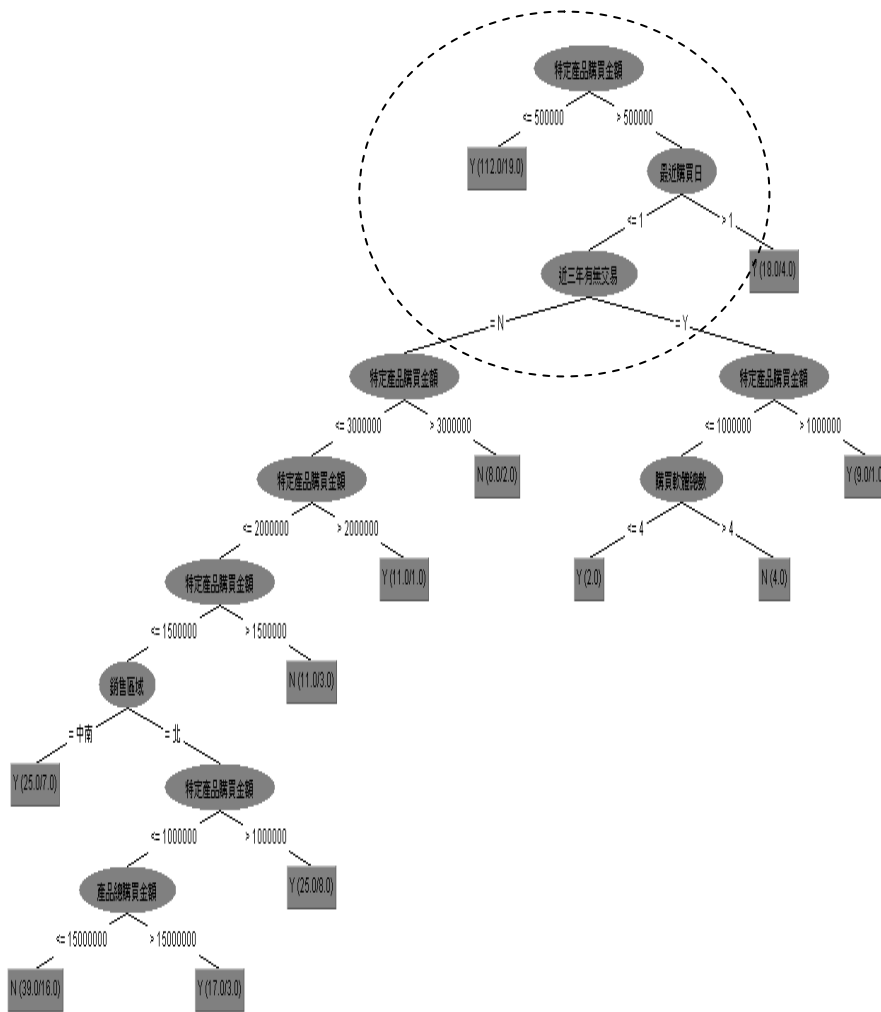
(七). 目標顧客預測規則的萃取

根據上述的變數處理過程，本研究共可獲得六個變數，將如表九所列。利用表九所列的變數，建構出的決策樹模式將如圖一所示。由圖一的決策樹的樹狀結構圖可知，本研究可利用「最近購買日(R)」、「近三年有無交易(R1)」、「購買軟體總數(F)」、「產品總購買金額(M1)」、「特定產品購買金額(M2)」、「銷售區域(Area)」等變數提供預測規則給軟體銷售人員參考。

利用圖一的決策樹，本研究共可萃取出 8 條可預測顧客再購行為的規則，以及 4 條可辨別顧客不會再購行為的規則。圖一決策樹中的每個葉節點，為了能夠充分表達分類至該葉節點的個案數本文將其所產生的結果以「個案數 S/分類錯誤數 E」的形式加以描述。

表九：建構決策樹預測模式的變數

變數	變數說明	內容值	統計特徵	說明
R	最近購買日	整數值	值域：1~8 平均數：1.25	預測再購的年度為 2009 年，而顧客最近採購軟體年度 2007 年，則 R 將編碼為 2
R1	近三年有無交易	Y、N	Y：20、N：261	2006 年至 2008 年三年內，任一年度有交易即為 Y，否則為 N
F	購買軟體總數	整數值	值域：2~12 平均數：4.86	10 年內購買的軟體總數
M1	產品總購買金額	連續值	值域：500 萬~4500 萬	
M2	特定產品購買金額	連續值	值域：50 萬~550 萬	
AREA	銷售區域	北、中南	北：209 中南：72	



圖一 決策樹結構圖

而為了能夠分析各條預測規則的重要性，本文採用關連法則中常用的「支持度(support)」與「信賴度(confidence)」的觀念計算出每條規則的「支持度」與「信賴度」，計算方式描述如下：

1. 支持度：支持度的計算公式為分子是指分類至該節點的個案數 S，而分母則為整個資料集的個數 N，透過  $S/N$  所呈現的值來表示支持度
2. 信賴度：信賴度的計算公式為分子是指該節點的正确分類個案數(S-E)，而分母則為該節點的個案數 S，透過  $(S-E) / S$  所呈現的值來表示信賴度。

以圖一中最簡潔的一條規則為例(圖一圓形虛線部分)，被分類至該節點的個案數為 112 個( $S=112$ )，其中分類錯誤的個案數則有 19 個( $E=19$ )；用以產生此決策樹的資料集 C，共包含了 281 筆個案( $N=281$ )。因此，該條規則的「支持度」為  $112/281=0.399$ 、「信賴度」為  $(112-19)/112=0.83$ 。

由圖一的結果得知，萃取出來的 12 條規則共包含了 8 條客戶再購的規則以及 4 條客戶不會再購的規則。若以考量平均規則的支持度 $>0.08$  而言(共 12 條規則，平均每一條支持度為  $281/12=23$ ， $23/281=0.08$ )，大部分的規則的支持度均低於 0.08，故描述 Y 類的一條規則，節點個案數 Y (112.0/19.0)，支持度 0.396，信賴度 0.83(圖一矩型實線部分)，以及描述的 N 類的一條規則，節點個案數 N (39.0/16.0)，支持度 0.139、信賴度 0.59 (圖一矩型虛線部分) 最具有參考價值。其規則如下：



Y 類別：當特定產品貢獻金額  $\leq 50$  萬時，節點個案數為 112/19，支持度為 0.399，信賴度為 0.83

N 類別：特定產品貢獻金額  $> 50$  萬 And 最近購買日  $\leq 1$  And 近三年有交易 = 'N' And 銷售區域 = '北' And 特定產品貢獻金額  $\leq 1$  百萬 And 產品總貢獻金額  $\leq 1$  千 5 百萬時，節點個案數為 39/16，支持度 0.139、信賴度 0.59

透過以上規則，在實地與資深業務人員訪談發現，透過此規則的分類，將有助於業務在找尋目標客戶時能夠有明確的劃分原則，這在過往的行銷程序上是有很大的助力，過往因為沒有這樣的分類規則，只能仰賴業務人員的經驗判斷，而藉由此規則的制定，不但可以減少人為的誤判，還可以大幅提昇業務人員作業效率，提昇客戶的再購率。

## 肆、 結論

一個公司的獲利需求主要來自於新顧客與持續購買的兩大類顧客，若企業能將顧客維持率提高達 5%，將可獲得 25% 到 85% 的利潤成長。隨著軟體業的快速蓬勃發展，資訊科技的應用將成為企業的核心競爭能力，而有效的應用資訊技術將可強化企業對市場的敏感度、提高決策品質強化與顧客間的良好關係。

有鑑於此，若能善加利用資料探勘的資訊技術，從公司多年經營的大量歷史記錄資料中，找出客戶的決策結果以及公司決策者對市場經營的決策過程，進而建構出有效的決策模式，不但能提昇決策者的決策品質，更可以透過模式的建立，找出客戶的消費習性，為公司帶來更大的利潤。

## 參考文獻

- [1] Berry, M. J. A. and Linoff, G. S., (1997), "Data mining techniques: for marketing, sales, and customer support." John Wiley & Sons, Inc.
- [2] Cheng, C. H. and Chen, Y. S., (2009), "Classifying the segmentation of customer value via RFM model and RS theory," *Expert Systems with Applications*, Vol. 36, pp.4176-4184.
- [3] Clark, P. and Niblett, T., (1989), "The CN2 Induction Algorithm," *Machine Learning*, Vol. 3, pp.261-283.
- [4] Chen, Y. L., Kuo, M. H., Wub, S. Y. and Tang, K., (2009), "Discovering recency, frequency, and monetary (RFM) sequential patterns from customers purchasing data," *Electronic Commerce Research and Applications*, Vol. 8, pp.241-251.
- [5] Han, J., Fu, Y. and Tang, S., (1995), "Advances of the DBLearn System for Knowledge Discovery in Large Databases," *Proc. of 1995 Int'l Joint Conf. on Artificial Intelligence (IJCAI'95)*, Montreal, Canada, Aug, pp.2049-2050.
- [6] Ha, S. H. and Park, S. C., (1998), "Application of Data Mining Tools to Hotel Data Mart on the Intranet for Database Marketing," *Expert Systems With Applications*, Vol. 15, pp.1-31.
- [7] John, A. M. and Hastak, M., (2007), "Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression," *Journal of Business Research* Vol. 60, pp.656-662.
- [8] Kalakota, R. and Robinson, M., (1999), "e-Business roadmap for success (1st ed.)," *New York, USA: Addison Wesley Longman Inc.*, pp.109-134.
- [9] Lee, J. H. and Park, S. C., (2005), "Intelligent profitable customers segmentation system based on business intelligence tools," *Expert Systems with Applications*, Vol.29, pp.145-152.
- [10] Liua, D. R. and Shiha, Y. Y., (2005), "Integrating AHP and data mining for product recommendation based on customer lifetime value," *Information & Management*, Vol.42, pp.387-400.
- [11] Ngai, E. W. T., Xiu, L. and Chau, D. C. K., (2009), "Application of data mining techniques in customer

- relationship management: A literature review and classification,” *Expert Systems with Application*, Vol.36, pp.2592-2602.
- [12] Peppard, J., (2000) , “Customer relationship management (CRM) in financial services.,” *European Management Journal*, Vol.18(3), pp.312-327.
- [13] Quinlan, J. R., (1986) , “Induction of Decision Tree,” *Machine Learning*, Vol. 1, pp.81-106.
- [14] Quinlan, J. R., (1993) , “Morgen Kaufmann Publishers, San Mateo, CA.,” *C4.5: Programs for Machine Learning*.
- [15] Seyed, M. H., Maleki, A. and Gholamian, M. R., (2010) , “Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty,” *Expert Systems with Applications* ,Vol.37,pp.5259-5264.
- [16] Wei, C., Piramuthu, S. and Shaw, M. J., (2003) , “Knowledge Discovery and Data Mining,” Chapter 41 in *Handbook of Knowledge Management*, Vol. 2, C. W. Holsapple (Ed.), Springer-Verlag, Berlin, Germany, pp.157-189.
- [17] 邱義堂，2001。通信資料庫之資料探勘：客戶流失預測之研究。國立中山大學資訊管理研究所碩士論文。
- [18] 孫文秀，1993。如何打開軟體行銷通路。資訊與電腦。
- [19] 許哲璋，2002。資料挖掘與統計方法應用於資料庫行銷之實證研究—以美妝保養品業為例。國立台北大學碩士論文。
- [20] 彭文正譯，2001。Michael J.A. Berry 以及 Gordon S. Linoff著，Data Mining資料採礦 客戶關係管理暨電子行銷之應用。數博網資訊股份有限公司。
- [21] 連惟謙，2004。應用資料分析技術進行顧客流失與顧客價值之研究。中原大學。
- [22] 張勳騰，1999。通信資料庫之資料探勘：目標行銷之應用」。國立中山大學資訊管理研究所碩士論文。
- [23] 薛新光，2006。運用資料探勘於資訊產品變裝較佳化組合設計之研究。大葉大學。
- [24] 劉宛晶，2005。植基於RFM分析法之產品購買週期的適時性推薦方法。朝陽科技大學。