

東 海 大 學
工業工程與經營資訊研究所

碩士論文

應用資料採礦技術於我國壽險業顧客價值
分析之研究



研究生：蕭乃綺
指導教授：洪堯勳 博士

中華民國九十五年七月

Application of Data Mining to Customer Value Analysis of Taiwan's Insurance Industry

By
Nai-Chi Hsiao

Advisor: Prof. Jan-Shin Hon

A Thesis
Submitted to the Institute of Industrial Engineering and Enterprise
Information at Tunghai University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in
Industrial Engineering and Enterprise Information

July 2006
Taichung, Taiwan, Republic of China

應用資料採礦技術於我國壽險業顧客價值分析之研究

學生：蕭乃綺

指導教授：洪堯勳博士

東海大學工業工程與經營資訊研究所

摘要

我國保險產業主要以人身保險為主，其特性在於以人為商品標的物，而人的需求不斷改變，複雜地近乎沒有規律。再加上近期政府甫通過之勞退新制，益愈擴大了原本就存在於業者與顧客間之資訊差距。壽險業者雖不斷推出新產品以為因應，但其經營方式多利用業務員之人際網絡開發新顧客。然若業者能依據 80/20 法則善用既有資料，則可從其中挖掘出既有顧客之潛在需求，並達投保率之提升與新顧客之開發。

本研究以人身保險產業為研究對象，利用資料採礦對個案金融控股公司之人身保險資料庫進行分析，首先將資料採礦與顧客價值之概念結合，以建立研究架構。其次，將初始的資料透過格式轉換與定義，轉變為分析所需要的結構化資料，接著利用 RFM 理論為顧客價值指標以概念區分顧客等級，並以資料採礦的決策樹演算法來驗證顧客等級之分類與欄位定義模式；最終，將決策樹驗證後之分類模式定義投入關聯法則 Apriori 演算法，以萃取出各顧客價值層級之潛藏法則。

本研究應用關聯法則於決策樹已驗證之分層定義，分別針對四個顧客價值層級作法則萃取，研究結果顯示，第一、四層級法則多等同於直觀判斷之特徵，惟第一層級之險種類別理出關聯；第二、三層級於保戶特徵及保單特徵上則皆具表現，未來壽險人員可應用該些法則研擬二次行銷策略，並以此作為新客戶開發之依據。

關鍵字詞：顧客關係管理、資料採礦、決策樹、關聯法則、保險業

Application of Data Mining to customer value analysis of Taiwan's Insurance Industry

Student: Nai-Chi Hsiao

Advisor: Prof. Jan-Shin Hon

Department of Industrial Engineering and Enterprise Information
Tunghai University

ABSTRACT

The insurance industry of Taiwan focuses on the product of Life Insurance which's most obvious character is using human being to be product target. However, people's needs change all the time and there is no rule to describe it. Additionally, government had approved a bill called Monthly Contribution Wages Classification of Labor Pension recently, and that can also make the increase of information gap which exists between customers and insurance companies. In order to solve this kind of problem, Life Insurance companies update products continually to custoermize for people' s need. The way most insurance companies use for contracts developing is by sales' relations. Nevertheless, according to 80/20 principle, if companies can utilize the dada which have be stored, they will find out the customers' lantent needs and use resources efficiently to maintain the high value customers.

This methodology aims to show that Data Mining can act as the substantial evidence for examining the data of insurance industry. It focused on the study of Life Insurance industry and analyze with the data which was from the database of case Financial Holding Co., Ltd. First, we will be combining the concept of data mining and customer relationship management in building the main structure of this study. Transforming the row data obtained through data switching and data identification into material which are good to be read by both analysts and the system is the second step. Using RFM Theory to decide which value level the customer is in and mark off every different level. After that, we will validate the clustered definition of customer's value level by the algorithm of Decision Tree and use the validated definition to the following. Finally, we will calculate out the most significant rules of every customer's value level for this company by Apriori algorithm.

According to Association rules, the result of this study reveals that the characters of level one and level four machth with the intuition, except for the kind of insurances which the customers of level one have bought. The characters of customers and contracts in level two and level three are obviously. Sales of Life insurance companies can apply these rules to second marketing strategy making and contract developing.

Keyword : Customer Relationship Management 、 Data Mining 、 Decision Tree 、 Association Rule

誌謝

本論文得以順利完成，首先要感謝的是學生的指導教授—洪堯勳博士，兩年來悉心指導與鼓勵，使學生在學術研究學習上深獲啟發，在情緒收納上亦是受益匪淺。

此外，由衷感謝口試委員黃開義博士與張炳騰博士不吝指正，在論文詳加閱讀後予以學生建議，使論文得以趨於完備。在學期間亦承蒙黃欽印老師在專業領域的教導，也感謝姚銘忠老師的不斷打氣，使學生在研究所的兩年中有許多的啟發與觸動。

特別感謝鄭大哥在人壽保險資料的提供與協助，使本研究得以順利完成。更要向研究室所有成員在這段時間的協助與關心說聲感謝，謝謝咬輪總是在我遇到瓶頸的時候引導我已經糾結的思路和情緒，謝謝管家、小他、阿賢、松竹與敏智學長在畢業後的關懷和打氣，謝謝小煌、小卡、建竣、泓毅給我這段相互勉勵甘苦與共的記憶，你們是最完美的夥伴。當然，還要謝謝永順與為任總是仔細地花時間與我討論問題，謝謝志銘不厭其煩地在軟體裝載方面給予協助，謝謝于倫陪我散步還有佳穎老載我兜風，謝謝采佩、momo、阿龔、孟儒貼心的微笑，謝謝子麟搞笑式的關懷，給我這一路滿滿的感激與感動。

更要感謝我的爹娘兄姊在求學的這段時間在精神上和經濟上給予無條件的鼓勵與支持，感謝大法予我的智慧，沒有這些就無法成就今天的我。謝謝芹、氫、君、莉、國慶、員魁、育霖、智浩、世琦、柏儀、佳珮、立翔、橘子、昕煒、智鈞等摯友在這段時間的陪伴與鼓舞，化作我強大的後盾。

最後，僅以這本論文獻給我所摯愛的家人與朋友，謝謝你們。

乃綺 謹誌於

東海大學工業工程與經營資訊研究所

價值鏈創新與管理研究室

中華民國九十五年七月

目錄

論文摘要.....	I
誌謝.....	III
目錄.....	IV
表目錄.....	VI
圖目錄.....	VII
第一章 緒論.....	1
1.1 研究背景與動機.....	1
1.2 研究問題與目的.....	2
1.3 研究流程.....	3
1.4 研究範圍與限制.....	4
1.4.1 研究範圍.....	4
1.4.2 研究限制.....	4
第二章 文獻探討.....	6
2.1 保險業.....	6
2.1.1 保險產品的分類.....	6
2.1.2 保險產業之現況與未來發展.....	7
2.2 顧客價值.....	10
2.2.1 顧客價值之定義與方法.....	11
2.2.2 顧客價值指標 RFM 模式.....	15
2.2.3 顧客價值於顧客關係管理之步驟.....	16
2.3 資料採礦.....	17
2.3.1 資料採礦的定義.....	17
2.3.2 資料採礦的功能.....	18
2.3.3 資料採礦的方式.....	19
2.3.4 資料採礦與統計方法.....	21
2.3.5 資料採礦的流程.....	24
2.3.6 資料採礦於保險業之應用.....	25
第三章 研究方法.....	28
3.1 研究工具.....	28
3.2 研究架構.....	28
3.3 資料蒐集處理與分析.....	29
3.3.1 資料蒐集.....	29
3.3.2 資料處理.....	30
3.4 RFM 顧客價值指標.....	31
3.5 決策樹學習法.....	31
3.5.1 C4.5 演算法.....	34
3.6 關聯法則.....	36

3.6.1 關聯法則分析之演算法.....	38
3.6.2 Apriori 演算法	38
第四章 研究分析.....	42
4.1 目標資料處理與敘述統計分析.....	42
4.1.1 目標資料轉換.....	42
4.1.2 敘述性統計分析.....	44
4.2RFM 顧客價值分群及決策樹驗證	50
4.2.1RFM 顧客價值分群	50
4.2.2 決策樹驗證分群.....	53
4.3 關聯法則挖掘及策略建立.....	58
4.3.1 資料前處理.....	59
4.3.2 資料後處理與法則分析.....	59
第五章 結論與建議.....	68
5.1 結論.....	68
5.2 未來研究與建議.....	71
參考文獻.....	72

表 目 錄

表 2.1 壽險業保險契約保額與件數統計.....	8
表 2.2 壽險業保費收入統計.....	9
表 2.3 新契約保費收入概況.....	10
表 2.4 顧客價值定義彙整.....	12
表 2.5 實施顧客價值的各種方法與步驟.....	14
表 2.6 資料採礦分析法與傳統統計分析法之比較.....	24
表 2.7 資料採礦在保險業之相關研究.....	25
表 3.1 決策樹演算法比較表.....	33
表 4.1 保單基本資料表.....	42
表 4.2 保戶基本資料表.....	43
表 4.3 保戶之年齡別、性別與婚姻別樞紐分析表.....	45
表 4.4 保戶繳別與職業別之樞紐分析表.....	46
表 4.5 保額、主附約與險別分類之樞紐分析表.....	47
表 4.6 保額、主附約與險別特性之樞紐分析表.....	48
表 4.7 保戶屬性與保單屬性交叉分析表.....	49
表 4.8RFM 指標欄位分數定義表.....	51
表 4.9RFM 指標分數表.....	52
表 4.10RFM 指標等級分布.....	53
表 4.11 決策樹之混淆矩陣分析表(Train).....	56
表 4.12 決策樹之混淆矩陣分析表(Validate).....	57
表 4.13 欄位變數列表.....	58
表 4.14 重量層級關聯法則表.....	61
表 4.15 中量層級關聯法則表.....	62
表 4.16 潛力層級關聯法則表.....	64
表 4.17 被動層級關聯法則表.....	65

圖 目 錄

圖 1.1 本研究流程圖.....	3
圖 2.1 保險結構圖.....	7
圖 2.2 顧客價值的三角模式.....	11
圖 2.3 顧客價值立方體.....	13
圖 2.4 顧客價值圖.....	14
圖 2.5 顧客關係管理的循環過程.....	16
圖 2.6 顧客關係管理之運作流程.....	17
圖 2.7 資料採礦技術發展歷程.....	22
圖 2.8 資料採礦組成結構圖.....	23
圖 2.9 資料採礦流程.....	24
圖 3.1 本研究資料採礦架構.....	29
圖 3.2 資料庫中的資料內容.....	30
圖 3.3 資料初步整理之 SPSS 介面.....	30
圖 3.4 決策樹的構造.....	31
圖 3.5 利用事後修剪法修剪決策樹.....	33
圖 3.6 高頻項目組之產生過程(第一次掃描).....	40
圖 3.7 圖高頻項目組之產生過程(第二次以後的掃描).....	40
圖 3.8 最後所得高頻項目組.....	41
圖 4.1SPSS 資料處理介面.....	43
圖 4.2 保戶之年齡別、性別與婚姻別樞紐分析圖.....	44
圖 4.3 保戶繳別與職業別樞紐分析圖.....	46
圖 4.4 保額、主附約與險別分類之樞紐分析圖.....	47
圖 4.5 主附約、保額與險別特性之樞紐分析圖.....	48
圖 4.6 保額、主附約與險別特性之樞紐分析圖.....	49
圖 4.7 顧客金字塔示意圖.....	52
圖 4.8RFM 指標分數分布圖.....	53
圖 4.9SAS Enterprise Miner 執行介面.....	54
圖 4.10Result of EM SAS Decision Tree 介面.....	55
圖 4.11 EM SAS 之決策樹混淆矩陣介面.....	56
圖 4.12 混淆矩陣示意圖.....	57
圖 4.13 EM SAS 決策樹 Output 報表之重要欄位介面圖.....	58
圖 4.14 資料區分圖.....	59
圖 4.15Weka 關聯法則介面.....	60

第一章 緒論

1.1 研究背景與動機

我國保險產業發展迄今約已五十多年，市場由人身保險與產物保險所劃分，2004 年台灣地區保險業總保費收入為新台幣 1 兆 4,240 億元，其中產險保費收入為 1,155 億元，而人身保險保費高於產險保費約 11 倍，為 1 兆 3,085 億元。依據壽險同業公會及台灣經濟研究院產經資料庫分析，人身保險七成以上保費由人壽保險部份所覆蓋，可知壽險保費為我國保險產業之主要收入來源。2005 年上半，壽險保費收入達新台幣 7,032 億元，較 2004 年同其成長 15.16%【25】。

截至 2005 年 7 月，國內共有 29 家壽險業者，而 2004 年總資產計達新台幣 5,501 億元，投保率為 166.30%【36】，相較於 2003 年投保率已達 455.98% 的日本而言，我國保險市場尚有發展空間。與日本投保率顯著差距的原因表現在各個方面，除表現在外部環境之國民所得較高、保險知識普及度、政府法規與保險產業發展時期較長等面向外，也同時突顯了內部環境中存在業者與保戶間之資訊不對稱，欲保者未能得到適當的保單資訊，而業者遺落了欲投保的顧客，這般對顧客價值的忽略自然致使保單張數無法提高。

幸而在資訊科技的推波助瀾下，各家業者紛紛仰賴統計軟體、資料庫、電子商務等眾多方法，投身各項價值提升研究。對過往交易資料的處理方式，也由單純資料儲存轉變為知識擷取，冀求取得有效資訊以反應市場狀況或消費者需求，以提升獲利。然而，儘管每種統計方法都能處理龐雜資料，但就解讀資料而言似乎仍不夠精確。保發中心董事長江朝國表示，在金融業經營功能以及保險業獲利能力提升的轉型過程中，資訊妥善運用常居於關鍵地位，而就保險業而言，資料採礦雖非萬靈丹，但對未來銀行保險與新保單的開發將居於重要的地位。透過過往研究【23】亦發現，將資料經依據過往成交紀錄所設計之模型處理分析，並針對所篩選出「最容易成交」之前 10% 客戶名單進行電話行銷，結果使得原先僅 13% 的成交率提升至 50%。可知資料採礦的確能使保險業者在行銷上更有效的應用時間與人力，因此，本研究欲以資料採礦方法針對

保險業者與保戶間所存在之資訊不對稱做探討。

美國調查研究公司 Forester Research 於西元 2000 年所做的研究也同時證實，下一世紀美國主要企業之最想增購之產品前三名，分別為顧客關係管理(CRM)、網站管理(Web Management)以及個人化服務(Personalize)。可知顧客的價值之於企業而言，的確居於舉足輕重的地位，而以人為主的保險產業自當更顯重視。因此，本研究欲針對人身保險市場部份，作顧客價值探討。期望為個案公司評斷出既有顧客之價值層級，並透過歷史資料找尋每層級顧客之關聯法則，以為保戶投資與保戶開發之依據，希望依循法則找出其潛在需求，發展二次行銷及保單開發。

1.2 研究問題與目的

別於他種產業，人身保險業之最大特性乃是以人為基礎，並以此作為保險商品標的物。整個產業的商品條件設計也自然由人所貫穿，從供需配適至業務推廣手段，無不依循「人」所自帶之特徵屬性。然而，人的健康狀況會隨時間改變，生活型態與投資習慣亦隨環境轉變，致使人的需求時時更替，再加上政府甫通過之勞退新制，益愈擴大了原本就存在於業者與顧客間之資訊差距。

面對這樣的問題，業者除了在產品與服務上推陳出新外，一套有效的顧客開發方法亦為不可或缺的一環。目前人身保險業者在經營上多利用業務員人際網絡以為新顧客開發，然若業者能依循 80/20 法則善用既有資料，則可從其中挖掘出既有顧客之潛在需求，落實商品客製化的同時提昇投保率。然而，客製化之施行必須在完善了解既存顧客的前提下，才得以準確實行，國內研究【9】【15】【16】【19】多針對產品屬性或保戶特徵做基本分類，並採用資料採礦方法來歸納屬性。但過往分析多半僅採以少量樣本，對保戶或商品屬性篩選或補值有時也不免流於主觀。

本研究嘗試利用資料採礦技術，協助個案金融控股公司在所提供之人身保險業務資料庫中採擷資料，且利用 RFM 理論依指標對顧客作客觀價值分群，並依層級分析歸納法則，用以辨識不同價值層級顧客，進而依據不同層級法則應用至策略施行。

具體而言，本研究的主要目的如下：

研究目的之一，根據資料採礦與顧客關係管理理論結合以推導出研

究架構，企圖作為保險產業顧客潛在需求與顧客價值分析實作的依據。

研究目的之二，協助區分顧客價值層級且佐以驗證，並針對各層級作法則探勘，以提升二次行銷及新保單開發策略施行之有效性。

1.3 研究流程

本研究分五大流程以圖 1.1 來表示：

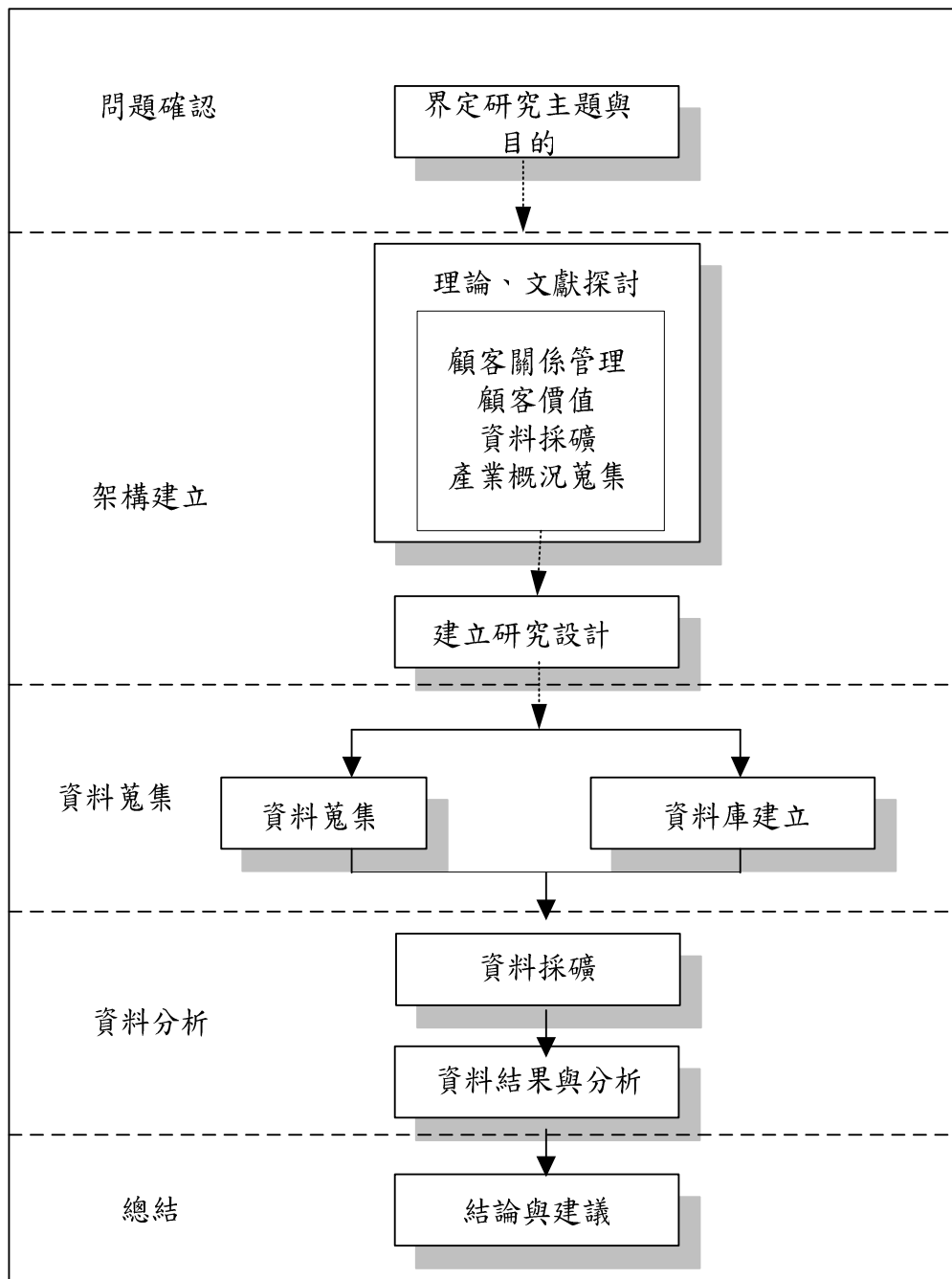


圖 1.1 本研究流程圖

資料來源：本研究整理

1. 問題確認：

瞭解及界定研究主題，說明研究背景及動機，並瞭解需研究的問題及目的。

2. 架構建立：

主要由三大部份之文獻探討做為歸納架構之建立依據，第一部分乃產業概況之蒐集，瞭解保險產業的相關特性以及變化；第二部分則針對行銷管理的重要工具--顧客關係管，作定義、執行步驟與架構加以說明；最後擬就資料採礦之定義、流程、功能與任務加以說明，並簡介其中多種常用的技術和使用時機，進而決定使用何種探勘方式分析資料。

3. 資料收集：

取得研究對象之初級資料，透過資料處理、過濾以及正規化，並建立欲研究資料的資料庫以利後續資料採礦分析。

4. 資料分析：

利用資料採礦技術的決策樹、關聯規則以及統計學上的卡方檢定，企圖在資料庫中找出有價值之資訊。

5. 總結：

根據本研究的實證分析結果，提出本論文的研究發現，並進一步對業界及學界提出可行的建議。

1.4 研究範圍與限制

1.4.1 研究範圍

本研究以個案金融控股公司之保險業務作為研究對象，針對民國 82 年 7 月至民國 94 年 12 月的人壽保險顧客基本資料作分析，以個案公司亂數選取提供的六萬多筆顧客記錄作為實證研究之對象，利用資料庫中的顧客基本資料及保單資料做資料採礦研究。

1.4.2 研究限制

本研究的研究限制如下：

1. 本研究乃以金融控股公司之保險業務顧客資料作為研究對象，故研究結果可能無法以此推論至其他不同型態的產業。

2. 本研究僅針對個案金融控股公司所提供之六萬多筆資料做分析，由於資料取得不易，未來若能將樣本擴大至整個產業，則研究結果將更能接近事實。

第二章 文獻探討

市場的激烈競爭使得顧客關係管理的落實，對各類產業而言益發重要，而其中不外乎保險業，逐漸著眼於顧客開發、保留與流失的部份，並將之視為企業獲利之關鍵因素。綜觀而言，企業利潤多由少部分最具價值之顧客所掌握，80/20 法則亦指出企業 80% 的利潤來自於 20% 之顧客

【21】。於是，如何真正掌握具有價值之顧客，投以等價值比利之資源以為經營，成為業者首重之事。

因此，本章節將對所聚焦之保險產業做概況了解後，深入探討顧客價值之意義以及落實步驟，並選取落實方法中較具潛力之資料採礦技術以為探討。

2.1 保險業

保險乃消化損失的一種制度，應用集腋成裘的原理，由保險公司集合一群人的金錢組成基金，一旦該群人中有人遇到意外遭受損失，便可從基金中提取款項解決經濟危機。因此，保險實乃一門分散風險的學問，亦是風險移轉的優良方法。在今日社會中，風險存在的方式不但新穎且急遽增多，使得現代人在生命以及財產等方面時時面臨著風險，為降低該些風險所造成不安全感及無法預期的影響，風險的角色也就益發重要。

2.1.1 保險產品的分類

一般而言，險種可分為「財產保險」以及「人身保險」。財產保險乃以財產為保險標的物，如房子、車子、貨物、工廠等種類繁多；而人身保險，則以人類生命或身體做為保險標的。人身保險又包括了人壽保險、健康保險、傷害保險以及年金保險，現今又多了利變型及投資型保險。就台灣地區而言，人壽保險公司皆為人身保險公司，而一般所稱的人壽保險事業亦即人身保險事業。由於保險種類繁多，本篇研究僅針對人身保險部分做研究，而保險的完整分類如圖 2.1 所示。

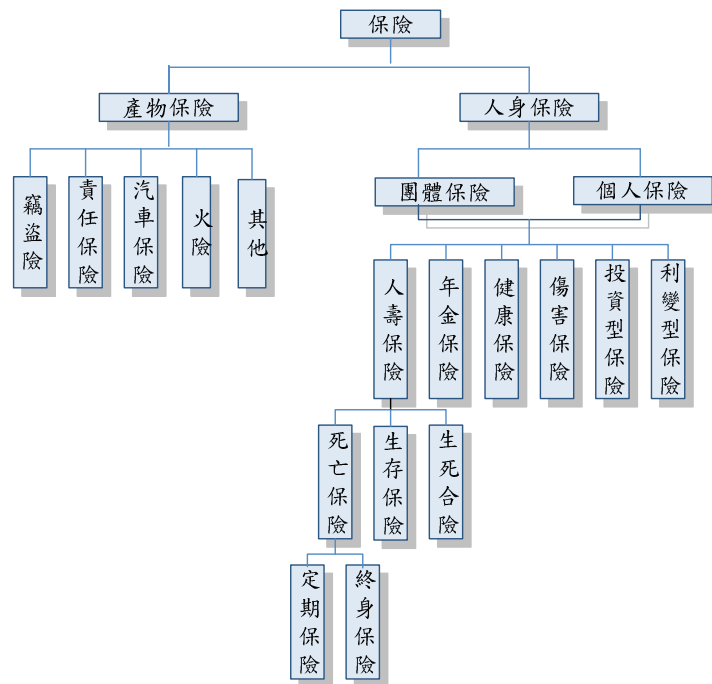


圖 2.1 保險結構圖

資料來源【16】

2.1.2 保險產業之現況與未來發展

近幾年台灣保險市場之發展，人身保險不論在保險密度、滲透度方面均維持高度成長趨勢，其中 2004 年人身保險密度已達 57,671 元，而人壽與年金保險投保率已達 166.30%，較 2003 年提高 7.43 個百分點，普及率更達 309.97%。

2004 年台灣地區保險業總保費收入為新台幣 1 兆 4,240 億元，其中人身保險保費收入為新台幣 1 兆 3,085 億元，產險保費收入為 1,155 億元，我們發現人身保險市場規模大於產險達 11 倍，且趨勢顯示差距逐漸擴大。2005 年壽險保費在人身保險市場佔足七成以上的比重，仍為保險業主要保費收入來源，而其在上半年之保費總收入達新台幣 7,032 億元，較 2004 年同其成長 15.16%【25】，如表 2.1 所示。

表 2.1 壽險業保險契約保額與件數統計

單位：新台幣百萬元；%

類別	2004 年				2005 年 1~6 月			
	保額	成長率	件(人)數	成長率	保額	成長率	件(人)數	成長率
新契約								
壽險	3,395,585	9.70	4,742,600	-15.27	1,884,883	17.05	2,661,344	15.68
傷害險	24,092,364	-7.37	14,449,241	-6.06	12,204,533	15.26	8,257,614	27.89
健康險	5,258,332	-39.90	10,662,743	-26.67	1,544,938	-58.83	4,520,128	-27.54
年金險	116,398	123.41	190,144	118.88	53,748	-0.97	87,835	2.72
有效契約								
壽險	28,783,571	8.11	37,452,430	4.58	29,594,611	6.28	38,615,645	4.57
傷害險	35,647,171	-18.49	41,508,877	-0.56	35,488,456	-20.87	42,961,229	1.95
健康險	15,064,095	2.50	54,237,615	5.26	15,200,293	0.96	55,512,415	4.74
年金險	165,576	203.03	280,377	180.11	212,192	104.47	359,503	101.40

資料來源【25】

2005 年上半年傷害險的保費收入在其新契約及有效契約件數均止跌回升，保費收入由衰退轉為小幅成長 2.57%。健康險年度保費收入較 2004 同期成長，使得該項保費收入於 2005 年上半年有 8.23% 的年成長率。年金險方面，雖 2005 年上半年受到勞退新制的影響，使其保費收入成長率由先前大幅成長趨勢轉為衰退，如表 2.2 所示，但預計在 2005 年第四季企業年金險正式推出後，對於年金險保費收入應有相當挹注，企業及個人二者可望於 2006 年成為年金險的發展重心，故年金險保費收入的發展仍以樂觀視之【25】。綜觀而言，壽險市場在未來的發展依舊可期。

表 2.2 壽險業保費收入統計

單位：新台幣百萬元；%

類別	2004 年			2005 年 1~6 月		
	金額	成長率	佔有率	金額	成長率	佔有率
壽險	955,015	9.85	72.99	536,803	20.59	76.33
個人	948,119	9.64	72.46	533,385	20.86	75.85
團體	6,896	0.86	0.53	3,418	-10.64	0.49
傷害險	57,891	-5.95	4.42	28,773	2.57	4.09
個人	48,704	-6.42	3.72	24,031	1.65	3.42
團體	9,187	-3.39	0.70	4,742	7.50	0.67
健康險	140,579	11.64	10.74	72,369	8.23	10.29
個人	134,238	11.57	10.26	68,749	7.50	9.78
團體	6,341	13.21	0.48	3,620	24.10	0.51
年金險	155,004	110.55	11.85	65,296	-7.52	9.29
合計	1,308,489	15.52	100.00	703,241	15.16	100.00

註：成長率係與上年同期比較

資料來源【25】

透過壽險商業同業公會與台灣經濟研究院產經資料庫於 2005 年 9 月份之報告可知，2005 年保險業存在諸多不利因素，如長期低利使得利差損壓力未明顯減輕、台外幣利差擴大加重保險業者海外投資避險成本、國內景氣表現不如預期使業者投資表現相對不佳等，致使我國保險業所面臨的經濟環境波動仍大。幸而保險業者積極推出新產品，並透過多元化擴大行銷通路，而政府在保險業相關法規的陸續修訂，亦解除對保險業的管制，使業者經營能更加靈活自由，紓解業者部份經營上的困境。

在新契約保費收入概況方面，2005 年第一季之投資型保險與傳統型保險市場呈現一消一漲，且傳統型保險比重領先投資型保險的差距明顯擴大。投資型保單收入在 2004 年不論是比重或規模，皆有突出的表現(見下表)，但 2005 年第一季卻一反 2004 年之成長趨勢，保費收入較 2004 年第一季衰退 9.95%，如表 2.3 所示，其中利率上揚使非投資型之分紅保單、利變型保單重要性逐漸增加，部分業者推出短年期的還本壽險熱賣，帶動傳統型壽險買氣增溫，使投資型保單熱潮消退，都是投資型保單成長減緩之因。

表 2.3 新契約保費收入概況

單位：新台幣百萬元；%

項目	2004 年		2005 年第一季		
	保費收入	比重	保費收入	比重	成長率
傳統型保險	291,622	65.36	96,710	75.48	35.64
投資型保險	154,580	34.64	31,412	24.52	-9.95
總保費收入	446,201	100.00	128,131	100.00	20.66

資料來源【25】

競爭情勢方面，截至 2005 年 7 月，國內壽險公司共有 29 家，其中本國業者為 21 家，外商在台分公司共 8 家。由保費收入市占率觀察，2005 年一月至五月前八大壽險公司排名與 2004 年相同，依次為國泰、南山、新光、中華郵政、美商安泰、富邦、中國以及三商美邦，且壽險整體保費收入表現均呈現成長趨勢，但影響 2005 年保費收入成長幅度大小原因，則著重於商品策略是否符合市場所需【25】。

就 2006 年景氣展望方面，87.50%的業者認為國內營運總額經增加，顯示在多元行銷通路、勞退市場商機、針對特定族群加強開法新產品及費率自由化下，業者認為未來營運額可望持續擴大，也連帶有 62.50%的業者對 2006 年的獲利率看法趨於樂觀，惟有 62.50%的業者同認為總成本亦將提高，連帶削弱業者對 2006 年整體保鮮業景氣之看法，加上多家業者資金運用操作之效益一直無法有效提升，故僅 37.50%的業者認為 2006 年總體保險業景氣可望再好轉，而有 37.50%認為景氣持平及 25.00%業者看壞。

2.2 顧客價值

顧客價值【2】【31】不外乎從三個方向審視：消費者盼望之產品品性、顧客所感受的服務品性以及以價值為基礎所訂定之價格，如圖 2.2 所示。當每件交易之前置與後端處理，能在此三個方面滿足甚或超越客戶期待，便是良好顧客價值之生成；反之，三個面向只消有任一面向未達客戶所願，就容易使得顧客價值低落。由於此三要素彼此所存在之重要性相同，買賣雙方施與受的配置彷彿一正三角形。就力學原理視之，只要其中任一邊「施」力不勻稱，三角形面積便會「受」到扭曲。故而若想滿足顧客需求，便須仰賴撐出一個等邊三角形的高度藝術。

顧客價值三角模式主要奠基於下五項要點：

1. 由顧客定義產品品性、服務品性與合理價格。
2. 顧客價值的期待乃相對於競爭者所提供。
3. 顧客價值是動態的，且總是往較高期望移動。
4. 產品品性與服務品性之傳遞非單是製造廠商所肩負，而乃整條行銷通路之責任。
5. 顧客價值乃整個組織之一致承諾。

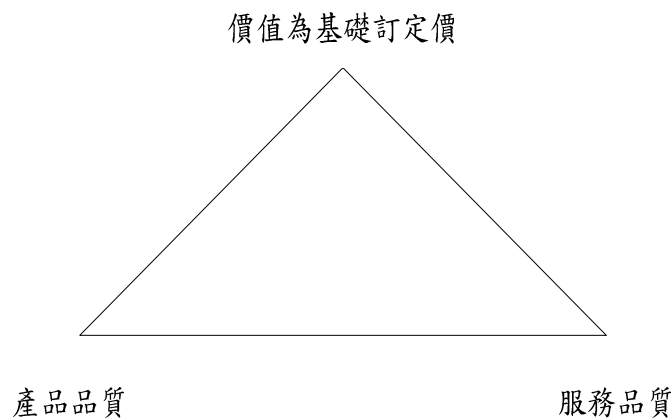


圖 2.2 顧客價值的三角模式

資料來源【31】

2.2.1 顧客價值之定義與方法

每位顧客對於企業所提供的產品、勞務及其他相關事項都抱持不同程度的期望，當顧客所獲得勝於先前所期望，對於品性之評價就相對得到提升。實際上，顧客對於衡量一件事物都存在不同之模式與價值，麻州劍橋策略規劃學院(Strategic Planning Institute of Cambridge)【13】曾針對許多企業調查，企圖透過數千種不同的元素，彙整出企業經營成功之關鍵策略，研究結論明白指出，真正贏得顧客並長久保有之道乃設法使顧客認知其所付出的成本是值得的。顧客價值的定義彙整歸納如表 2.4。

表 2.4 顧客價值定義彙整

作者/年代	顧客價值定義
大前研一(1983)【22】	策略的首要目的，策略的首要目的，不是打敗競爭對手，而是提供顧客真正需要的價值。
Gale (1994)【24】	認為消費者價值是市場上所認知品質調適到該產品的相對價格。
Kotler (1994)【30】	任何創新的服務都會被模仿，因此服務業者要如何保持領先優勢，領先的市場，必須創造新特色即新的顧客價值。
Naumann (1995)【30】	提出顧客價值的三角模式即產品品質、服務品質與價格，且認為顧客透過此三種向度來認知所得到的是否有真正的價值。
Engel, Minird and Blackwell (1995)【7】	提出價值是決定產品的使用，並幫助消費者在不同的產品中做出抉擇，為了解決衝突，而訂定決策的標準。
Butz and Goodstein (1996)【4】	認為消費者價值是他們使用卓越產品或服務後，感受到有附加價值時所引發消費者與生產者間之感情連結。
Woodruff (1997)【34】	認為顧客價值是顧客認知與偏好與產品屬性、屬性績效及使用過後達成目標及目標的結果等之評估。
Solomon (2000)【28】	由市場行銷來發展，認為價值是提供市場區隔的基礎。

資料來源：【24】

Magson【21】認為定義價值時需包含時間以及價值衡量對象二項關鍵元素，並進一步的提出顧客價值立方體，指出完整的顧客價值分析，除在顧客層面要能計算個體或某一區隔之價值外，還要衡量行銷活動方面的價值，如圖 2.3 所示。

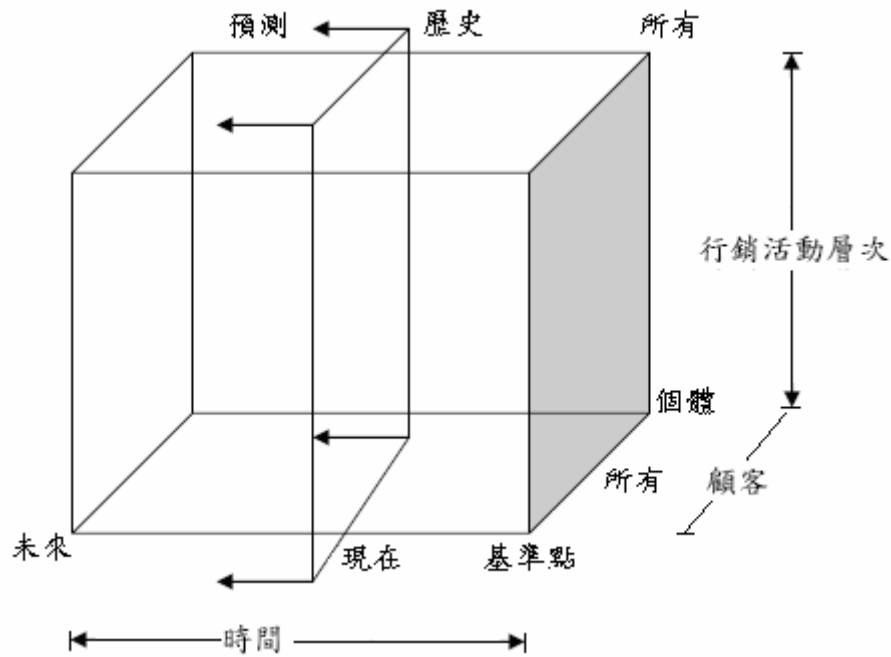


圖 2.3 顧客價值立方體

資料來源【21】

Kotler 等【20】提出顧客價值乃顧客從產品或服務中得到的總價值(Total Customer Value)，包括產品價值(Product Value)、服務價值(Service Value)、個人價值(Personnel Value)、形象價值(Image Value)。而顧客為取得產品與服務價值，所要花費的成本含括金錢成本(Monetary Price)、時間成本(Time Cost)、精力成本(Energy Cost)及心理成本(Psychic Cost)，合稱為顧客成本(Customer Value)。把顧客價值與顧客成本兩者相加就是顧客真正獲得的價值(Customer Delivered Value)，如圖 2.4 所示。

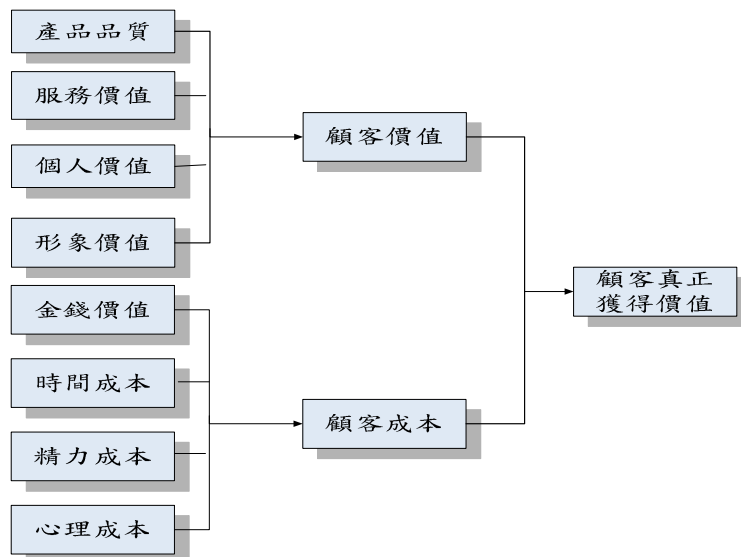


圖 2.4 顧客價值圖

資料來源【20】

綜觀所述，許多學者依顧客所求針對不同的顧客定義價值，將實施顧客價值的各種方法與步驟予以歸納成表 2.5。

表 2.5 實施顧客價值的各種方法與步驟

學者	焦點	內容	行銷導向		
			顧客	市場	價值
Band (1991)	創造顧客價值的方法	衡量顧客價值的認知 擁護創造價值的顧客 經營與顧客的關係 完成高品質的作業績效	√		√
Berry (1991)	建立顧客價值的方法	提高財務性利益、提高社會性利益、增加結構性的關係	√		√
Kotler (1992)	如何尋找顧客新的價值	實證過程、直覺過程、辯證過程、需要層級過程	√	√	√
Gale (1994)	顧客價值管理的演進	一致性的品質、顧客滿意、市場認知的品質與價值、顧客價值的管理	√	√	√
Jasionwski And Hamrin (1995)	超越顧客價值的方法	促使員工以顧客為焦點，建立後續及衡量顧客服務的系統講究細節、最高主管直接介入、持續不斷的改進	√		√

資料來源【26】

簡而言之，創造顧客價值乃使顧客兼顧產品屬性(所感受的實體品質)、服務水準(包括售後服務)與價格(相對反應出的價值)三方面，而廠商勢必得著眼於此思維基礎上時時與競爭商品比較，假若只重視顧客價值「三角模式」中任一或二個面向，便即有可能面臨顧客流失之窘境。

2.2.2 顧客價值指標 RFM 模式

根據 Arthur Hughes 【12】的研究，RFM 通常以平均數來代表。利用此三指標可衡量企業與顧客之間的關係，進而判定每位顧客價值度，據以擬出顧客關係策略。企業常利用 RFM 指標將顧客量化評分，其目的在量化顧客消費行為，使其符合科學行銷公式應用。RFM 為「最近購買時間(Recently)」、「購買頻率(Frequency)」、「購買金額(Monetary Amount)」三個字的縮寫，為評量顧客忠誠度與顧客貢獻度時，最常使用的一個評估法【34】。以下為 RFM 之說明：

1. 最近購買時間

即計算由最後一次購買時間至現在之時間，為測量時間的量度。購買產品時間距現時愈近，隱含著該顧客再次選購此產品之購買程度愈高；反之，則表示著此顧客之購買行為可能改變或已經改變。

2. 購買頻率

乃一個期間內顧客購買的總次數，例如一季、一個月，甚至一個星期，亦可用於測量顧客在此期間內與公司之互動程度。購買頻率愈高代表此顧客與公司互動程度愈高，持續購買動機較強。

3. 購買金額

決定在某一時段內購買的總金額。顧客購買某類產品之總金額代表著對此類產品之興趣指標，亦可視為對公司實質金錢之貢獻。Kahan 【18】。提及 RFM 是應用非常廣泛的行為分析技術，利用 RFM 可以更簡單、更快速地分析公司的顧客。Kahan 並指出，從行為的觀點來看 RFM 是最常用來測量與顧客價值強度的方法之一。在 RFM 的使用上，如 Suh et al. 【30】，將近期從大到小排列，前百分之二十歸為第一類，依次分成五類；同樣的，將頻率從大到小排列，前百分之五十分歸為第一類，共分成兩類；金額(Monetary)同於頻率的分法，分成兩類。如此一來，即獲得每一顧客的 RFM 分數組合，由最差的顧客到最好的顧

客，再將不同的分數組合分別給予從 0 到 1 不同的分數，最後針對分數較高的顧客採取行銷的策略。

Sung and Sang【29】是將 RFM 的值在沒有經過轉換的情況下，直接當成建立模式的輸入變數，透過集群的分析方法將顧客分成不同的顧客群，再針對不同的顧客群採行不同的策略。Bauer【4】將 RFM 用於郵寄時的目標顧客選擇，認為使用 RFM 做為直效行銷工具之優點為：

1. 增加回應率
2. 降低每份郵寄廣告單之成本
3. 有助於顧客個人化。顧客對某產品之忠誠度愈高(RFM 值愈高)，亦即蘊涵該顧客下次持續購買該商品之機率愈大。

2.2.3 顧客價值於顧客關係管理之步驟

顧客關係管理乃一種即時將顧客資訊不斷轉變成顧客知識之過程，其步驟分為知識發掘(Knowledge Discovery)、市場行銷計畫(Marketing Planning)、顧客互動(Customer Interaction)以及分析與修正(Analysis and Refinement)四個階段，並循環往復直至找出最適合自身的最佳管理方法，如圖 2.5 所示。而 Kalakota and Robinson【17】認為，要妥善的管理顧客生命週期，就必須以三個階段來實行顧客關係管理，這三個階段分別為：獲取新顧客(Acquisition)、強化既有顧客的獲利性(Enhancement)、以及維持現有顧客(Retention)，如圖 2.6 所示。

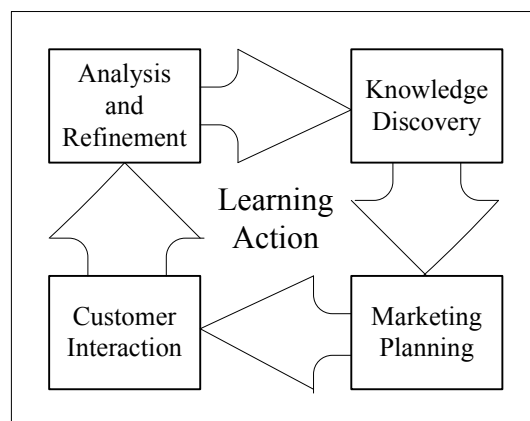


圖 2.5 顧客關係管理的循環過程

資料來源：【30】

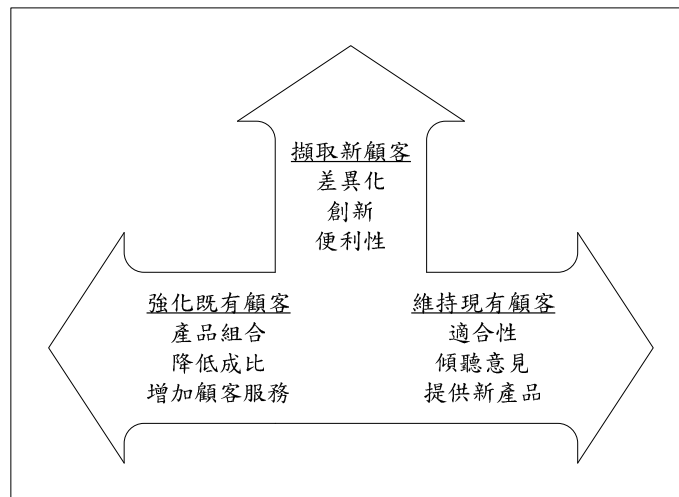


圖 2.6 顧客關係管理之運作流程

資料來源：【17】

2.3 資料採礦

隨時間累積，現今企業體之電腦系統中經常儲存大量資料，包括市場、顧客、供應商、競爭對手以及未來趨勢等重要資訊，但過多的資訊與無組織化，使得企業決策單位常無法有效地利用現存資訊，甚至使得決策行為產生混亂淆與誤用。而資料採礦技術之採用，乃從巨量的資料庫中，挖掘出不同的資訊與知識以為決策支援之用，進而產生企業的競爭優勢，致使獲得更龐大之客群，在從客群中採擷知識，如此循環往復。

2.3.1 資料採礦的定義

對於資料採礦的定義，多位知名的資料採礦領域學者，都曾給予資料採礦不同的意義。較早的學者如 Frawley、Grupe 和 Owrang 等僅視資料採礦為資料庫知識發現過程（KDD；Knowledge Discovery in Databases）中的一個步驟，但後來的學者有將兩者劃上等號的趨勢，給予資料採礦較為寬鬆的定義，認為資料採礦是一資料庫知識發現過程。

資料採礦【18】是指找尋隱藏在資料中的趨勢(Trend)、特徵(Pattern)及相關性(Relationship)等訊息過程，資料採礦是一種習慣稱法，我們也可稱這種從資料庫中挖掘知識的技術或系統為『從資料庫挖掘知識』(Knowledge mining from database)、『資料萃取』(Knowledge extraction)、『資料模型分析』(Data Pattern Analysis)、『資料考古學』(Data archaeology)或『資料撈取』(Data dredging)等等。有些人亦將資料採礦視為知識發現的同義辭，目前已被許多研究人員視為結合資料庫系統與機器學習技術的

重要領域，許多產業界人士也認為此領域是一項增加各企業潛能的重要指標。

而李昇暉【6】認為所謂的「資料挖掘」(Data Mining)，就是從大量的資料庫中找出相關的模式(Relevant Patterns)並自動地萃取出可預測的資訊。而資料採礦的目的是建立預測模型，也就是從資料中找出對應的各項關聯。

2.3.2 資料採礦的功能

Michael J.A. Berry & Gordon S. Linoff【12】認為，Data Mining 的功能包含了六大功能：分類(classification)、推估(estimation)、預測(prediction)、關聯分組或同質分組(association or affinity grouping)、群集化(clustering)、描述及視覺化(description and visualization)。分述如下：

1. 分類(classification)

依分析對象的屬性分門別類並加以定義，建立類組(class)。如將保單貸款者的風險屬性，區分為高、中、低度風險申請者。使用的技巧有決策樹(decision tree)，記憶基礎推理(memory - based reasoning)等。

2. 推估(estimation)

分類的結果乃是不連續的，而推估所得的結果則是連續性的數值。如按照信用申請者之教育程度、行為別來推估其信用卡消費量。使用的技巧包括統計方法上之相關分析、迴歸分析以及類神經網路方法。

3. 預測(prediction)

根據對象屬性之過去觀察值來推估該屬性之未來值。例如由顧客過去之刷卡消費量預測其未來之刷卡消費量。使用的技巧包括迴歸分析、時間數列分析與類神經網路方法。

4. 關聯分組(association or affinity grouping)

從所有物件決定那些相關物件應該放在一起。例如超市中相關之盥洗用品(牙刷、牙膏、牙線)，放在同一間貨架上。在客戶行銷系統上，用來確認交叉銷售的機會以設計出吸引人的產品群組。

5. 群集化(clustering)

將異質母體中區隔為較具同質性之群組(clusters)，相當於行銷術語中的區隔化(segmentation)。群集話事先並未對於區隔加以定義，使資料自然產生區隔。使用的技巧包括 k-means 法及 agglomeration 法。

6. 描述及視覺化(description and visualization)

資料探勘之目的有時只是單純的想利用更容易了解的方式，來描述一個複雜的資料庫中所存在的狀況。對一種尚未被認識的狀態而言，一個正確的描述足以啟發許多對該狀態的認識與解釋，而資料視覺化即為一種強而有利的描述方式，透過人類所擅長讀取的圖像表示方法來呈現，比一千條列表的規則更有價值。

2.3.3 資料採礦的方式

資料探勘非是一種新技術，而是結合多項專業技術的研究【26】它包括機器學習、統計方法、資料庫、專家系統及資料視覺化的領域。Curt【10】認為資料探勘的技術至少包括資料視覺化、機器學習技術、統計技術及資料庫四種技術。Y. Fu 於 1997 年【8】將資料探勘的技術依不同研究領域區分說明：

1. 統計方法 (Statistical Approaches)：資料探勘使用許多統計工具，包括 Bayesian 網路、迴歸分析、相關分析及分群分析 (Cluster Analysis)，通常統計模式是經由訓練資料集 (Training Data Sets) 來建立，然後從模式中尋求規則及特徵。
2. 機器學習方法 (Machine Learning Approaches)：像統計方法一樣，機器學習方法尋求一個最佳模式來符合測試資料 (Testing Data)。不同之處在於大多數機器學習方法利用資料自動化學習過程，自動歸納出分類規則及建立模組。資料探勘最常使用的機器學習方法包括類神經網路以及決策樹歸納。
3. 資料庫導向方法 (Database-oriented Approaches)：資料庫導向方法不像前兩個技術在尋找最佳模式，轉而著重在處理現有的資料，是一種屬性導向歸納 (Attribute-oriented Induction)，反覆從大量的資料中找出共同的規則及模式。

4. 視覺探索 (Visual Exploration)：將多維度的資料轉換成視覺化物件，如點、線及區域，讓使用者可以動態檢視及探索有興趣的部分，進而分析出資料的模式。
5. 其他方法：如類神經網路 (Neural Network) 運用於資料的分類及預測，約略集合 (Rough Set) 運用於分類及分群。這些技術可以互相整合來處理複雜的問題及提供替代方案，如統計方法可以結合視覺探索技術，用以輔助系統處理資料模式及趨勢分析。

另外，常見的資料採礦方法如下：

1. 購物籃分析：運用關聯規則 (Association Rule)，著眼點在於找出可以一起銷售的商品組合，它能顯示商品組合的售出率有多高並且形成規則，當交易是非匿名時，它可以加上一個模組進行跨時性分析。
2. 記憶基礎理解 (Memory-based Reasoning)：比對新資料和現存資料庫原有資料的相似度，以決定新資料的類別或預測值的工作，它是以距離函數判斷資料間的相似度，再以數個最相似的資料投票來決定歸類結果。
3. 群集偵測：針對要分析的資料，利用幾何學、統計、類神經網路等方法，將資料分成多個群內同質、群間異質的群組，使各群組的特徵能有效突顯出來，連結分析搜尋資料間的關聯，並且據此發展出分析模型，這是圖像理論在資料挖掘中的應用。
4. 決策樹：主要用在資料分類與預測功能上，它能將訓練資料集的紀錄區分為獨立的子群，每一子群都有自己的規律，彼此是互斥的。利用決策樹技術，容易獲得清楚易懂的分類規則並找出關鍵屬性，所以本研究將加以利用決策樹技術來進行資料挖掘的工作。
5. 類神經網路：一種平行分散式的計算模式，以大量簡單的相連人工神經元，模仿生物神經網路的資訊處理系統，使電腦能夠模擬人類的神經系統結構，進行資料的處理。
6. 基因演算法：應用選擇、雜交、突變等物競天擇和基因演化的機制，將此機制結合電腦語言，經過世代繁衍，得到最後留下的最佳方案，通常用於找尋預測功能的最佳參數。

本研究應用購物籃分析、集群偵測以及決策樹的方法，透過保戶交易以及保單資料，來做顧客行為特徵與商品特徵分析。

2.3.4 資料採礦與統計方法

一般而言，資料採礦在統計上可由兩種不同的角度來看：

1. 以摘要為主：即對資料庫有整體地了解。對傳統的統計方法而言，統計抽樣和模型的建立是其中的骨幹，目的乃對母體中有興趣的部分先建立模型再做推論，以對資料庫的正常狀況有所了解，並區分常態及異常。

2. 以預測性為主：如哪類型之顧客潛在願付之保額較高？哪類型顧客投保某一類型保單之機率較高？哪類型的保戶可以使用保單貸款？回答這一類的問題時，我們必須建立一個基本的模型和公式【29】。這套模型公式在資料庫中不單只是重要的參考指標，更甚者能幫助我們發覺出哪些是所有可能的顧客之一，當我們自這個角度來尋求結果時，所需的樣本並不是一般的統計抽樣所能給的。

從技術上來看，資料採礦可以追溯至三大始祖【19】：

- A. 傳統統計學(Classical Statistics)
- B. 人工智慧(Artificial Intelligence，簡稱 AI)
- C. 機器學習(Machine Learning)

其中傳統統計學可說是資料採礦的技術中最基礎的部份，提供了資料採礦所必須之基礎技術。硬要區分 Data Mining 和統計的差異是沒有太大意義的，一般將之定義為 Data Mining 技術的 CART、CHAID 或模糊計算等等理論方法，也都是由統計學者根據統計理論所發展衍生，換另一個角度看，Data Mining 有相當大的比重是由高等統計學中的多變量分析所支撐。

資料採礦之第二個源頭乃人工智慧，其必須搭配運算速度非常快的電腦來使用。人工智慧精確的定義是：予以一電腦系統人類的知識和行為，並具有學習、推理判斷來解決問題，且能記憶知識和了解人類自然語言的能力。在經多年的研究後，推展出人工智慧之生成過程乃累積人

類針對問題及相關事物所引起的反應，以及因其所引發的推理、判斷、解決問題、學習、及思考決策等過程，並將這些過程分解成一些基本步驟，再透過電腦程式，將這些解決問題的過程模組化或公式化，使得電腦能夠有一個結構的方法來設計或應付更複雜的問題。這套能夠應付問題的軟體系統，即稱之為人工智慧系統。

機器學習為資料採礦之第三個源頭，其發展可謂之是由人工智慧延伸而成，是統計學與人工智慧的結合，融合了人工智慧的啟發式學習及高等統計分析，嘗試讓電腦程式從資料中學習，並依不同的資料品質來使用統計的基本概念及人工智慧啟發，以做出不同的決策【33】。圖 2.7 清楚表達上述整個發展歷程，且成功地應用在不同的領域。

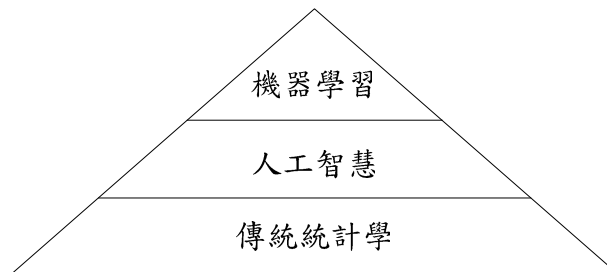


圖 2.7 資料採礦技術發展歷程

資料來源【33】

綜觀上述，資料採礦就即從大量的資料中發掘出潛在且有用的知識，以『正確掌握資訊』。當新的記錄進入資料庫，抑或在欄位中加入新的元素時，會使原生資料(raw data)不斷地增加，使資料量成長，而漸漸地成為一個龐大的資料庫，資料採礦可謂為資料庫知識發現的核心【8】。隨著資料庫技術的提高，以及資料處理能力的發展，資料倉儲(Data Warehouse)成為資料採礦的主要基礎，資料倉儲能從企業的大量資料中擷取有用的資料，透過整合的過程，提供給企業管理者及決策者所需的資料，所以資料倉儲是資料處理的一個過程【32】。對於資料倉儲的定義而言，資料倉儲是以更好的方式提供企業組織的決策分析，且回答使用者對系統所詢問的各種交互參照問題，並滿足不同使用者的需求；其收集資料的面向包括主題導向的、整合的、不可更改的、隨時間不斷變化的資料集合，它要從大量的資料庫中抽取出資料，經過清理、分類、轉換、管理，形成統一的儲存格式，給決策者和使用者們提供更好的支持。綜

上所述，圖 2.7 可知資料採礦結合資料庫系統與機器學習等技術，根據資料之間的特性、相互關係來發掘出有意義的訊息。

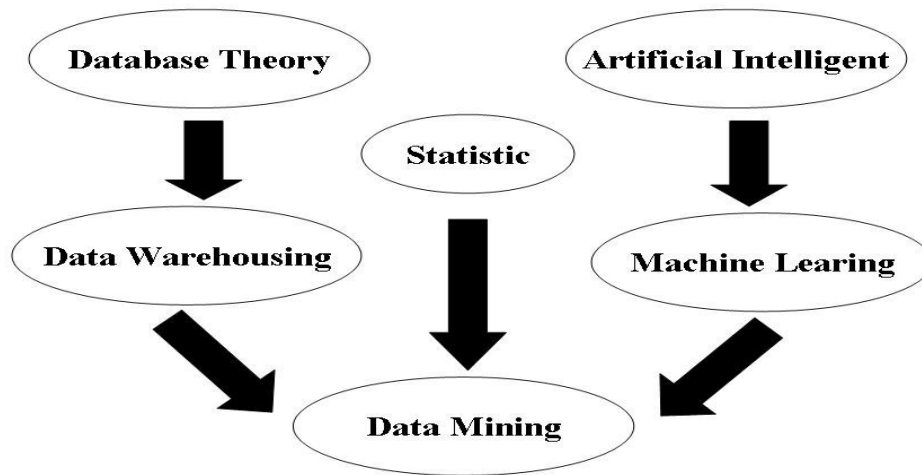


圖 2.8 資料採礦組成結構圖

資料來源【32】

從圖 2.8 中可看出，統計學是建立大部份資料採礦的技術中最基礎的部份，如包括其中的迴歸分析、標準分配、標準差、變異數、群聚分析及內部一致性等，都提供了資料採礦必須的基礎技術。因此，傳統統計學在發展資料採礦的工具及技術中扮演非常重要的角色。學者 Witten 和 Frank (2000)【33】曾指出：資料採礦和統計學兩者的共同特點都是企圖從資料中找到結構與知識，而前者是將尋找所有可能假設的過程加以一般化，後者為先提出假設，再考驗假設是否正確。

資料採礦比傳統統計分析方法更具有三點優勢【28】：

1. 資料採礦分析方法不需要對既有的函數作假設，研究者只需選擇適當的變數。
2. 資料採礦分析方法具有自我學習和處理非線性問題的能力，能針對任一確定模式作良好的函數逼近，以彌補有時應用在非線性的模式建構上時假設太強的缺點。
3. 資料採礦可以根據自我學習所得到的規則，來推論類似輸入資料的輸出結果的歸納能力。表 2.6 為資料採礦分析法與傳統統計分析法

的比較結果。

表 2.6 資料採礦分析法與傳統統計分析法之比較

	資料採礦法	統計方法
對分析資料屬性清楚	必須	必須
對解決問題目標明確	必須	必須
提供分析演算法	統計分析方法、人工智慧、決策樹、類神經網路	統計分析方法
模式建立	提供多種模型，可在短時間內決定合適者	需要分析者逐一分析變數重要性、模式才能建立
相關變數	可以找出多個變數之相關性	一次只能檢查一個變數對結果的影響
可以預期分析結果	不可以	可以
執行方式	不斷循環、不斷修正的過程	可以問題為導向，相關問題通常只需分析一次

資料來源【32】

2.3.5 資料採礦的流程

資料採礦流程又可稱為工作循環如圖 2.11 所示，

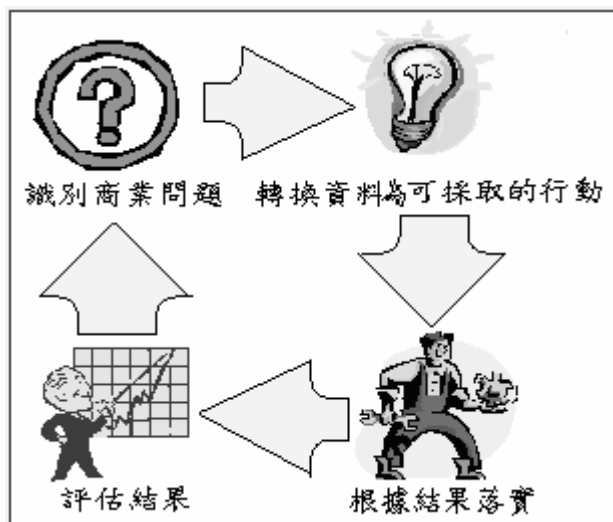


圖 2.9 資料採礦流程

資料來源【13】

其包含了四種主要的商業程序：

1. 識別商業問題：瞭解問題的所在，需從什麼方向、角度來分析此問題。

2. 將資料轉換到可採取行動的結果上：使用資料採礦方法，把資料轉換到可採取行動的資訊上。通常此階段最為繁雜，又可分為四個步驟：
 - (1). 選取輸入資料：資料探勘的第一步驟，是指定要探勘和分析的資料，資料來源不一定是特定資料庫中的所有資料，探勘的資料可能從一個或多個資料庫中，取得表格、概略表或記錄文字檔。
 - (2). 轉換資料：為降低資料量，首先將收集的資料作整理、清除重複或無效的資料記錄，並且確保消費者資料的完整。
 - (3). 資料採礦：使用資料採礦方法如分類、趨勢分析、分群、關聯及循序特徵等，從轉換後的資料中發掘存在的多種特徵及資訊。
 - (4). 解釋結果：經過資料探勘後，一般以文字及圖形來顯示結果。
3. 根據結果採取行動：依據分析的資訊做出行動。
4. 評估結果：對於行動後所得到回應做一評估，再度定義新的問題，重複此工作循環。

2.3.6 資料採礦於保險業之應用

隨著保險業的競爭激烈，國外保險集團透過資料探勘的技術建置資料庫，並成功達到資料庫行銷的案例，亦逐漸受到國內業者的注意與參考。以美國保險業者於資料採礦之應用來說，主要乃針對客戶流失的挽回、保單的升級與再銷售以及找出並預防詐騙的可能性三個方面，而近年資料挖掘於我國保險業之研究彙整如表 2.7 所示。

表 2.7 資料採礦在保險業之相關研究

出版年	作者	研究目的	研究重點	演算法則選擇
2002	涂靜儀【14】	運用人工智慧之類神經網路與基因演算法進行之是擷取，期能得知資料潛藏之意義，了解客戶之風險特徵，以輔助企業訂定行銷策略，區分目標市場，技進行目標市場行銷	明定保護之風險特徵，進行目標市場行銷	自組織映射圖網路(SOM)、基因演算法
2002	林育宏【10】	利用 SOM 將業務員及其	探討業務員	自組織映射圖網

		客戶之保單資料，做適當的分群分類。探討個案公司業務員的基本特性與其客戶保單間之關係，藉以從中獲取重要的行銷資訊	和保戶之前的關係	路(SOM)
2003	李宏谷【5】	建置一簡易的顧客關係管理行銷系統，有效率的自動將顧客資料分群	顧客關係管理之分群研究	自組織映射圖網路(SOM)
2004	張愷芬【16】	針對人身保險做交叉銷售的分析，以過往的顧客選購保險中的資料，運用關聯法則找出一個銷售的最佳保險組合給顧客	探討購買險種之組合(以主、副約來分析)	關聯法則、決策樹
2004	江茂榮【4】	針對過往資料之人身保險做交叉分析，找出業務員與客戶投保之相關性，來幫助業務員提高招攬成功的機率	找出業務員與客戶投保之關連	自組織映射圖網路(SOM)
2004	林明宏【11】	發展合適的早期保單失效率之預測模型，並由單一公司的運用推廣至各保險公司，以提供業界及早採取積極的契約包全工作，維持較佳之保單繼續率。	預測保單之師效率模型	類神經網路
2004	李家旭【7】	針對不同主約保單客戶，深入探討附加保單狀況的分析，且建構出預測購買附加保單之模式，並找出因藏在資料中的關聯規則	探討主附險間的關連，並兼購預測模型	類神經網路、決策樹(C5.0)、聯法則
2004	羅國輝【35】	訪談個案公司導入顧客關係管理系統之應用，以問卷調查探討不同個人背景因素對於使用顧客關係管理系統的滿意度是否有所影響，並分析業務人員使用顧客關係管理系統之績效，因哪些因素而有所差異。	探討壽險人員運用顧客關係管理系統滿意度與績效間的關連	多變量分析
2004	李珮榕【8】	針對個案公司保單做資料探勘，找出其過往資料中貸款保單的行為特徵模式	探討台北地區保單貸款特性	類神經網路、決策樹(CART及C4.5)
2005	張家鳳【15】	將既有客戶進行市場區隔析，欲找出投資型保單客	探討投資型保單特質，並	關聯法則、決策樹

		戶特質，並找出保單所購買產品險種的關連性，以設計差異化聯合銷售方案。	找出險種間的關連性	
2005	Kuo-Lung Tseng【31】	藉由負向關連性的性質，找出存在於不同群組之保戶間，具有顯著差異的負向對比集，以制定良好的行銷策略及產品規劃	Mining negative contrast sets from data with discrete attributes	關聯法則
2005	Chien-Hsing Wu、Shu-Chen Kao、Yann-Yean Su、Chuan-Chun Wu【6】	建構新的顧客回應模式之知識探勘機制，並以建構的顧客回應模式協助壽險公司擬定粗銷策略	Targeting customers via discovery knowledge for the insurance industry	決策樹(ID3)
2005	陳佳鈴【19】	應用資料探勘於客戶關係管理之分群研究，以制定行銷策略、商品組合或培訓人才。	找出以分類好的保單客戶與保單型態間的關連	自組織映射圖網路(SOM)

資料來源：本研究整理

由上述歸納可知，資料採礦於我國保險業之研究當廣泛，除流失率、保單貸款風險、保戶加保預測、業務員掌控因素與顧客關係管理系統滿意度外，多就關聯法則或決策樹此二演算法進行法則探勘與分類，研究主要包含險種特徵屬性間、保戶特徵屬性間以及險種與保戶特徵屬性間三個區隔，且多直接將清理完之資料投以演算法做分類。然而，在進行顧客價值分析與顧客關係管理時，所蒐集資料含括範圍甚廣，如行銷活動的顧客反應度、運銷和產品供應之相關數據資料、銷售與購買之資料、客戶資料、客戶網站註冊資料、相關服務之數據、產品市場資料及網路銷售數據等。而資料欄位特徵屬性之定義方式繁多，若定義方式未若妥當，則分類與法則之準確度則需予以琢磨。因此，本研究依據 RFM 理論做客觀之顧客價值分層後，欲採以決策樹驗證其欄位特徵屬性之定義方式，並加入過去研究資料所缺乏之欄位共同討論，期望經由個案的探討和實務的分析，作為人身保險業者於顧客價值分析之實作依據。

第三章 研究方法

本章將對本研究的研究方法加以闡述，第一部份對於所使用的應用工具做描述，第二部份描述論文整體的研究流程，第三部份針對本研究所應用資料採礦之分群技術以及關聯法則之演算法作解說。

3.1 研究工具

本研究所使用的硬體架構為 AMD Athlon(tm) XP 2000+ 1.8Ghz，搭載 512MB 的記憶體與 160G 的硬碟空間，採用 Microsoft Windows XP SP2 作業系統。

初始資料原始檔案為 Microsoft Access 格式。首先，將資料初步過濾導入 Microsoft SQL Server 2000 做資料轉換，並讓處理過的資料輸入分析資料庫，作為日後資料探勘來源。已儲存的資料在 Microsoft Excel 及 SPSS 作完敘述性統計分析後，匯入 Enterprise Guide 轉換成適合 SAS Enterprise Miner 唯讀的格式，再將轉換過的資料導入 SAS Enterprise Miner，以該軟體作決策樹學習法之分類驗證，而後透過*.txt.格式協助程式編碼，並以資料採礦軟體 weka 針對每一價值層級客做關聯法則採擷。

3.2 研究架構

本研究目的乃運用資料採礦技術，針對個案保險公司之人身保險資料進行價值分群與投保行為分析，並依據分析結果支援個案公司進行二次行銷與客戶開發。

由於原始顧客資料雜亂，無法直接進行分析，因此必須先將資料做正規化處理。首先，依顧客資料屬性做統合，再匯入本研究所建立之資料庫。經 SQL 轉換後的資料，先於 Excel 和 SPSS 進行 RFM 顧客價值指標運算，而後依據指標分數將顧客分群。

以資料採礦技術來說，經常被利用在保險產業的技術包括決策樹 (Decision Tree)、關聯法則 (association Rule) 與類神經網路 (Neural Network)。不同的演算法存在不同的處理方式，對於可以分別處理的資料型態也不盡相同；然而，透過過往文獻【7】以及本研究所獲資料之初始測試發現，對於以名義變數為主要探勘的資料而言，決策樹演算法的預測及分類效率優

於類神經網路。因此，本研究對 RFM 分類效率之檢驗乃採用決策樹演算法。

若透過決策樹演算法所檢測之群組分群錯誤率過高，則返回資料處理步驟，重新定義變數區間並計算 RFM 指標分數，循環往復直至決策樹演算法之分類錯誤率達最低。而後乃採以分類錯誤率最低之欄位定義方式，接續下列研究步驟；同時，對於決策樹演算法所篩選出之重要欄位予以記錄，且以該紀錄欄位做關聯法則之採擷，並依循每一價值層級之關聯法則予以層級策略建議。本章節後方將對決策樹演算法以及關聯法則多做說明。初步研究架構如圖 3.1 所示。

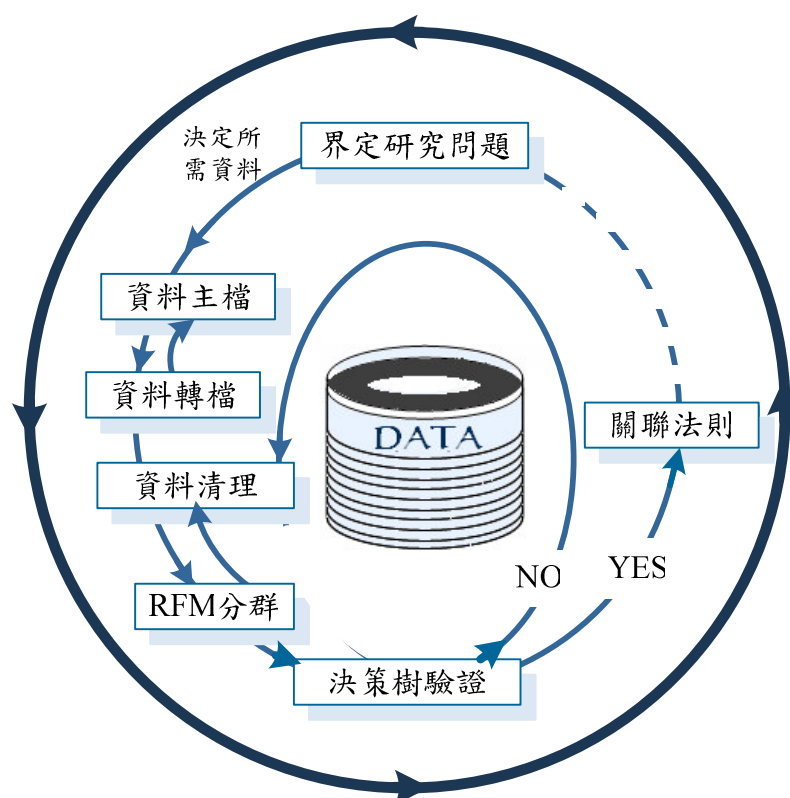


圖 3.1 本研究資料採礦架構

資料來源：本研究整理

3.3 資料蒐集處理與分析

3.3.1 資料蒐集

本研究所取得之初步資料，乃個案金融控股公司之人身保險業務保戶相關資料，資料蒐集期間為民國 82 年 7 月至民國 94 年 12 月，針對該顧客基本資料和過去保戶投保行為做分析。本研究將一張保單設定為一個顧客，整理後共 63512 筆資料。

3.3.2 資料處理

資料處理分為資料格式轉換及資料清理兩個階段。前者乃因個案業者所提供之資料格式無法完全適用於分析軟體，且部份資料分析和法則採擷需要特定的資料格式；後者則是在了解資料領域後，界定資料範圍，並針對不一致的資料或遺失值進行處理，以增加資料的正確性。圖 3.2 與圖 3.3 為資料庫中的資料內容以及資料初步整理之 SPSS 介面。

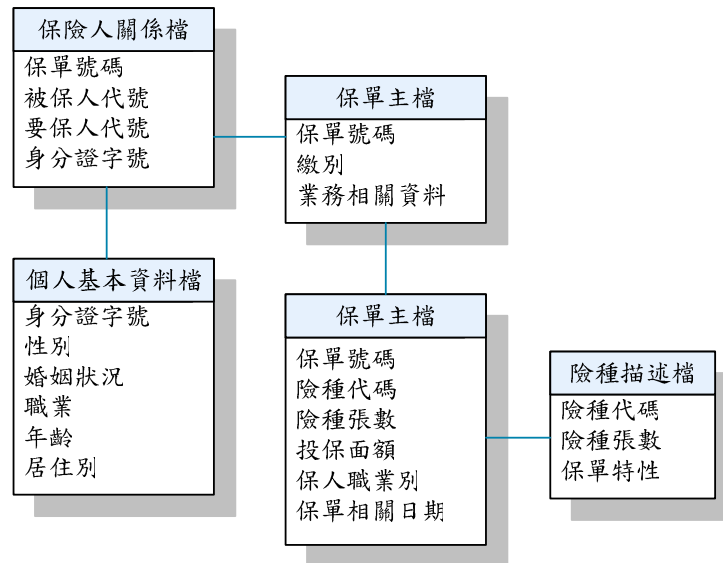


圖 3.2 資料庫中的資料內容

資料來源:本研究整理

no	how_many	main_sub	year	month	pay_year	age	gender	pay_way	money	money_bl	log_money	log_f_money	f_money	first_pro	Guar	marit	depart	zo	charact	lis
1	1	1	89	9	6	18	1	1	10_01_02	1.00	3.47	3.4	2945	13	0	10	00	0	20	
2	2	1	85	7	10	47	1	2	80_01_02	1.90	4.76	4.5	57144	13	1	10	00	0	30	
3	3	1	85	7	10	47	1	2	70_01_02	1.35	3.78	3.4	8083	1	1	10	00	0	30	
4	4	1	88	9	20	52	1	2	30_01_02	1.48	4.01	4.5	10520	1	1	10	00	0	30	
5	5	1	88	9	20	52	1	2	90_01_02	1.95	4.15	4.5	14082	1	1	10	00	3	52	
6	6	1	86	1	20	13	1	2	120_02_03	2.08	3.19	3.4	1564	1	0	10	00	3	50	
7	7	1	87	10	10	14	1	1	200_02_03	2.30	3.05	3.4	1126	1	0	30	00	0	30	
8	8	1	85	7	20	30	1	2	100_02_03	2.00	4.13	4.5	13400	1	1	10	00	0	30	
9	9	1	85	7	20	30	1	2	60_01_02	1.78	3.40	3.4	2520	2	1	10	00	3	52	
10	10	1	90	5	1	51	1	4	50_01_02	1.70	5.88	5.6	****	2	1	10	00	0	20	
11	11	1	85	2	15	46	1	3	150_02_03	2.18	4.65	4.5	44881	2	1	10	04	0	20	
12	12	2	85	2	15	46	1	3	180_02_03	2.26	3.70	3.4	5029	1	1	10	04	3	50	
13	13	1	90	6	20	34	1	2	30_01_02	1.48	4.11	4.5	12960	1	1	10	00	0	30	
14	14	1	90	6	20	34	1	2	5_00_01	.70	3.51	3.4	3260	1	1	10	00	3	51	
15	15	1	90	6	20	34	1	2	60_01_02	1.78	4.03	4.5	10726	1	1	10	00	3	52	
16	16	1	91	10	1	49	1	4	66_01_02	1.82	6.18	6.7	****	1	1	10	00	0	20	
17	17	1	87	12	30	48	1	4	20_01_02	1.30	5.44	5.6	****	3	1	10	07	0	30	
18	18	2	90	12	20	33	1	2	10_01_02	1.00	3.70	3.4	5040	3	0	10	00	0	30	
19	19	2	90	12	20	33	1	2	10_01_02	1.00	3.79	3.4	6180	3	0	10	00	0	31	
20	20	2	90	12	20	33	1	2	60_01_02	1.78	3.78	3.4	6054	1	0	10	00	3	52	
21	21	1	88	6	20	14	1	2	50_01_02	1.70	3.82	3.4	6550	1	0	10	07	0	30	
22	22	1	88	6	20	14	1	2	10_01_02	1.00	3.58	3.4	3820	1	0	10	07	3	51	
23	23	1	88	6	20	14	1	2	60_01_02	1.78	3.03	3.4	1064	1	0	10	07	3	52	
24	24	1	88	6	20	45	1	2	60_01_02	1.78	4.18	4.5	15180	6	1	10	00	0	30	
25	25	1	88	6	20	45	1	2	10_01_02	1.00	3.76	3.4	5760	6	1	10	00	3	51	
26	26	1	89	2	20	59	1	2	60_01_02	1.78	4.06	4.5	11594	1	1	10	00	3	50	
27	27	1	89	2	20	59	1	2	10_01_02	1.00	3.87	3.4	7990	1	1	10	00	3	51	
28	28	3	91	12	20	52	1	2	60_01_02	1.78	4.01	4.5	10211	1	1	20	00	3	50	
29	29	1	88	11	20	56	1	2	30_01_02	1.48	4.10	4.5	12450	1	1	10	00	0	30	
30	30	1	88	11	20	56	1	2	5_00_01	.70	3.54	3.4	3430	1	1	10	00	3	51	

圖 3.3 資料初步整理之 SPSS 介面

資料來源:本研究整理

3.4 RFM 顧客價值指標

初步整理完的資料，透過 RFM 顧客價值指標來計算客戶的價值，並參考學者 Jay Curry【26】的顧客金字塔模型整合個案保險業者對顧客價值的定義，且針對顧客價值指標的分數作客戶等級的分群。

3.5 決策樹學習法

本研究將利用決策樹學習法，歸納由RFM顧客價值指標所分出不同顧客群之特徵屬性，並搭配關聯法則找尋集群屬性，作為未來目標行銷工作的參考。以下就先針對決策樹學習法之生成、修剪以及讀取規則等作介紹。

決策樹 (Decision Tree) 學習法應用的層面相當廣泛，其主要功能乃藉由分類已知的事例來建立樹狀結構，並從中歸納存在事例中的規律；此外，所建構出的決策樹，亦可用以樣本外之預測。

在決策樹演算結果的樹狀圖中，每一內部節點 (Internal Node) 代表對某屬性的測試 (Test)，其下的每一分支 (Branch) 代表此屬性的一個可能值或多個可能值的集合。最後每一葉節點 (Leaf Node) 所對應的是一個目標類別 (Target Class)。圖3.4乃決策樹基本構造的例子，首先根節點為「體溫超過攝氏38度」，根據此判斷將資料切割成兩個分支，而體溫未超過38度的集合已達決策樹的結尾，即稱其為葉子；而體溫超過38度的集合形成新的決策節點，繼續上述步驟直到達到葉子。

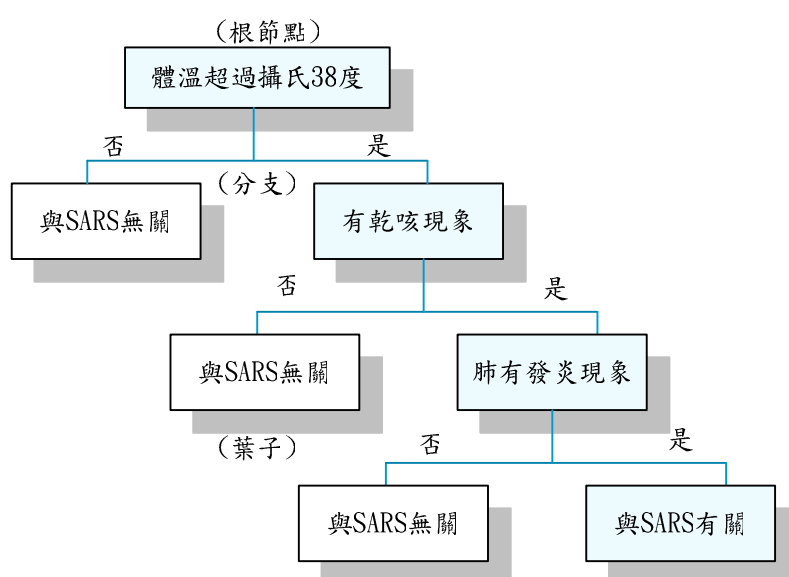


圖 3.4 決策樹的構造

資料來源【7】

傳統決策樹的根部在頂端，欲建立決策樹之資料從根部進入後，應用測驗方法選擇進入下一層的某個子節點。不同的演算法採以不同的方法檢測，但降低子節點內的凌亂度（Disorder），是選擇測試屬性的共同目標。如此不斷重複該過程，直到資料到達葉部節點為止。

若資料不完整或含有雜訊時，所建構的決策樹通常產生過配合資料（Overfits the Data），以致於生成的決策樹過於複雜。通常產生過適現象（Overfitting）的原因有二，一為樣本本身的屬性太多，決策樹學習法容易選用到和種類不相關的屬性，即自由度太大。另一為偏移（Bias），不同演算法在尋找測試屬性時，都有自己的偏好，所以在篩選時可能會依尋演算法所偏好之屬性，而非真正和種類相關的屬性，因此決策樹建構完成後還需要做適當的修剪（Pruning）。

進行決策樹修剪時有二個常見方法：

1. 預先修剪（Pre-Pruning）：

預先修剪以提早停止決策樹生長來達到修剪的目的，直到樹停止生長時，末端節點即成為樹的樹葉。樹葉的標籤（Label），為該節點訓練集合（Training Set）中佔有比例最大的類別。停止決策樹生長的時機是在決策樹建立前，事先設定好一個閾（Threshold），當分支節點滿足該閾值的設定，就停止該分支繼續成長。

2. 事後修剪（Post-Pruning）：

事後修剪是先建立一個完整的樹，再將其分支移除的作法，移除分支的依據是計算該分支的錯誤率(Error Rate)，未被移除的分支節點就變成樹葉。

然而，許多學者認為預先修剪法所需設定的閾值，難免過於主觀，使得目前事後修剪法比較受到歡迎。因此本研究在實證分析修剪決策樹時，也將採用事後修剪法，其概念以圖3.5 說明將更為清楚：

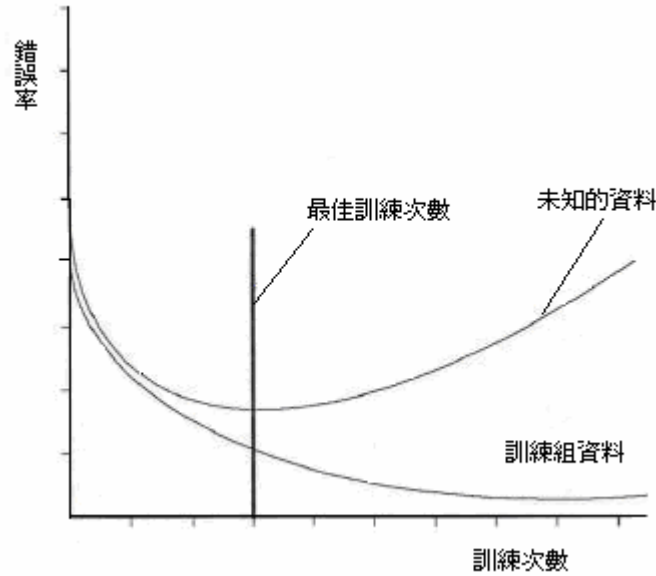


圖 3.5 利用事後修剪法修剪決策樹

資料來源【17】

目前被廣泛使用的數個決策樹演算法，包括CHAID（Chi-Square Automatic Interaction Detector）、CART（Classification and Regression Trees）、ID3(Interactive Dichotometer 3)與C4.5及C5.0等幾種。下表為決策樹各種演算法之特色列表。

表 3.1 決策樹演算法比較表

演算法類別	CHAID	CART	ID3	C4.5	C5.0
年代	1980	1984	1986	1993	1998
提倡者	Kass	Breiman、Friedman、Olshen、Stone	Ross Quinlan	Ross Quinlan	Ross Quinlan
特色	以統計分析(卡方檢定)為主要的分類方法	產生分類樹、回歸樹	產生分類樹	改良ID3而來	為C4.5的商業版本
決策樹的型態	Non-Binary-Tree	Binary-Tree	Non-Binary-Tree	Non-Binary-Tree	Non-Binary-Tree
應用資料型態	分類型	分類型、連續型	分類型	分類型	分類型
決策樹的修剪	Pre-pruning	Post-pruning	Post-pruning	Post-pruning	Post-pruning
變數/分裂點的選擇	卡方檢定、F檢定	Gini Index	Information Gain	Information Gain	Information Gain

資料來源【15】

由於在近年來的實證研究中，透過C4.5演算法來建構決策樹之成果顯示相當成功，故本研究採C4.5演算法來建構決策樹。以下針對C4.5演算法，做更詳細的介紹。

3.5.1 C4.5 演算法

C4.5 演算法是由其前身 ID3 學習法改進而來【4】。ID3 為一決策樹歸納技術，在建構決策樹的過程中，ID3 以資訊獲利 (Information Gain) 為準則，選擇最佳屬性當成決策樹的節點，使最後得到的為一最簡單(或接近最簡單)的決策樹。資訊獲利 (Information Gain) 是當某一屬性為決策樹節點時，其分支節點的加權熵值 (Entropy) 與原物件集合熵值的差距。若某一物件集合 C 分屬於 j 個不同類別，則此物件集合的熵值 $E(C)$ 為：

$$E(C) = - \sum_j p_j \times \log_2 p_j$$

其中 P_j = 屬於類別 j 的物件數 C 的總物件數

此時若選擇某一屬性 A_i 為決策樹節點，便意謂著在此節點下建立 m 個子節點 (假設屬性 A_i 有 m 個可能值)，而被分配到每一個子節點所包含的物件，其屬性 A_i 的值皆相同。因此以屬性 A_i 為決策樹節點時，其分支節點的加權熵值可計算如下：

$$E(A_i) = \sum_{k=1}^m \left(\frac{n_k}{n} \right) \times E(C_k)$$

其中 C_k 為物件集合 C 中其 iA 屬性值相同的物件子集合 k

$E(C_k)$ 為物件子集合 C_k 的熵值

n 為物件集合 C 的總物件數

n_k 為物件子集合 C_k 的物件數

而資訊獲利 (Information Gain) 是指選擇屬性 A_i 為決策樹節點時熵值的改變程度，也就是原來物件集合的熵值，與以 A_i 為決策樹節點時，其分支節點加權熵值的差距，即 $G(A_i) = E(C) - E(A_i)$ 。

建構決策樹的過程中，針對每一個節點，ID3 選擇能達到最大資訊獲利之屬性為此節點的分割條件，一直持續到無法有效降低資料的凌亂度為止，其詳細的演算法如下：

- (一) 從決策樹的根節點 (Root Node) 開始，設定根節點為目前節點 C ，此時所有物件都屬於 C 的物件集合。
- (二) 如果 C 中所有的物件都屬於同一類別 (Class)，則設定節點 C 為此類別然後停止，否則繼續執行步驟三。
- (三) 對 C 的物件集合，計算其熵值 $E(C)$ 。
- (四) 針對所有尚未出現在從根節點到目前節點之路徑的屬 iA_i (稱之為候選屬性)，以 A_i 分別對 C 的物件集合進行分割，並計算其分支節點的熵值 $E(A_i)$ 與資訊獲利 $G(A_i) = E(C) - E(A_i)$ 。
- (五) 選擇資訊獲利最大候選屬性，將之當成節點 C 的分類屬性。
- (六) 在節點 C 下建立子節點 C_1 、 C_2 、...、 C_m (假設所選定的分類屬性有 m 個屬性值)，並依據分類屬性值，將 C 中的所有物件分配至適當的子節點中。
- (七) 將每個子節點 C_i 視為節點 C ，然後由步驟二重複執行。

ID3 選擇分類屬性的方法，對於一般學習問題已有不錯的成果，但以資訊獲利為選擇分類屬性的準則，偏向選擇屬性值較多的屬性。

為了彌補這個缺點，Quinlan【17】發展 C4.5 演算法時，提出將資訊獲利正規化的方法，來緩和分成過多子樹之效應。由此可知 C4.5 演算法改以獲利比 (Gain-Ratio Measure) 作為選擇屬性的衡量準則：

$$GR(A_i) = G(A_i) / IV(A_i)$$

其中 $G(A_i)$ 為屬性 A_i 對 C 物件集合進行分割的資訊獲利

$$IV(A_i) \text{ 為屬性 } A_i \text{ 的資訊值} = - \sum_m P_m \times \log_2 P_m$$

且 P_m 為 C 物件集合中，屬性 A_i 之值為 a_m 的物件數 / C 的物件數和 CART 演算法一樣，C4.5 演算法也需要做事後修剪的工作，但二者有差異，C4.5 演算法是根據預估錯誤率 (Predicted Error Rate) 作為修剪條件，而非整體錯誤率。它的方法是從樹的底部 (樹葉) 往上測試每一個節點所形成的子樹，若是將子樹以一個樹葉代替後得到的預估錯誤率較低，則將子樹修剪為樹葉，否則將保留原來的子樹。

所謂的預估錯誤率，是指由訓練資料的錯誤率來推估其它非訓練資料

的錯誤率。舉例來說，假設被分到某子樹的所有訓練資料個數為 N ，其中有 E 個分類錯誤，則訓練資料的錯誤率就是 E/N 。而直接用觀察到的錯誤率，去推估新資料被測試時也有相同錯誤率，顯然過於樂觀，因此還需另外推算一個較高的錯誤率才合理。計算預估錯誤率的方法有數種，C4.5 演算法採用的是 (N, E) 二項分配 (Binominal Distribution) 機率值，作法是給定一個信心水準 (Confidence Level, CF) 後，計算在二項機率分配下，錯誤率 E/N 的信賴區間，並取上限值作為預估錯誤率，寫為 $U_{CF}(N, E)$ 。其中 N 乃所有訓練資料中分入該子樹的個數，而 E 則是 N 中分類錯誤的資料個數。

3.6 關聯法則

經決策樹驗證之 RFM 顧客價值分群錯誤率達最低後，接續透過關聯法則來分別採擷四個顧客價值層級之關聯屬性，以作為未來二次行銷工作與客戶開發之參考。以下乃先就本研究所採之關聯法則做介紹。

挖掘關聯式法則【11】料採礦中相當重要的一個議題，一般交易資料庫中都儲存著數量龐大的交易資料，而每一筆交易記錄也都記載著相關的項目(item)，包含使用者相關資料及交易的項目或時間等；而資料採礦的技術便可從這些大量的資料中，挖掘出各個項目之間的關聯法則。

關聯規則是指由於某些事件的發生而引起另外一些事件的發生，發現關聯的目的是為了要在交易資料庫中發現項目之間的關係，找出同一個交易中隱含其它項目之存在的項目。其起源在於分析市場購物籃資料(Market Basket Data)中的大量商品集合(Large itemsets)的關聯程度，因此，也稱為市場購物籃分析(Market Basket Analysis)。關聯法則通常應用在市場購物籃分析上，藉由銷售點系統所紀錄的消費者購買紀錄，分析出購買商品之間的關聯性，根據這些關聯性，可以了解顧客的消費習慣，知道顧客是誰、買了些什麼以及為什麼這樣買。藉由此些資料，企業可以得知哪些商品是會被同時購買，及其被購買的相關先後順序；而這些資訊對於企業在做產品銷售計劃及決策時相當幫助。

關聯法則的主要目的是要找出資料項目間的關聯性，其型式為 $X \rightarrow Y$ ，就是當客戶買了某項產品 X 時，同時會去購買 Y 產品，這兩種產品 X 、 Y 之間就是一種關聯。其中 X 與 Y 分別表示在資料庫中不同的資料項目組，

也就是若 X 出現時，同時也可能出現 Y。例如，顧客買了牛奶和麵包時，也會購買果醬；其表示方式如下：

牛奶&麵包→果醬 [support=40% ， confidence=70%]

40% 和 70% 則是代表該條規則的支持度 (Support) 和信賴度 (Confidence)。也就是在所有交易資料中，會購買牛奶和麵包的人佔了總交易量的百分之四十；而在這百分之四十的交易中，會再購買果醬的人佔了百分之七十。這種交易行為的百分比，必須要高於某個程度的比例這條關聯性的法則才算成立。因此以下介紹關連法則裡面重要的參數名詞：

1. 項目組(Itemset)

相異物品的項目組。當一筆紀錄中有{A, B, C}三個項目，則項目組有{A}、{B}、{C}、{A, B}、{B, C}、{A, C}、{A, B, C}，若物品項目有 n 項，則項目組有 $2^n - 1$ 種。因此，當項目有很多時，項目組是很可觀的。

2. 支持度(Support)

即某項目集在資料庫中出現的頻率(frequency)，通常以 $Support(X)$ 表示之，X 為 item set。例如：若資料庫中有 100 筆交易紀錄，其中有 40 筆交易發現有購買牛奶，則此 item set 的支持度為 $40/100=40\%$ 。其支持度越高，代表這個 item set 越值得深入探討

3. 信賴度(Confidence)

指兩個項目組之間的信賴程度，通常以 $\frac{Support(X \cap Y)}{Support(X)}$ 表示之，X, Y 為 item set。也就是在 X 出現的機率下，Y 也會出現的條件機率

4. 支持度門檻值(Min Support)

若某個項目組的支持度太低，則不具參考性，因此，在進行關聯法則前會先訂定一個支持度的門檻值(threshold)，此支持度的門檻值稱為最低支持度(Min Support)。若項目組超過 Min Support 值則稱為高頻項目組(Large itemset, 或稱為 Frequent itemset)。因此，如何去訂定適當的 Min Support 值是很重要的，因為太低的 Min Support 值會探勘出過多的規則而無所適從，太高的 Min Support

值又容易錯失一些具有價值的規則。

5. 信賴度門檻值(Min Confidence)

與 Min Support 意義類似。在進行關聯法則前亦先訂定一個信賴度的門檻值，此信賴度的門檻值稱為最低信賴度(Min Confidence)。若信賴度太低，則不具參考性，且在探勘過程中，未達到 Min Confidence 的關聯法則將會被刪除。

Megiddo and Srikant 【22】指出支持度門檻值及信賴度門檻值之設定是很重要的，當支持度門檻值設太低時，會將重要性較低之項目也包含進來，而設太高又怕因此而失去某些重要規則。信賴度門檻值之設定太低則較不具意義，Han and Kamber 【11】認為門檻值的設定需依據使用者的需求而定。因此關聯法則是否可被保留使用是透過以上係數的設定，必須要滿足支持度是否超過支持度門檻值(Min Support)及信賴度是否大於最小的信賴度門檻值(Min Confidence)而定【1】【2】。

3.6.1 關聯法則分析之演算法

由上述說明可知所謂關聯法則就是某些項目會引發其他項目出現的規則，例如消費者購買雜誌後，通常有 75%的人會同時買咖啡，此時「買雜誌→買咖啡」便是一個法則。而法則的挖掘，就是在龐大的資料中，把一些資料項目的相關性找出來。以交易資料庫為例，每日存在之交易量將積累而成龐大資訊，無法由人進行分析來找出商品之間的關聯性，然而這些交易記錄事實上卻隱含了許多有用的資訊（如消費者購買習慣），若能運用適當的方法將其採擷出來，便可能發現商機創造利潤，挖掘關聯法則的演算法也就是在這種需求驅使下所產生出來的方法。以下就本研究所使用的 Apriori 演算法做一介紹。

3.6.2 Apriori 演算法

Agrawal 等學者於 1994 年提出了 Apriori 演算【1】，目前已是探討關聯法則最具代表性的演算法，其後雖又許多針對不同情況所提出的演算法，但大多都依據 Apriori 加以延伸改進。Apriori 演算法的探勘過程主要分成兩個階段：

1. 辨識常出現的項目集合

先找出滿足最小支持度的所有項目組，也就是在交易資料庫中包含項目組的交易數量必須大於或等於所指定的最小支持數量。然後，這些滿足最小支持數量的項目組就稱之為高頻項目組(large itemsets or frequent itemsets)，若一個項目組包含有 K 個項目，稱之為 K 項目組(K-itemsets)；若某 K-項目組滿足最小支持數量，即稱之為高頻 K 項目組(large K-itemsets)。

2. 產生關聯法則

根據前階段所找出的高頻項目組及以最小信賴度為條件，計算出所有符合的關聯規則；例如 ABC 為高頻 3 項目組， $A、B、C \subseteq I$ ，假如關聯規則 $AB \rightarrow C$ 滿足最小信賴度，則此關聯規則成立。Apriori 演算法表示式如下：

```

 $L_1 = \{\text{Large 1-itemsets}\};$ 
For (k=2;  $L_{k-1} \neq \emptyset$ ; k++)
  { $C_k = \text{Set of Now Candidates}$ ;
  for all transactions t  $\in D$ 
    for all K-subsets s of t
      if (s  $\in C_k$ ) s.count++;
   $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsupp}\}$ 
  Set of all large itemsets =  $\bigcup_k L_k$ ;
  }
```

舉個例子來說明 Apriori 演算法推導關聯法則的產生過程。在資料庫 D 中有 4 筆已排序的交易紀錄，如圖 3.6 所示。另外設 $\text{Min Support} = 50\%$ ，也就是超過 $4 * 50\% = 2$ 筆的資料可視為 Large itemsets； $\text{Min Confidence} = 80\%$ 。在第一次掃描時，如圖 3.2 所示，找出資料庫中所有 1-itemset 並 count 出現次數放在 C_1 中，在 C_1 中 count 次數超過 Min Support (超過 2 筆紀錄)，則視為 Large itemsets，把結果放在 L_1 中。

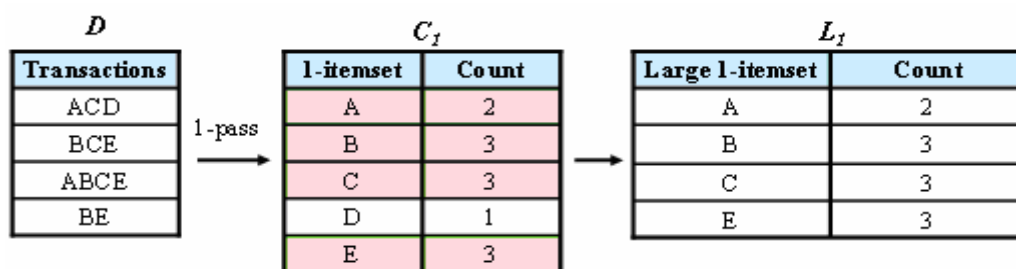


圖 3.6 高頻項目組之產生過程(第一次掃描)

資料來源【1】

接下來的掃描，如圖 3.7 所示，用聯集(join)的方式，如第二次掃描的過程，即將 $L_1 * L_1$ (*為 join 的運算元)，來產生 C_2 的 candidate itemsets，並篩選(prune)出超過 Min Support 的 itemsets，放到 L_2 ；直到 join 出來無法產生新的 Large itemsets。

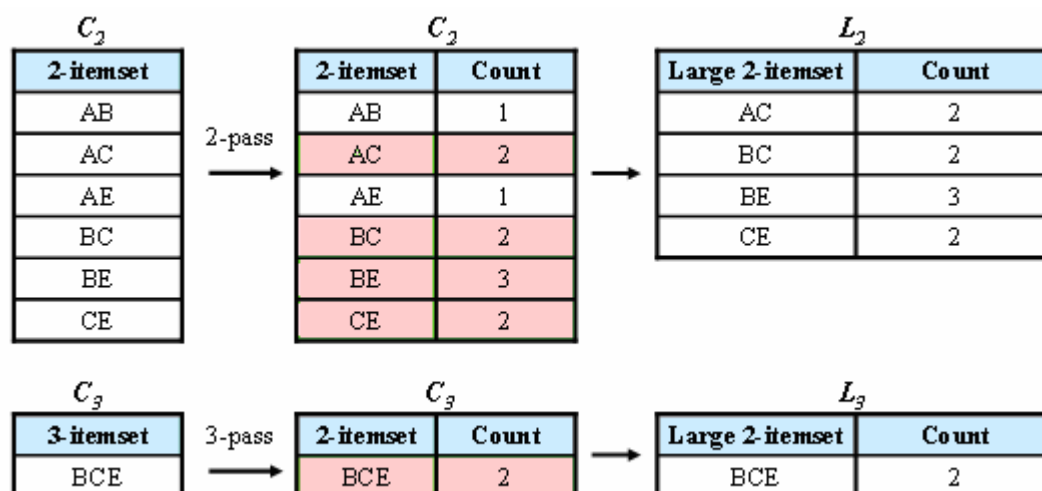


圖 3.7 圖高頻項目組之產生過程(第二次以後的掃描)

資料來源【1】

當找完所有的 Large itemsets 後，將 $L_1 \sim L_3$ 中所有的 itemsets 彙整，查若 $(k-1)$ -itemsets 為 k -itemsets 的子集合，則將 $(k-1)$ -itemsets 刪除；例如， $\{A, B\}$ 、 $\{A, D\}$ 、 $\{B, D\}$ 是 $\{A, B, D\}$ 的子集合，所以 $\{A, B\}$ 、 $\{A, D\}$ 、 $\{B, D\}$ 會被刪除。最後加入 Min Support 來考慮，所得之 Large itemsets，如圖 3.8 所示。

Large itensets	Count
BCE	2
AC	2

圖 3.8 最後所得高頻項目組

資料來源【1】

在了解整體研究工具以及所採用之演算法後，本研究接續將導入個案金融控股公司所提供之人身保險業務資料，搭以研究架構進行顧客價值研究。先將轉換與清理後之原始檔匯入資料庫，透過基本統計方法做基本分析及價值分群並進行驗證，而後再對已驗證之分群進行單一層級法則萃取，佐以二次行銷與顧客開發之策略建議。

第四章 研究分析

本章節將依據文獻中探討過的資料採礦流程逐一進行知識挖掘。先確立採礦的目標並界定資料範圍，再對資料進行處理和轉換程序，透過敘述性統計之剖析，篩選所需資料用以劃分顧客價值分群，佐以驗證後萃取法則。

4.1 目標資料處理與敘述統計分析

4.1.1 目標資料轉換

若對資料來源之含括範圍未有充足認識，容易使得資料在定義與處理過程間出現失當，致使往後所萃取出之知識不具可用性，因此，首先將對取得資料做初步篩選以及瞭解。

本研究透過個案金融控股公司取得 65,261 筆之人身保險資料，將其整合為保單與保戶兩部份，如表 4.1 保單基本資料表以及表 4.2 保戶基本資料表所示，資料經篩選後餘有 23 個欄位，共 63,512 筆完整資料。

表 4.1 保單基本資料表

項目	欄位名稱	轉換後名稱	描述	欄位型態
1	保單號碼	NO.	投保人身分證字號轉換	數值
2	保單識別碼	ID_NO.	保單識別碼	數值
3	主附約	Main_Sub	保單位主約或附約	類別
4	生效年	Year	保單始生效年	數值
5	生效月	Month	保單始生效月	數值
6	業務單位	Depart	投保單位	類別
7	繳費年期	Pay_Year	總繳費期間	數值
8	保額	Money	投保金額(萬元)	數值
9	首期保費	First_Money	首期保費繳額(元)	數值
10	險別分類	Kind	保單型	類別
11	特性分類	Character	保單類別	類別
12	保單終止日	End	保單終止日	數值
13	險種版數	Edition	險種版數	數值

資料來源：本研究整理

表 4.2 保戶基本資料表

項目	欄位名稱	轉換後名稱	描述	欄位型態
1	被保人代碼	NO.	投保人身分證字號	數值
2	要保人代號	ID_NO	被保人身分證字號	數值
3	投保年齡	Age	投保年齡	數值
4	婚姻別	Marry	婚姻狀態	文字
5	保單張數	How_Many	該保人擁有保單張數	數值
6	職業別	Career	職業級別	文字
7	繳費別	Pay_Way	繳費方式	文字
8	性別	Gender	性別	類別
9	吸煙別	Somke	吸煙與否	類別
10	地區別	Zone	居住地區	類別

資料來源：本研究整理

為方便資料後續處理作業，本研究採用 SQL 轉換資料，先將資料轉換成.txt 的格式，並採以 Excel 活頁簿與 SPSS 進行資料初步分析。

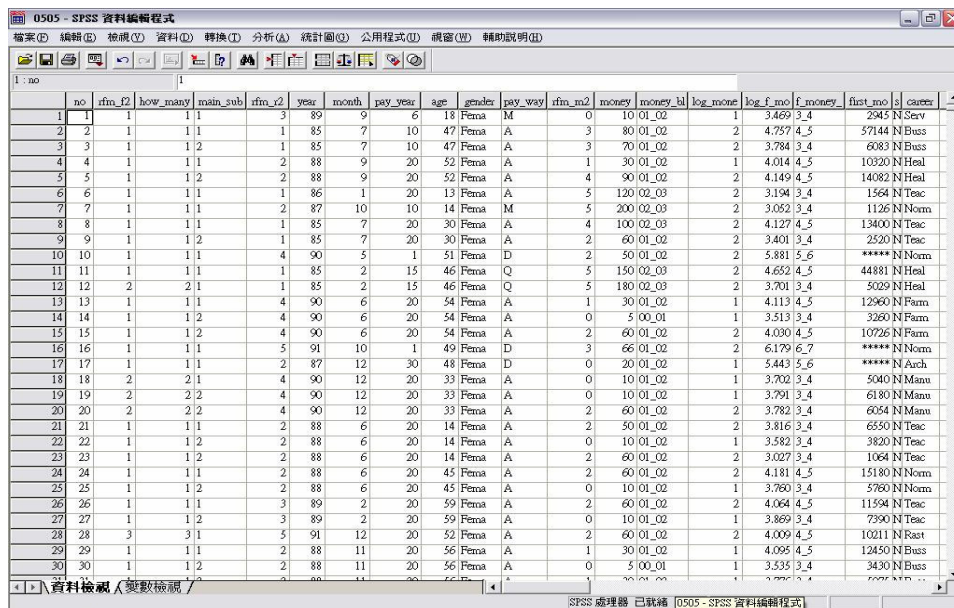


圖 4.1 SPSS 資料處理介面

資料來源：本研究整理

4.1.2 敘述性統計分析

本研究乃著眼於顧客價值分群並針對每一層級資料作關聯法則萃取，然若對資料狀況掌握越深，資料所呈現的訊息則將益愈豐富，對於每筆資料接續之處理方式也會更為恰當。因此，本研究首先以 SPSS 與 Excel 圖表分別對保單屬性變數與保戶屬性變數探討。

在整份顧客交易資料中，保戶年齡主要集中於 40 歲以下，其中又以 0 至 10 歲與 21 至 30 歲之人數為多，31 歲以上保戶以已婚者居多，且男性多於女性，如圖 4.2 以及表 4.3 所示。

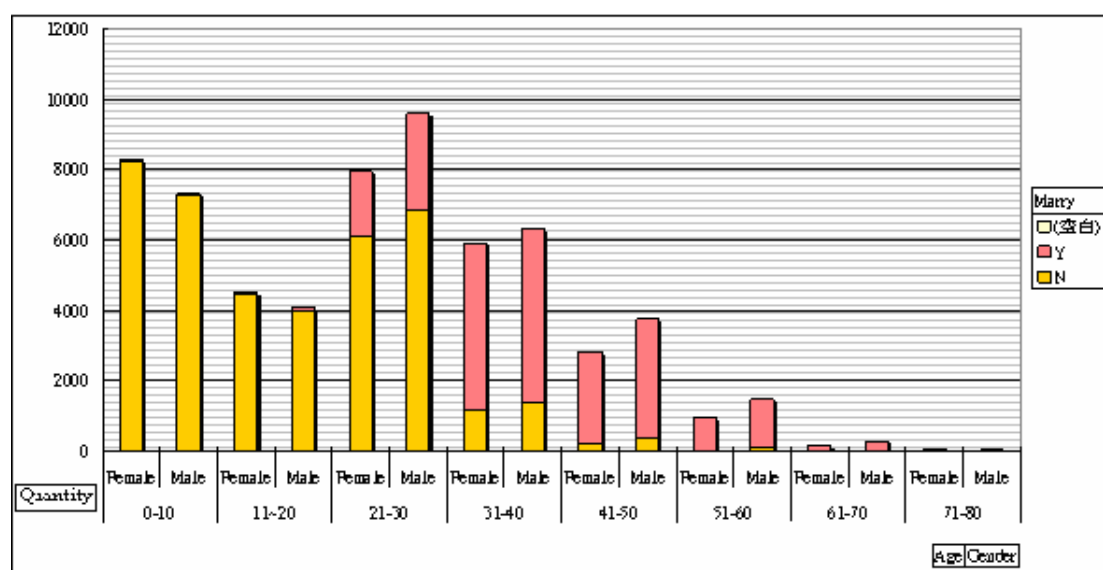


圖 4.2 保戶之年齡別、性別與婚姻別樞紐分析圖

資料來源：本研究整理

表 4.3 保戶之年齡別、性別與婚姻別樞紐分析表

Age	Gender	N	Y	總計
0-10	Female	8235	32	8267
	Male	7268	38	7306
0-10 合計		15503	70	15573
11~20	Female	4476	23	4499
	Male	3962	107	4069
11~20 合計		8438	130	8568
21-30	Female	6102	1834	7936
	Male	6857	2724	9581
21-30 合計		12959	4558	17517
31-40	Female	1174	4742	5916
	Male	1390	4945	6335
31-40 合計		2564	9687	12251
41-50	Female	195	2622	2817
	Male	376	3368	3744
41-50 合計		571	5990	6561
51-60	Female	17	964	981
	Male	91	1381	1472
51-60 合計		108	2345	2453
61-70	Female	3	168	171
	Male	3	295	298
61-70 合計		6	463	469
71-80	Female	3	63	66
	Male	1	53	54
71-80 合計		4	116	120
總計		40153	23359	63512

資料來源：本研究整理

為使資料完整呈現，在保戶職業別方面，本研究選取所有職業別而無作其他篩選，欄位定義採英文單字前四個字母以為辨識，由上至下分別為建築(Arch)、商業(Buss)、娛樂(Ente)、農(Farm)、漁(Fish)、健康事業(Heal)、家管(Home)、資訊(Info)、製造(Manu)、軍事(Mili)、礦採(Mine)、新聞(News)、一般(Norm)、安全事業(Poli)、公共事業(Publ)、餐廳(Rast)、宗教(Reli)、服務(Serv)、文教(Teac)、交通(Traf)、木頭(Wood)等二十一個項目；其中以文教業與一般工作者較多，而商業、服務業以及家管居次。繳費方式上定義年繳(A)、躉繳(D)、月繳(M)、季繳(Q)、半年繳(S)五種繳別，保戶多採用年繳，其次為月繳與躉繳(即一次將保費繳清之繳費方式)，如表 4.4 與圖 4.3 所示。

表 4.4 保戶繳別與職業別之樞紐分析表

計數	Pay_Way					總計
	A	D	M	Q	S	
Career	A	D	M	Q	S	總計
Arch	687	24	98	55	104	968
Buss	3007	174	499	301	345	4326
Ente	63	5	12	9	16	105
Farm	247	12	21	16	47	343
Fish	38	1	1	2	4	46
Heal	1137	181	160	84	86	1648
Home	2948	702	526	244	339	4759
Info	495	25	97	42	35	694
Manu	2272	89	450	203	211	3225
Mili	615	19	322	45	53	1054
Mine	9		3			12
News	73	5	10	3	18	109
Norm	7103	1335	2700	551	542	12231
Poli	171	7	24	21	27	250
Publ	499	37	47	21	23	627
Rast	651	26	150	90	81	998
Reli	50	19	6	1	1	77
Serv	4039	230	855	354	396	5874
Teac	19045	1445	2820	959	1255	25524
Traf	418	12	125	35	41	631
Wood	6		5			11
總計	43573	4348	8931	3036	3624	63512

資料來源：本研究整理

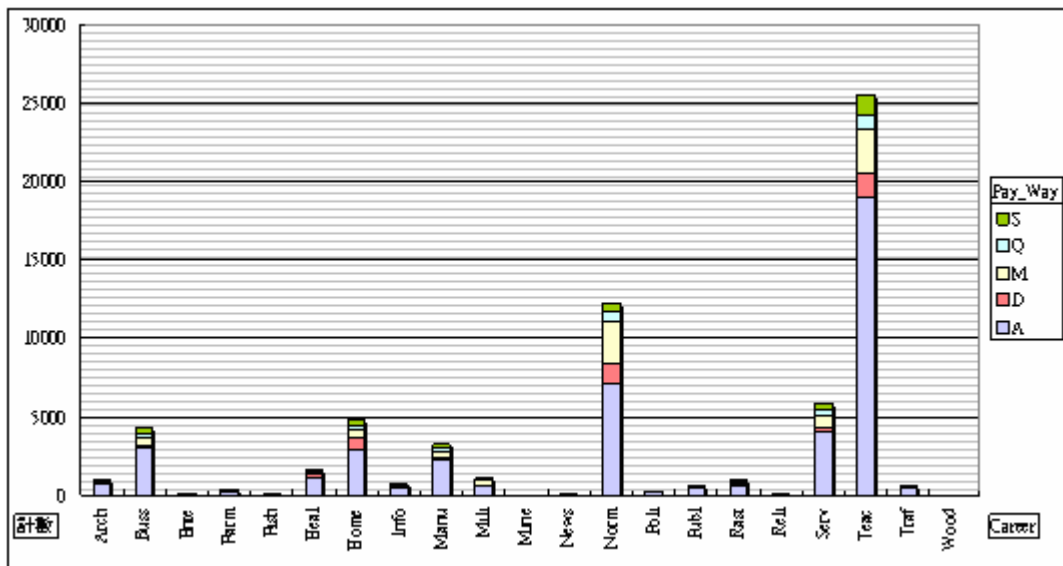


圖 4.3 保戶繳別與職業別樞紐分析圖

資料來源：本研究整理

在保單變數定義方面，由於保戶投保額度之全距過大，因此，在保額及首期保費之欄位處理上予以對數處理，將保額區隔為一萬元以下

(00_00)、一萬至十萬元(00_01)、十萬至一百萬元(01_02)、一百萬至一仟萬元(02_03)、一仟萬元以上(03_04)等五個區間，並定義約種(Main_Sub)之主約為 1 附約為 2。而在險別分類(Kind)上，一般將其區分為生存險(10)、生死合險(20)、死亡險(30)三種，然為依循原始資料，本研究加入投資型(1)、健康險主約(50)、健康險附約(51/52)與年金險(11/12)四個類別共同分析。

由圖 4.4 與表 4.5 可知，保戶投保金額主要集中於十萬至一百萬之間，但主約保額高於一百萬者達 15,867 人，占總保戶之四分之一。主約投保人數以生死合險、死亡險及健康險主約為多，附約保戶則集中健康險附約。

表 4.5 保額、主附約與險別分類之樞紐分析表

計數的Smoke		Money Block					總計
Main_Sub	Kind	00_00	00_01	01_02	02_03	03_04	
1	1			478	99		577
	10			15			15
	11		341	127	4		472
	12		12	11	1		24
	20		221	10184	2557	36	12998
	30			9438	7331	19	16788
	50			6621	5875		12496
1 合計			574	26874	15867	55	43370
2	20			45	5		50
	30	138	645	1117	579		2479
	51		2517	6731			9248
	52			8365			8365
2 合計		138	3162	16258	584		20142
總計		138	3736	43132	16451	55	63512

資料來源：本研究整理

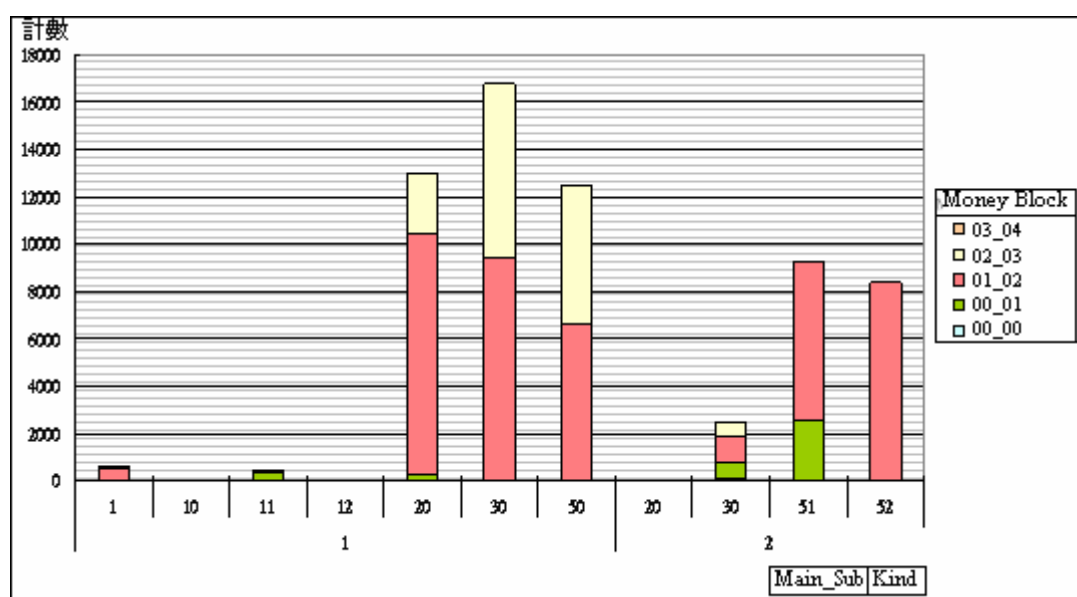


圖 4.4 保額、主附約與險別分類之樞紐分析圖

資料來源：本研究整理

另一方面，在特性上定義險別為六個特性種類，分別為一般壽險(A)、理財壽險(B)、一般年金(C)、健康險(E)、投資型商品(H)與年金險(I)。其中，以一般壽險保戶最多，而健康險居次，各約占總保戶四成七之比例，而投保金額亦集中於十萬至一百萬之間，如圖 4.5 與表 4.6 所示。

就成交量而言，民國 87 至 92 年之成交數目較為顯著，以民國 89 年達其高峰，而該些成交約種主要乃一般壽險以及健康險，然健康險多以附約形式簽訂。如圖 4.6 所示。

表 4.6 保額、主附約與險別特性之樞紐分析表

Main_Sub	Character	00_00	00_01	01_02	02_03	03_04	總計
1	A			18530	9059	27	27616
	B		221	1107	829	28	2185
	C		353	138	5		496
	E			6621	5875		12496
	H			478	99		577
	I						
1 合計			574	26874	15867	55	43370
2	A	138	645	1162	584		2529
	E		2517	15096			17613
2 合計		138	3162	16258	584		20142
總計		138	3736	43132	16451	55	63512

資料來源：本研究整理

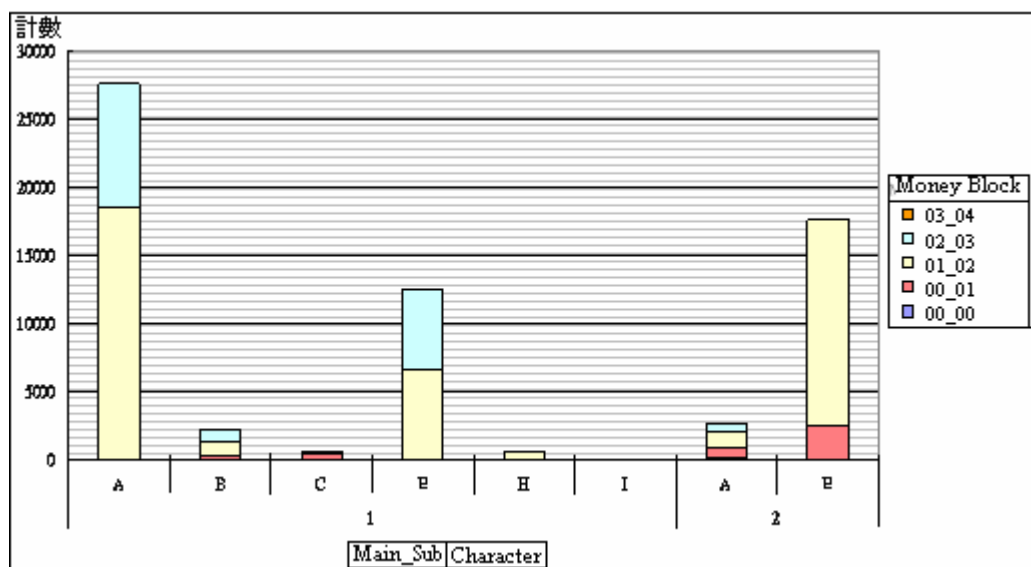


圖 4.5 主附約、保額與險別特性之樞紐分析圖

資料來源：本研究整理

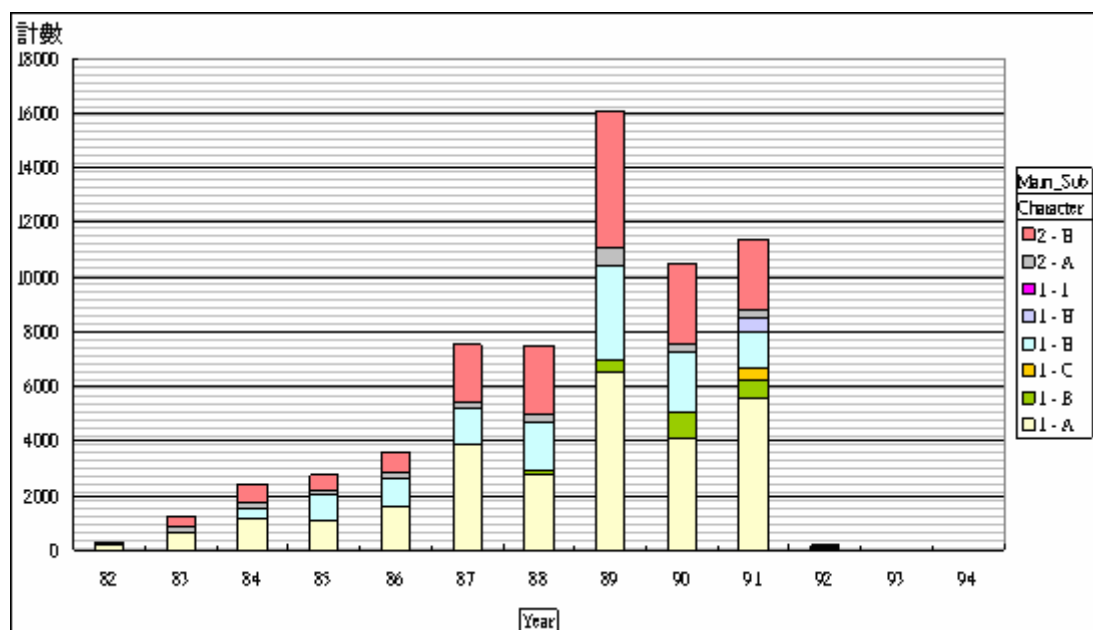


圖 4.6 保額、主附約與險別特性之樞紐分析圖

資料來源：本研究整理

綜觀而言，保額平均約在 60 萬元左右，而首期保費平均約為 20 萬元，且保費約在 16 年半左右繳清；男性在首期保費之繳納能力較女性為強，但就保額平均數而言則男女不相上下；婚姻別上，已婚者之保額與首期保費均較未婚者高；職業別方面，保額平均數約落於 60 萬元左右，唯娛樂事業、漁業及健康事業工作者保額較高，然繳費期間仍呈現一致，約為 20 年不等。如表 4.7 所示。

表 4.7 保戶屬性與保單屬性交叉分析表

CAREER	PAY_YEAR	First Money	MONEY
Arch	19.12	43774.68	67.06
Buss	18.29	47712.77	61.72
Ente	16.72	333525.23	81.52
Farm	18.12	109105.59	63.22
Fish	22.23	95906.00	90.21
Heal	16.93	194377.29	77.77
Home	16.49	226517.61	62.93
Info	18.08	51792.83	59.83
Manu	18.56	53268.64	63.78
Mili	17.93	19947.18	50.36
Mine	20.00	8375.67	46.42
News	18.58	17513.98	60.15
Norm	16.89	152598.88	62.19

CAREER	PAY_YEAR	First_Money	MONEY
Polh	18.69	28398.36	63.49
Publ	18.30	40934.57	59.52
Rast	18.68	26878.13	58.65
Reli	14.58	297413.03	57.65
Serv	18.30	35905.58	63.74
Teac	16.17	44093.27	59.70
Traf	18.76	26879.21	63.89
Wood	19.55	4033.55	52.73
GENDER	PAY_YEAR	First_Money	MONEY
Female	17.20	68651.82	62.92
Male	16.86	96289.96	60.55
MARRY	PAY_YEAR	First_Money	MONEY
N	16.67	45818.19	59.87
Y	17.64	146679.13	64.81

資料來源：本研究整理

透過統計分析，能以各類圖表呈現資料基本訊息，然而，一旦資料變數或欄位達三個以上，單一統計方法所能表達之訊息就相當有限，儘管能將所有維度都納入一併討論，但過多的資訊卻使人難以解讀，因此，本研究接續採用資料採礦方法，做進一步分析。

4.2RFM 顧客價值分群及決策樹驗證

4.2.1RFM 顧客價值分群

RFM 為「最近購買時間(Recently)」、「購買頻率(Frequency)」、「購買金額(Monetary Amount)」三個英文字之縮寫，用以評量顧客忠誠度、顧客貢獻度以及未來的消費能力，利用此三指標來衡量企業與顧客的關係，進而判定顧客價值【26】。因此，根據個案公司定義，本研究分別以目標資料中最近一次投保時間、保單張數、近年保額花費，代表個案公司保險部門之 RFM 指標。而後根據資料量度，定義欄位選項之指標分數，並將所定義欄位之分數予以加總，作為顧客 RFM 指標值。經驗證後最佳 RFM 指標分數定義如表 4.8 所示。

表 4. 8RFM 指標欄位分數定義表

		欄位分數
Recently 近一次投保時間	民國 82~83 年	0
	民國 84~85 年	1
	民國 86~87 年	2
	民國 88~89 年	3
	民國 90~91 年	4
	民國 92 年以後	5
Frequency 近年保單張數	一張以下	0
	二張	1
	三張	2
	四張	3
	五張	4
	六張以上	5
Monetary Amount 近年保額花費	0~20 萬元	0
	20~40 萬元	1
	40~60 萬元	2
	60~80 萬元	3
	80~100 萬元	4
	100 萬元以上	5

資料來源：本研究整理

派伯斯和羅傑【26】認為，一旦確認了顧客且擁有顧客資料，下一個步驟就是以他們對公司的價值作衡量，並對其進行區隔。意即當論及顧客對一企業之長期利潤度時，面對所有顧客並非一視同仁。顧客關係管理之目標在於留住顧客，但留住的該是最有價值的顧客。

Jay Curry 和 Adam Curry【26】亦曾對此闡述，他們認為若想了解顧客對於公司收益以及利潤影響之差異性，最好的方法就是利用顧客價值去建構「顧客金字塔」，一如圖 4.7 所示的。

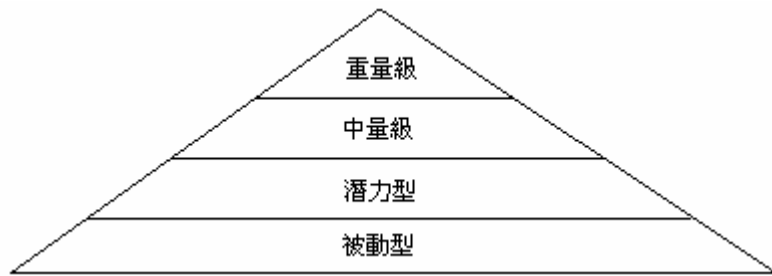


圖 4.7 顧客金字塔示意圖

資料來源【26】

Jay Curry【26】對於如何建構顧客金字塔，給予下列建議：

首先乃對公司頂端 5%之顧客做確認，因從銷售收益角度觀之，該 5%顧客之營收往往占總收入之 45%~50%，故將之視為重量級顧客。其次為中量級顧客，即接續重量級顧客後之 15~25%的顧客歸之，其消費額約占總收入之 30%左右。而後，將剩餘部份之前 60%左右顧客歸為潛在顧客與隱性顧客，該類別顧客對總收入貢獻度約 15%。最終，定義底層 10%左右之顧客為被動型顧客。儘管其曾與企業有業務往來，然近期並無消費紀錄，對該企業產品及服務缺乏需求者，消費約占總收益的 5%左右。

根據個案保險業者資料與前述之 RFM 指標所給予的指標分數，可計算出個案保險業者之 RFM 指標分數以及分布圖，如表 4.9 與圖 4.8 所示。

表 4.9 RFM 指標分數表

RFM_Score	次數	百分比	有效百分比	累積百分比
2	531	.8	.8	.8
3	4337	6.8	6.8	7.6
4	9580	15.0	15.0	22.6
5	9132	14.3	14.3	36.9
6	11626	18.2	18.2	55.2
7	9832	15.4	15.4	70.6
8	7870	12.3	12.3	82.9
9	4770	7.5	7.5	90.4
10	3666	5.7	5.7	96.1
11	1648	2.6	2.6	98.7
12	510	.8	.8	99.5
13	187	.3	.3	99.8
14	108	.2	.2	100.0
15	22	.0	.0	100.0
總和	63819	100.0	100.0	

資料來源：本研究整理

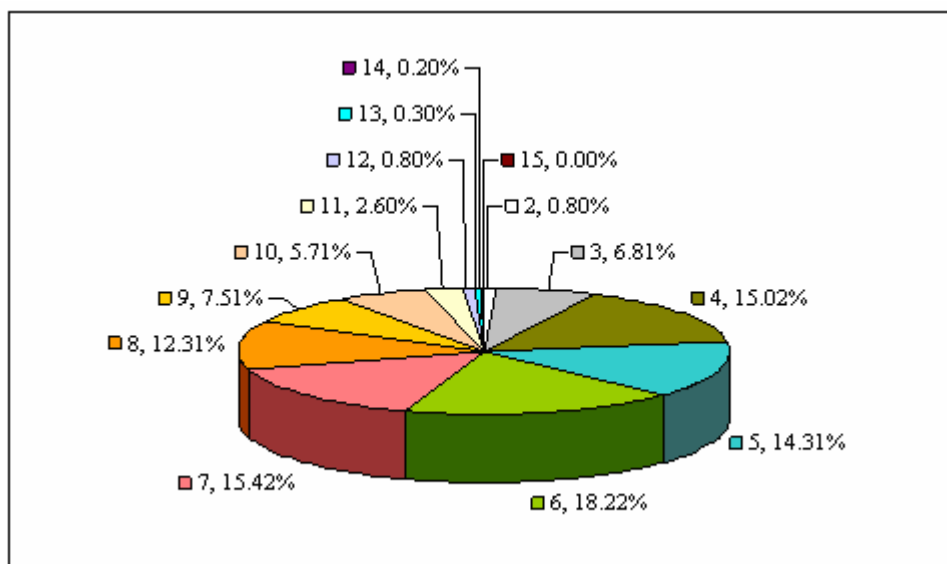


圖 4. 8RFM 指標分數分布圖

資料來源：本研究整理

在了解 RFM 指標分佈後，本研究依據前所提及之顧客金字塔理論以及個案保險業者對該顧客價值定義做整合，將 2 至 15 之 RFM 指標分數依等級比例切割為四個層級，如表 4.10 所示。

表 4. 10RFM 指標等級分布

顧客等級	4	3	2	1
RFM 指標分數分布	2~3	4~7	8~10	11~15
人數	4868	40170	16306	2475
累積人數	4868	45038	61344	63819
百分比	7.6	62.9	25.5	3.9
累積百分比	7.6	70.5	96	99.9

資料來源：本研究整理

4.2.2 決策樹驗證分群

由於保單張數、投保金額以及投保時間三個欄位在進行 RFM 顧客價值分群之時，以為分類指標欄位，因此，在進行決策樹演算法的同時將該三項欄位予以排除，以增加顧客層級分群驗證之準確性。

為測試依據 RFM 理論所做之層級區別是否真能將顧客準確區隔開來，本研究以決策樹之混淆矩陣作為評判依據。利用預測分類和實際分類編排出混淆矩陣，並計算分類正確率，以其作為分類適切與否之評估指標。

分類正確率公式如下：

$$1. \text{分類正確率} = \frac{\text{觀察值總數} - \text{分類錯誤之觀察值總數}}{\text{觀察值總數}}$$

2. 混淆矩陣(Confusion matrix)

矩陣縱行表目標分類，橫列則表推論分類，而每 j 縱行第 i 橫列的元素值，代表當屬於第 j 種分類，被推論為第 i 種分類的範例數。若混淆矩陣對角線上之元素值大而非對角線上的元素值小，則代表分類結果良好。

應用軟體對於葉節點之預設值為 2，因素信賴度為 0.2，即葉節點例子個數在 2 以上且因素信賴度在 0.2 以上者，會表現在類別路徑 (path) 之內。如此設定會使得決策樹之生成過於龐大，並產生過適化現象，因此，對於兩參數之設定本研究認為有作調整的必要。經反覆測試，將決策樹規範在葉節點最小例子個數為 4 以上、因素信賴度最低為 0.25 且決策樹最大深度 (層) 為 6 的設定下，所產生之節點表現最佳。

在分類正確率測試方面，乃針對已定義且分類好的資料做測試。若矩陣之分類錯誤率過高，則回溯至資料處理步驟，重新定義並篩選資料進行分類，而後再經決策樹演算法測試；反之，若矩陣分類錯誤率已達最低，則接續研究流程對每層級顧客作關聯法則之萃取。圖 4.9 為本研究在 SAS Enterprise Miner 測試各類資料定義之於混淆矩陣正確率之作業執行介面。

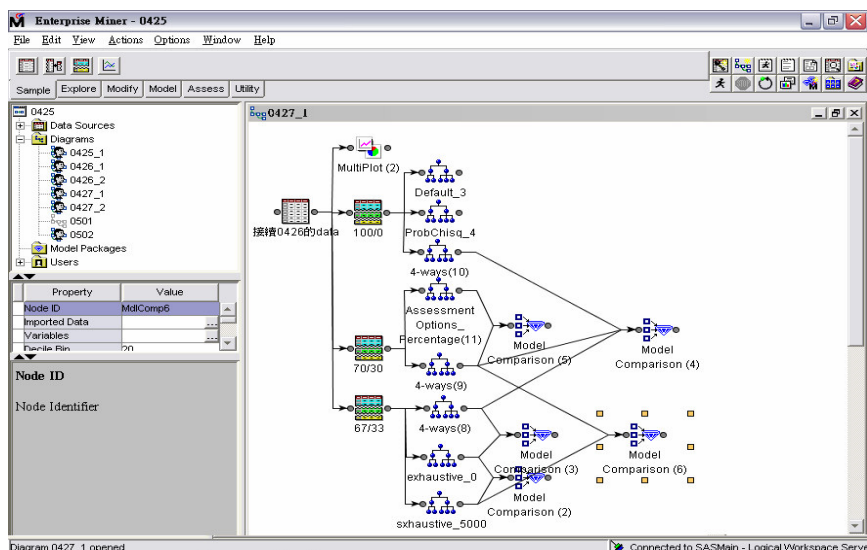


圖 4.9 SAS Enterprise Miner 執行介面

資料來源：本研究整理

一般而言，任一次決策樹演算法之執行會產生 Score Rankings、Tree Map、Tree、Leaf Statistics、Fit Statistics 與 Output 六張結果顯示圖表，如圖 4.10 所示。透過 Score Rankings 所呈現之收斂曲線能決定資料最適訓練 (Training) 比例以及驗證 (Validate) 比例，一般以資料之三分之二用以訓練，然經測試後本研究選取表現較佳之 70% 與 30% 為訓練及驗證比例；Tree Map 與 Tree 則予以決策樹兩種不同型態之圖示，但 Tree 對於枝葉的表示較詳盡明白；Leaf Statistics 乃每枝葉所占總資料比例；Output 則將決策樹各類細部資料列以報表，而其中包含混淆矩陣。如圖 4.11 所示。

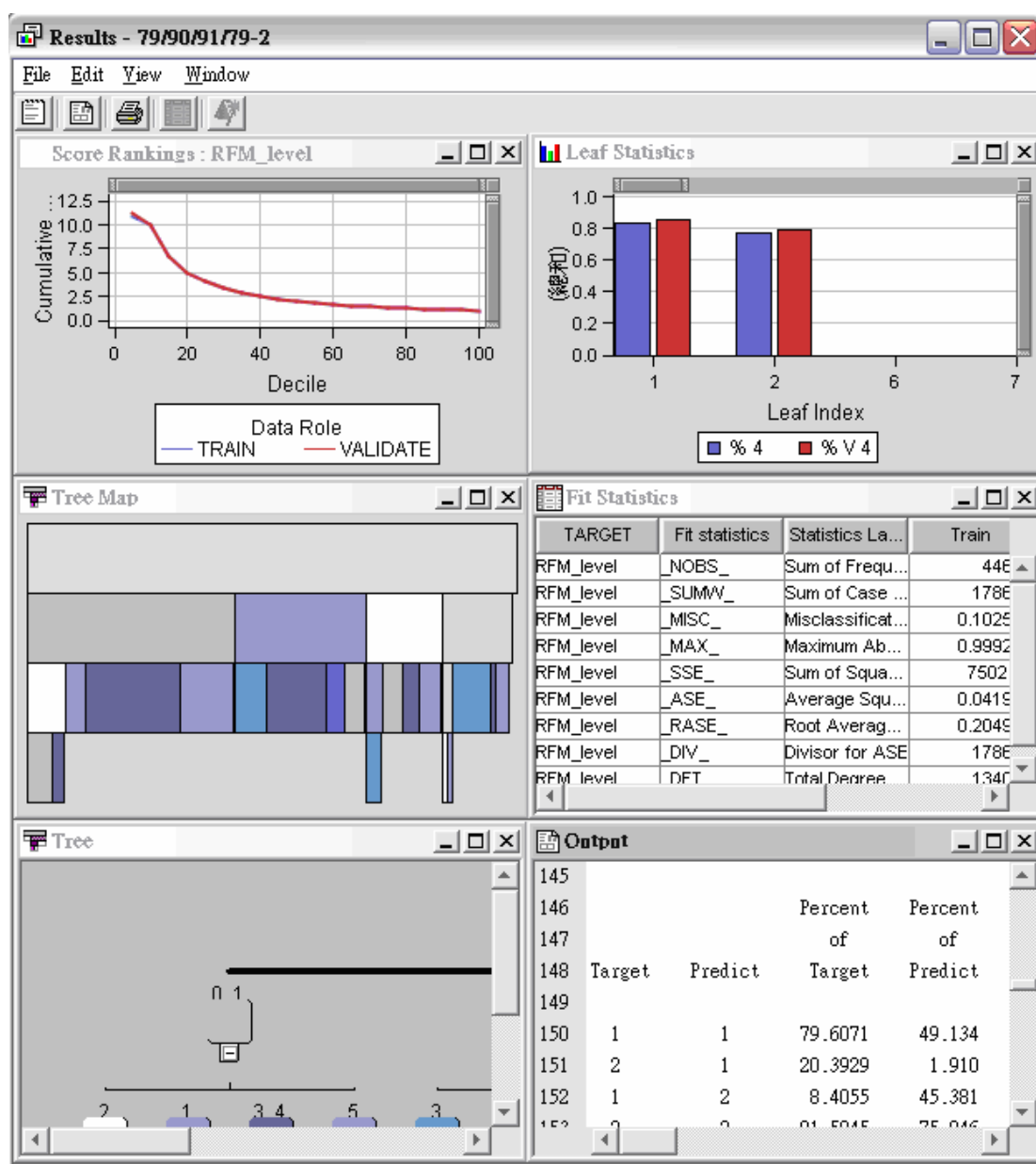


圖 4.10 Result of EM SAS Decision Tree 介面

資料來源：本研究整理

Target	Predict	Percent of Target	Percent of Predict	Count	Percent of Total
1	1	79.6071	49.134	851	1.9051
2	1	20.3929	1.910	218	0.4880
1	2	8.4055	45.381	786	1.7596
2	2	91.5945	75.046	8565	19.1739
1	3	0.3168	5.485	95	0.2127
2	3	8.7693	23.044	2630	5.8876
3	3	90.9139	96.970	27266	61.0387
3	4	20.0047	3.030	852	1.9073
4	4	79.9953	100.000	3407	7.6270

圖 4.11 EM SAS 之決策樹混淆矩陣介面

資料來源：本研究整理

由於最佳資料訓練比例為七比三，因此，已定義之 63,512 筆資料將依該比例投入訓練與驗證階段。經 32 次測試後，表現最佳之分類混淆矩陣如表 4.11 與表 4.12 所示。訓練資料集 1 至 4 等級之顧客價值分類正確率分別為 79%、91%、90%及 79%，平均分類正確率達 84.75%，而平均分類錯誤率為 15.25%；而驗證資料集之等級分類正確率分別為 81%、91%、91%及 82%，平均分類正確率達 86.25%，平均分類錯誤率為 13.75%。圖 4.12 乃透過 EM SAS 將混淆矩陣分析表所轉成之示意圖，紅色為分類正確之部分，而藍色則為分類錯誤部分。於軟體操作介面上，可透過游標移動查看每一分類正確及分類錯誤之細部資訊。

表 4.11 決策樹之混淆矩陣分析表(Train)

真實類別\預測類別	1	2	3	4	總數
1	851	786	95	0	1732
2	218	8565	2630	0	11413
3	0	0	27266	852	28118
4	0	0	0	3407	3407
總數	1069	9351	29991	4259	44670
分類正確率	79%	91%	90%	79%	

資料來源：本研究整理

表 4. 12 決策樹之混淆矩陣分析表(Validate)

真實類別\預測類別	1	2	3	4	總數
1	342	350	51	0	743
2	78	3727	1088	0	4893
3	0	0	11737	315	12052
4	0	0	0	1461	1461
總數	420	4077	12876	1776	19149
分類正確率	81%	91%	91%	82%	

資料來源：本研究整理

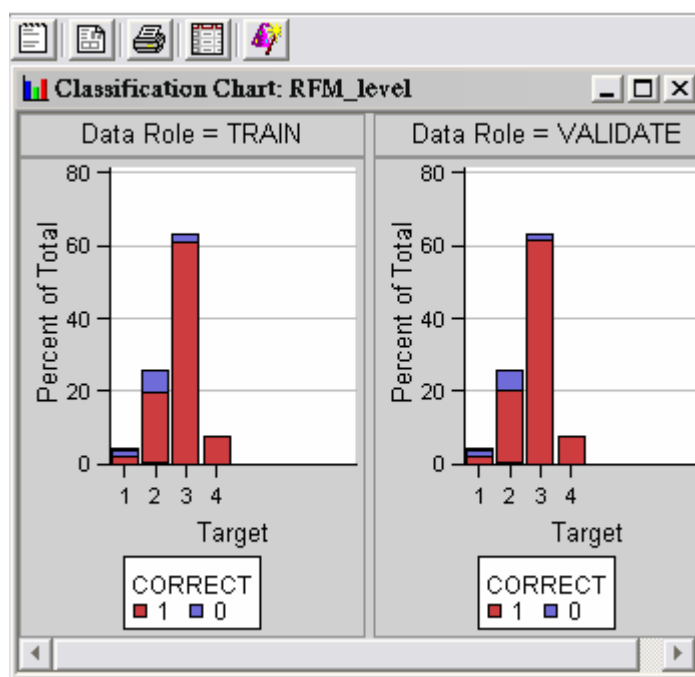


圖 4. 12 混淆矩陣示意圖

資料來源：本研究整理

在 EM SAS 決策樹演算法之 Output 報表中，同時亦陳列了變數的處理結果。軟體預設值乃以卡方檢定衡量名義型變數，對於連續型變數則採以 T 檢定，分別透過不同的統計方法，評判該變數在整體資料中重要與否。而在排除與 RFM 顧客價值分群相關之保單張數、投保金額與投保時間三項欄位後，投入決策樹演算法之剩餘的 20 個欄位如表 4.13 所示。透過 EM SAS 決策樹之 Output 報表可知，在該些變數之中，居於重要地位者包含主附約、生效月、繳費年期、首期保費、險別分類、特性分類、投保年齡、職業以及繳費別等九項欄位變數，如圖 4.13 所示。

表 4.13 欄位變數列表

項目	欄位名稱	欄位型態	項目	欄位名稱	欄位型態
1	保單號碼	數值	11	險種版數	數值
2	保單識別碼	數值	12	被保人代碼	數值
3	主附約	類別	13	要保人代號	數值
4	生效月	數值	14	投保年齡	數值
5	業務單位	類別	15	婚姻別	文字
6	繳費年期	數值	16	職業別	文字
7	首期保費	數值	17	繳費別	文字
8	險別分類	類別	18	性別	類別
9	特性分類	類別	19	吸煙別	類別
10	保單終止日	數值	20	地區別	類別

資料來源：本研究整理

The screenshot shows a SAS window titled "Results - Decision Tree (24)". The "Output" pane displays a table of Variable Importance. The table has columns for Observation (Obs), Variable Name (NAME), Label (LABEL), Importance, VImportance, and Ratio. The variables are ranked by importance, with 'Kind' being the most important variable.

Obs	NAME	LABEL	IMPORTANCE	VIMPORTANCE	RATIO	
71	Variable Importance					
72	1	Kind	Kind	1.00000	1.00000	1.00000
73	2	F_Money_Block	F_Money_Block	0.79303	0.77877	0.98201
74	3	Depart	Depart	0.75794	0.73553	0.97044
75	4	Age	Age	0.53291	0.49425	0.92746
76	5	Pay_Way	Pay_Way	0.32911	0.26489	0.80486
77	6	Main_Sub	Main_Sub	0.19369	0.18326	0.94616
78	7	Career	Career	0.13999	0.15998	1.14277
79	8	Month	Month	0.12210	0.14008	1.14726
80	9	Character	Character	0.10796	0.06577	0.60923

圖 4.13 EM SAS 決策樹 Output 報表之重要欄位介面圖

資料來源：本研究整理

4.3 關聯法則挖掘及策略建立

本研究之目的乃依顧客價值將其正確分類，並針對不同價值層級之顧客提出差異化行銷之建議。因此，在透過決策樹演算法驗證分群準確性後，接續將針對不同層級之顧客資料，進行關聯法則萃取，以為行銷策略之依循。

4.3.1 資料前處理

在利用決策樹演算法時，本研究得知較為重要之欄位囊括主附約、生效月、繳費年期、首期保費、險別分類、特性分類、投保年齡、職業以及繳費別等九項，然採以關聯法則之主要目的，乃針對每一價值層級之顧客做關聯法則分析，除不具意義或具有交互作用之欄位需予以剔除外，其他欄位對關聯法則仍存在貢獻之可能性。因此，本研究重新選取性別與婚姻別二項欄位進入分析步驟以為輔助，但對關聯法則之擷取仍以決策樹演算法所判定出之重要欄位，為法則主要採擷對象。

而在驗證分類正確率時所予以排除之保單張數、投保日期以及投保金額三個欄位變數，亦於關聯法則分析前重新選取加入。三個欄位先前之欄位定義方式主要目的乃為正確區分顧客層級所編排，在顧客層級近乎正確分類後則可將其加入，一併討論。然為確保該三項欄位定義不影響法則擷取，於加入關聯法則萃取前，將欄位重新定義與編碼，與篩選出之九項重要欄位以及後來加入之二項欄位共同分析。

完成欄位變數選取後，乃依 RFM 指標分數層級將 63,512 筆資料區分為 RFM Score Level 1~4 之四個資料集，如圖 4.14 所示，同時，轉換每一資料集之 Excel 檔案格式為 *.arff 格式，以利分析軟體 weka 之讀取。

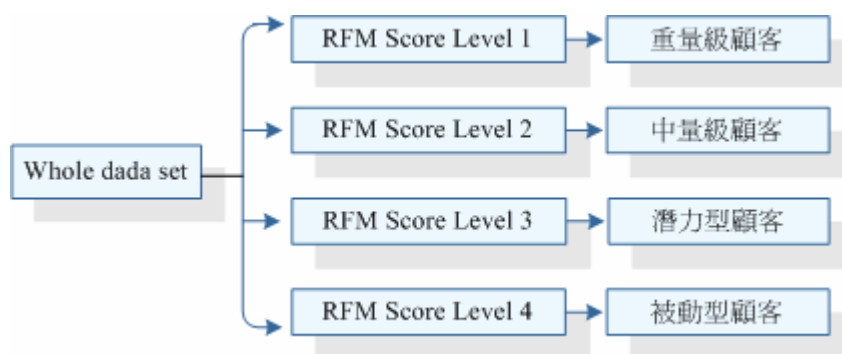


圖 4.14 資料區分圖

資料來源：本研究整理

4.3.2 資料後處理與法則分析

在資料採礦的過程中，能從資料庫中挖掘出許多知識以及法則，然而，其中必然同時涵蓋了具重覆性、直覺性或是無意義之關聯法則。因此，在眾多關聯法則中須找出具有意義者，以提供決策者參考依據。Klemettinen

【3】等學者曾指出，並非所有高信度及高支持度的法則都是有用的，因為有些法則的意義可能重覆或是理所當然，而有些沒有意義的法則是由不相關的屬性所構成，於這般情況下，在建立關聯法則之後，還須挑選出具有意義且可解釋的法則。

本研究關聯法則所使用的工具為紐西蘭 Waikato 大學所開發出來的 Waikato Environment for Knowledge Analysis 軟體，使用介面如圖 4.15 所示。以下將針對其關聯法則參數的設定做一說明。

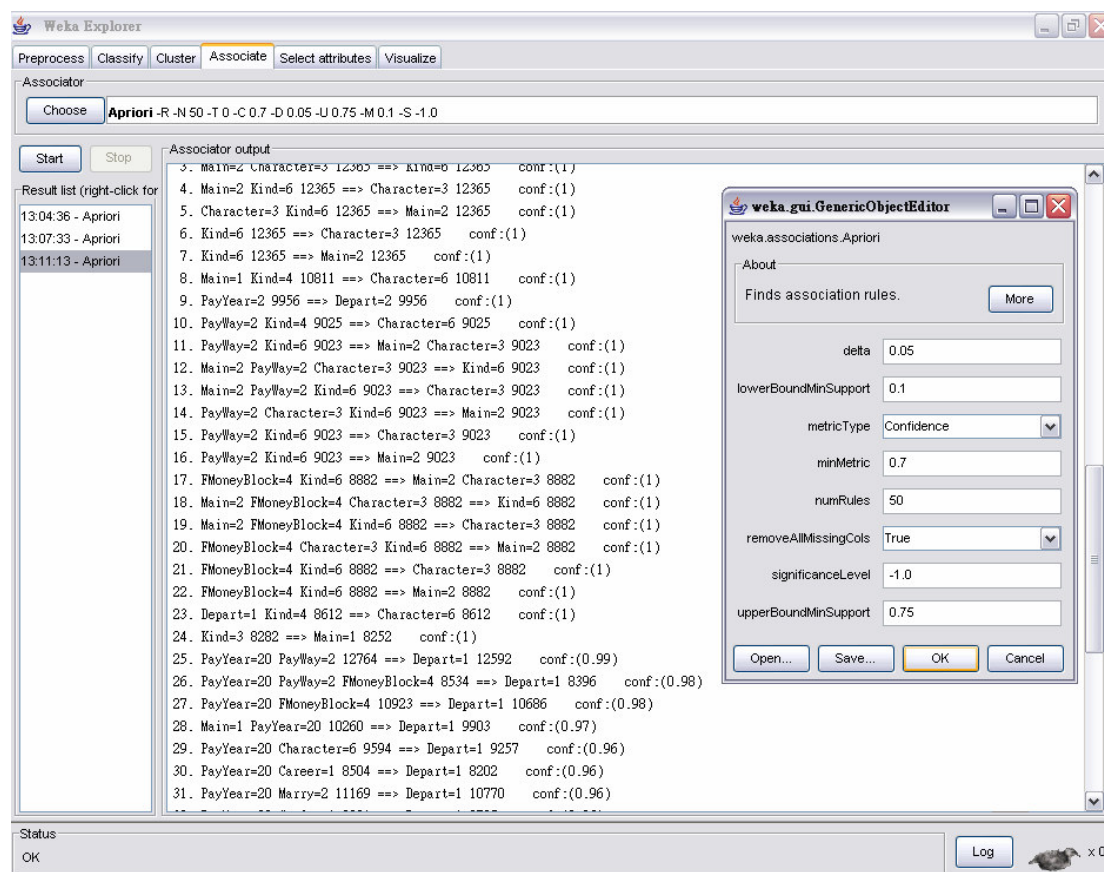


圖 4.15 Weka 關聯法則介面

資料來源：本研究整理

關聯法則探勘之參數設定以支持度門檻值(Min Support)與信賴度門檻值(Min Confidence)最為關鍵，此二參數之設定多依賴經驗或依使用者需求而得。在資料量較小時，最小支持度門檻值可設定較低，然最小支持度門檻值過低，將使得電腦運算負擔過重而無法執行，因此須依資料特性作調整；反之，最小支持度門檻值若設定較高，信賴度愈高，則表欄位變數關係愈益強烈。

軟體支持度門檻值之預設值為 1.0，而信賴度門檻值預設為 0.9，即表

報表所呈現之關聯規則乃介於支持度 1%~100%間，且法則信賴度至少高於 90%。然而，在此設定下所呈現之法則多為無意義之關連。因此，本研究認為支持度的範圍有向下調整的必要，經反覆測試，規範支持度介於在 1%~50%之間且信賴度高於 75%者，所呈現之法則較具意義。以下將擷取每一價值層級關鍵之關聯法則做一整理及分析，並針對該層級予以行銷策略建議。

1. 重量層級

表 4. 14 重量層級關聯法則表

1	Year=6 Fmblock=6 ==> Mblock=5 Kind=20 Main=1 sup:(0.06) conf:(0.99)
2	HowMany=3 Gender=0 Character=3 ==> Mblock=5 Kind=50 PayWay=2 sup:(0.05) conf:(0.92)
3	Age=3 Character=1 Main=1 ==> Fmblock=6 Kind=20 sup:(0.07) conf:(0.87)
4	Year=6 Marry=2 Kind=50 ==>Mblock=5 Character=3 Main=1 sup:(0.09) conf:(0.97)
5	HowMany=5 Age=3 Marry=1 ==> Main=1 134 sup:(0.05) conf:(0.98)
6	Character=1 Kind=20 ==>Fmblock=6 Main=1 sup:(0.07) conf:(0.88)
7	HowMany=2 Age=1 Mblock=5 Kind=50 ==> Marry=2 Character=3 sup:(0.07) conf:(0.89)
8	Kind=50 PayWay=2 HowMany=3 Gender=0 ==> Mblock=5 sup:(0.05) conf:(0.92)
9	Gender=0 Mblock=5 Career=2 ==> Main sup:(0.06) conf:(0.99)

(1)Year=6 Fmblock=6 → Mblock=5 Kind=20 Main=1 sup:(0.06) conf:(0.99)

於 6%的重量層級顧客間，若交易時間為 90 年且首期保費在新台幣 50 萬元以上者，其投保商品為主約性質之生死合險商品，並且保險金額大於新台幣一百萬存在 99%的可能性。

(2)HowMany=3 Gender=0 Character=3 → Mblock=5 Kind=50 PayWay=2 sup:(0.05) conf:(0.92)

在 5%的重量層級中，有 92%的保戶為女性，以健康型保單為主要投保項目，並以擁有 3 張保單，保費繳交方式多唯一次繳清。

(3)Age=3 Character=1 Main=1 → Fmblock=6 Kind=20 sup:(0.07) conf:(0.87)

重量層級顧客中有 7%的年齡介於四十至六十歲間，若其以主約型態投保投資型保單，採生死合險並且首期保費在五十萬以上者有 87%的可靠性。

由上述法則觀測可知，90 年投保且首期保費高於 NT\$500,000 元之重量級顧客，所投保項目為保額高於 NT\$1,000,000 元以上之生死合險；保單張數三張以上之健康型男性保戶，保額高於 NT\$1,000,000 元且採躉繳繳費方式；40~60 歲之理財型主約保戶，投保首期保費在 NT\$500,000 以上之生死合險；20~40 歲且擁有 2 張以上保額於 NT\$1,000,000 之保單保戶，投保健康型主約且已婚。綜觀而言，首期保費重量層級之關鍵利潤由 40~60 歲且近期投保之保戶掌握，無論男性或女性皆擁有多張保單，首期保費高於 NT\$500,000 元，且保險金額介於 NT\$1,000,000 至 NT\$10,000,000 元間，採躉繳一次將保費繳清之繳費方式，投保險種以理財型與健康型主約表現顯著，並採主約型態之生死合險作為商品選擇主要考量，而鮮少搭配附約。

就重量層級保戶而言，已屆高齡而擁有多張保單且經濟狀況表現優渥，然以躉繳之繳費方式購買健康型與理財型生死合險以為避稅。建議業者可派理財專員專為其作投資型保單規劃，抑或是設置理財餐會建立保單理財概念。進行商討與教育的過程間，提升派員對客製化保障項目刪改的權利，透過授權當下允諾之方式增加保戶購買之感性訴求因素，並選擇較完善之主約型態為其做設計或搭配。此外，對於重量層級保戶除著重被保人信函或訪問等情感聯繫外，亦可往被保人之直旁系血親橫縱兩向深植。以往我國業者在建立保單資料時，被保人或要保人多為資料庫輸入主軸，然由既有保戶所發展延伸之保單開發項目亦應多加註解，建議可以介紹人欄位設計保額累計項目發展優質保戶專案，期望達到以戶養戶之功效，而達優質保戶標準者，可予以連鎖商店消費優惠、停車優惠、場地租借優惠或他種實體優惠。

2. 中量層級

表 4.15 中量層級關聯法則表

1	Gender=1 Marry=2 Mblock=5 ==>FMblock=1 Character=3 Kind=50 sup:(0.10) conf:(0.86)
2	Mblock=4 Career=1 ==>Main=1 Kind=30 Character=0 sup:(0.06) conf:(0.85)
3	Month=12 Age=2==>Main=1Character=0 sup:(0.06) conf:(0.93)
4	Year=5 Age=1 Marry=2 Kind=50 ==> Character=3 Main=1 sup:(0.050) conf:(0.97)
5	Gender=0 Mblock=4 Marry=2 Kind=30 ==> Main=1 Character=0 sup:(0.05) conf:(0.98)
6	Age=2 Character=0 Kind=30 ==> Main=1 Mblock=4 sup:(0.08) conf:(0.93)
7	FMblock=0 Kind=50 Main=1 ==> Marry=2 Character=3 sup:(0.07) conf:(0.86)

8	FMblock=1 Main=1 Character=3 ==> Gender=1 Marry=2 Mblock=5 sup:(0.10) conf:(0.86)
9	Year=5 FMblock=1 Kind=50 ==> Mblock=5 Character=3 sup:(0.07) conf:(0.91)
10	Career=6 Marry=2 ==> Main=1 901 sup:(0.06) conf:(0.85)

(1)Gender=1 Marry=2 Mblock=5 →FMblock=1 Character=3 Kind=50
Main=1 sup:(0.10) conf:(0.86)

中量層級顧客中有 10%的已婚女性保戶投保金額大於新台幣壹佰萬，首期保費項目有八成六之顧客介於一千元至五千元間，且以健康型之主約約種為投保標。

(2)Mblock=4 Career=1 →Main=1 Kind=30 Character=0 sup:(0.06)
conf:(0.85)

有 6%的中量層級保戶職業為文教，投保金額介於新台幣七十五至一佰萬間，而該保戶投保險種乃主約型態壽險型之死亡險有 85%的信賴度。

(3)Month=12 Age=2→Main=1Character=0 sup:(0.06) conf:(0.93)

有 6%之中量層級保戶其年齡介於 20 與 40 歲之間，且於 12 月份完成投保，其中存在 93%之保戶投保主約型態壽險型商品。

透過法則可以發現，已婚女性保戶而保額大於 NT\$100,000 元者，投保首期保費介於 NT\$1,000~5,000 間之健康型約種；保戶為保額介於 NT\$75,000 ~NT\$100,000 之文教業者，投保死亡型壽險主約；投保於 12 月份之 20~40 歲保戶，投保保額為 NT\$75,000 ~NT\$100,000 之壽險主約；首期保費在 NT\$5,000 元以下之健康型主約，已婚男性保戶投保保額介於 NT\$75,000 ~NT\$100,000 之間。總括來說，中量層級顧客雖在首期保費上偏低，然其投保金額之選取仍高，首期保費與投保金額之額度分別為 NT\$5,000 元以下及 NT\$75,000 ~NT\$100,000 之間，主乃因保戶多從事文教產業，儘管每月收入並非十分可觀，但薪資足以應付生活開銷與儲蓄投資，再加上收入平穩且福利優渥等因素，使得此層級之顧客十二月份成交量大且投保期間長，投保險種方面則以較完整之主約型態做選擇，約種著重於壽險型生死合險與健康型生死合險。

服務業者偏向死亡型健康型商品主約，文教業者在收入與環境方面具有安定因素，但穩定的特質讓其選擇小額繳費但總保額額度仍高之長期保單經營，並將險種之選擇界定在壽險型。由於職業賦予此層級保戶單純但

具防衛的特性，建議業者可針對文教業保戶做風險性商品開發，開發前期依據每一保戶所評估之風險接受度編撰投資理財概念引導手則，透過書面之方式削弱防衛及抗拒心理，且同時能提供或教育其投資型與利率變動型商品相關知識，後期則以派員或講師定期輔導解說，透過縮減資訊不對稱之恐懼感來刺激需求；此外，可設計休養生息之回饋度假專案，提供於限定日期內締約之中量層級保戶免費豪宅或度假小木屋數日體驗之回饋。

3.潛力層級

表 4.16 潛力層級關聯法則表

1	Character=3 Kind=52 Main=2 ==> Mblock=3 FMblock=2 sup:(0.08) conf:(0.76)
2	HowMany=1 Marry=1 ==>Kind=30 Character=0 Main=1 sup:(0.22) conf:(0.86)
3	FMblock=2 Character=0 ==>Kind=30 PayWay=4 Main=1 sup:(0.09) conf:(0.85)
4	Mblock=3 Kind=52 ==> Marry=2 Character=3 sup:(0.07) conf:(0.69)
5	Age=2 Mblock=2 Character=0 ==> HowMany=1 Main=1 sup:(0.07) conf:(0.72)
6	FMblock=0 Character=3 Main=2 ==> Kind=52 sup:(0.08) conf:(0.66)
7	Year=5 Kind=20 ==>HowMany=1 Character=0 Main=1 sup:(0.07) conf:(0.76)
8	PayYear=2 Character=0 ==> Main=1 Kind=30 sup:(0.09) conf:(0.85)

(1)Character=3 Kind=52 Main=2→Mblock=3 FMblock=2 sup:(0.08) conf:(0.76)

潛力型層級保戶有 8%投保附約型態第二版健康型商品，而該些保戶有七成六之投保金額涵蓋新台幣五十萬與七十五萬之範圍，且首期保費介於新台幣五千元至一萬元間。

(2)HowMany=1 Marry=1 →Kind=30 Character=0 Main=1 sup:(0.22) conf:(0.86)

有 22%之潛力型層級保戶未婚，若該保戶擁有一張保單，則其投保主約型態之壽險型死亡險的可靠度達 86%。

(3)FMblock=2 Character=2 PayWay=2 →Kind=30 PayYear=4 Main=1 sup:(0.09) conf:(0.85)

在潛力型層級保戶間，有 9%的保戶投保壽險型商品，如果該些顧客採季繳之繳費方式繳交保費，則其投保主約型態之死亡險有 85%的信賴度。

觀察上述法則可知，此層級健康型第二版型附約成交約種落於保額

NT\$500,000~NT\$750,000 間，首期保費於 NT\$5,000~NT\$10,000 之內；未婚保戶持有單張保單，保單險種為死亡型壽險主約；首期保費 NT\$5,000~NT\$10,000 之壽險型約種，為季繳之死亡型主約；投保保額為 NT\$100,000~NT\$250,000 壽險之 20~40 歲未婚保戶，僅擁有單張保單且為主約；首期保費介於 NT\$0~NT\$1,000 元之健康型附約，為第二版型健康型約種；繳費期間六至十年之壽險約種，為健康型約種之主約。綜上所述，潛力層級保戶分布於 20~40 歲之間，投保約種區分為主約型態壽險型生死合險與主約型態壽險型死亡險二區塊，然兩者皆近期投保並搭以附約型態第二版健康型商品，且首期保費多較中量層級為高，約當 NT\$5,000~NT\$10,000 元，保額上則反低於中量層級而介於 NT\$250,000~NT\$750,000 元間，投保年期六到十年不等，繳費時間以季為單位。綜觀而言，中量層級近期投保之保戶著重壽險型之死亡險與生死合險，並兼備最新之第二版健康型附約，而非直接選擇主約型態之健康型商品，可知潛力層級保戶具備一定程度的保險知識，並且能跟上業者產品開發之概念以及產品更新的速度；保額項目反應此層級保戶投保能力較前兩層級為弱，但由首期保費與投保險種顯現其雖投保能力較弱，但對保險概念及商品認識之更新速度快速，目前雖僅擁有單張保單，但仍突顯此層級既有顧客之開發可期。

建議業者依循其更新速度快速之特性，間或透過網路、信函或派員深植，以強化其對新開發之概念或商品的想法及認識，由興趣引發需求。另一方面，運用契約撤銷權與聯名卡的使用來增加產品附加價值，前者以限定毀約期限來誘發其對保險商品的購買嘗試，而後者則利用累積保單或保額來計算續約或轉約之折扣扣抵，抑可與商家策略聯盟來突顯聯名卡之附加價值。亦可針對此層級保戶層級保戶僅願投資附約型態商品之特性，開發虛擬主約專案，由於附約需依賴於主約之下，然又須依循此特性，故開發一虛擬主約並授予派員數種附約之單獨締結權，讓附約簽約型態得以成立。此外，今年 6 月政府甫通過之勞退新制，除推波助瀾地增加保戶對保險的知識建立外，亦將使得年金型態商品以及退休金帳戶建立成為保戶關注的重點。因此，業者可利用此良好時機協助保戶重新檢視保單，於過程中作年金型商品概念教育與開發，並建立情感及信任依賴。

4.被動層級

表 4.17 被動層級關聯法則表

1	Year=3 Mblock=0 ==>Character=3 Kind=51 Main=2 sup:(0.17) conf:(0.89)
2	Year=4 Fblock=0 ==>Mblock=0 Marry=2 Character=3 Kind=51 sup:(0.21) conf:(0.88)
3	HowMany=1 Gender=0 Kind=51 ==> Mblock=0 Character=3 Main=2 sup:(0.20) conf:(0.96)
4	Marry=2 HowMany=1 Kind=51 ==> Fblock=0 Mblock=0 sup:(0.29) conf:(0.81)
5	Year=3 Main=2 ==>Mblock=0 HowMany=1 sup:(0.17) conf:(0.9)
6	Fblock=0 ==> HowMany=1 Character=3 Main=2 759 sup:(0.18) conf:(0.85)

(1)Year=3 Mblock=0 →Character=3 Kind=51 Main=2 sup:(0.17) conf:(0.89)

有 17%之被動型層級保戶於 87 年投保，若投保金額在新台幣十萬元以下，則其投保項目為附約型態第一版健康型商品之可靠度達 89%。

(2)Year=4 Fblock=1 →Mblock=0 Marry=2 Character=3 Kind=51 sup:(0.21) conf:(0.88)

在被動型層級之保戶中，有 21%的保戶於 88 年投保，若首期保費於新台幣壹仟元以下，則其間有八成八的保戶已婚，且投保第一版健康型附約，投保金額乃於新台幣十萬元以下。

(3)HowMany=1 Gender=0 Kind=51 → Mblock=0 Character=3 Main=2 sup:(0.20) conf:(0.96)

有 20%被動型層級之保戶為女性且擁有一張保單，若其投保型種為第一版健康型商品，則其中有九成六乃投保附約型健康型商品，且商品投保金額在新台幣十萬元以下。

由上列法則可知，於 87 年投保且保額低於 NT\$100,000 者，投保約種為健康型態之舊版健康險種附約；88 年投保而首期保費低於 NT\$1,000 者，保戶已婚且投保保額低於 NT\$100,000 元之健康型舊版附約；保戶已婚且擁有單張健康型舊版附約，約種保額與首期保費分別於 NT\$100,000 與 NT\$1,000 以下。綜觀而言，被動層級保戶以女性為主，於民國 87 年前後投保，故而投保險種多為舊版商品，其中以附約型態之舊版健康型商品表現最為顯著，保額在 NT\$100,000 元以下，且首期保費低於 NT\$100,000 元，無論是保額或首期保費額度皆不高，從投保期間起算至今僅擁有單張保單，且無加保意願。

被動型保戶多為對保險認識不深或經濟上較無力投保群集，儘管保留該部分顧客並無法大幅提升公司利潤，然若業者資源允許，亦可對此層級保戶作開發。此層級保戶之保單特性散落，維其欲以投保首期保費與保額皆低之附約險種型態，而其缺乏新版商品之狀況可知其缺乏保險概念追蹤意圖，因此，建議業者可針對此層級顧客做訪談，了解此層級保戶加保意願低落之原因並以此擬定改善方針，或許可針對經濟匱乏之層級，設計保險金額與首期保費較低之人道關懷保障項目，從長期的角度來經營。

第五章 結論與建議

5.1 結論

本研究乃以保險產業之人身保險為研究資料來源，欲結合 SAS Enterprise Miner 與 weka 兩分析軟體建立一資料採礦架構，期望業者可透過此程序區隔出值得深植之顧客層級，並充分利用客戶與消費行為之紀錄從事法則探勘，以在企業資源有限的情形下，精準地運用時間、成本以及策略，進而提高保戶增購保單之比例。

針對提高保戶增購保單之目的，本研究分為兩個步驟來進行。第一部分乃就既有保戶資料做描述且區隔每一顧客價值層級，並依據 RFM 理論測試欄位分群定義正確與否。第二部份則利用關聯法則演算法，分別對已達錯誤率最低之每一顧客價值層級進行關聯法則萃取，並擬以二次行銷策略。

本研究之研究流程乃於取得各案金融控股公司之人身保險資料後，先對原始資料做格式轉換以利軟體分析讀取，而後進行資料清理，除去資料夾帶之雜訊與干擾。在資料完成轉換與清理之後，透過敘述性統計概括陳述資料屬性，以輔助下一步驟 RFM 價值分群之欄位定義。當 RFM 價值分群之定義終了，則透過決策樹 C4.5 演算法針對該定義方式佐以驗證，若分類錯誤率已達最低，則接續以關聯法則之 Apriori 演算法，找出每一層級顧客潛藏之關聯法則，並針對不同價值之客群擬定策略。

本研究應用關聯法則於四個顧客價值層級之研究結果顯示，第一、四層級法則多等同於直觀判斷之特徵，層級高低與保額以及首期保費兩欄位變數成正比，惟理出第一層級 40~60 歲已婚保戶投保險種為理財型險種，並透過躉繳的方式節稅；第二、三層級於保戶特徵及保單特徵上則皆具表現，且兩層級所分別表現之保戶與保單特徵，皆於另一特徵已成立之前提下表現，意即保戶特徵表現在保單特徵已成立之情況下，且保單特徵抑表現於保戶特徵已成立之情形下。在保戶特徵方面，中量層級 20~40 歲之未婚文教業者投保死亡險之壽險主約，而 20~40 歲未婚服務業者投保死亡險之健康主約，並鮮少搭配附約。潛力層級已婚保戶投保之主約法則紛亂，惟其所投保之附約多為新版附約。於保單特徵方面，中量與潛力層級在首

期保費與保額法則上呈現相反趨勢，中量層級顧客首期保費低但保額表現高，而潛力層級則於首期保費高而保額表現低。

業者對於首期保費之建置，實乃於不損及顧客保障權利的前題下所發展之三贏策略。對保戶而言，首期保費高則保戶享有之利率回饋則高；就業者而言，首期保費高則業者擁有之現金流量高；而由業務員角度觀之，首期保費高則依其按比例發放之獎勵佣金則高。然而，業務員由於佣金所致，容易對於首期保費較高之第三層級顧客予以較高重視，但就長期來評判，較具價值且應較用心經營之顧客，實乃首期保費較低但保額高之第二層級顧客，建議業者將資源正確配置，則保單開發將更具效率。然若業者能使第二層級之中量層級顧客，導引為首期保費亦高之特徵模式，營運績效則必更佳。

保單開發可就舊有保戶潛在需求與新保戶需求兩方面做努力，對於二次行銷與新保戶開發之策略擬定，本研究乃依循各層級保戶特徵屬性與保單屬性所萃取之關聯法則，各層級法則之策略建議綜述如下。

1.對於重量層級之保單增購策略

- (1) 理財餐會：就重量層級保戶而言，多半已屆高齡且擁有多張保單，故策略以情感聯繫為主，可定期拜訪或舉辦餐會，除表達對客戶年節之祝福外，更能同時告知顧客新的商品或活動資訊，使其感受重視而提升業者形象，此外，提升派員對客製化保障項目刪改的權利，透過授權當下允諾之方式增加保戶購買之感性訴求因素，並選擇較完善之主約型態為其做設計或搭配。
- (2) 專業投資理財之保單規劃：此層級保戶經濟狀況表現優渥，因此可針對投資型商品、理財型商品與利變型商品做開發，建議可訓練專業理財專員專為其建立投資型及理財型保單規劃，並間或給予其客製化之保障項目。
- (3) 發展優質保戶專案：可依循保戶之信任與肯定，往被保人之直、旁系血親橫縱兩向深植，建議可累計保額介紹人累計方式發展專案，達優質保戶標準者，予以連鎖商店消費優惠、停車優惠、場地租借優惠或他種實體優惠。。

2. 對於中量層級之保單增購策略

- (1)設計投資理財概念手則：中量級保戶可針對環境安定且收入穩定之文教業者做深耕，建議業者前期依據每一保戶所評估之風險接受度編撰投資理財概念引導手則，透過書面之方式削弱防衛及抗拒心理，且同時能提供或教育其投資型與利率變動型商品相關知識，後期則以派員或講師定期輔導解說，透過縮減資訊不對稱之恐懼感來刺激需求。
- (2)休養生息回饋度假專案：提供於限定日期內締約之中量層級保戶免費豪宅或度假小木屋數日體驗之回饋。

3. 對於潛在層級之保單增購策略

- (1) 虛擬主約專案：針對此層級保戶層級保戶僅願投資附約型態商品之特性，然附約需依賴於主約之下，故依循此特性開發一虛擬主約並授予派員數種附約之單獨締結權，讓附約簽約型態得以成立。
- (2)契約撤銷權：以限定毀約期限來誘發其對保險商品的購買嘗試
- (3)聯名卡：利用累積保單或保額來計算續約或轉約之折扣扣抵，抑可與商家策略聯盟來突顯聯名卡之附加價值。
- (4)商品更新快報：經由本研究發現，潛力層級顧客能跟上業者保險商品汰換速度，可根據此特性深植，建議業者透過網路、信函或派員間或給予已新開或欲開發之保險概念以及商品知識，使該些保戶能更方便快速地立即掌握有興趣之商品資訊並深入了解，由興趣引發需求。而由於此一區塊之保戶對保險商品敏感度較高，亦可邀請其參與商品開發之腦力激盪或是多加給予業者各方建議。
- (4)協助保單檢視：業者可搭配今年六月政府甫通過之勞退新制，於協助保戶作保單檢視之過程中，達其情感加溫且教育其勞退新制，並予以退休金及年金型商品概念之教育與開發，為日後投保奠定基礎。

4. 對於被動層級之保單增購策略

- (1)人道關懷商品：被動型保戶多為對保險認識不深或經濟上較無力投保之群集，儘管保留該部分顧客並無法大幅提升公司利潤，然若業者資源允許，建議業者可針對此層級設計保險金額與首期保費較低之人道關懷保障項目，從長期的角度來經營。

5.2 未來研究與建議

本研究利用資料採礦技術欲提高人身保險保戶增購保單之比例，然於整個資料匯集、轉檔、定義、驗證、執行之工作循環中仍存在許多可以改進之處，以下列出幾點建議以作為後續學術界以及業界相關研究者的參考：

1. 本研究將一張保單視為一個資料分析單位，即將一張保單視為一投保險種、一個保戶等，且未將險種之時間序列(定期保單與非定期保單)納入考量，如此將保單狀況簡化之情況也許會影響資料表現，可於未來研究中做修正。
2. 本研究之欄位定義方式依據決策樹混淆矩陣驗證結果決定，然並未考量欄位組距不等之情況，未來研究可對此多加著墨。
3. 本研究由於資料取得不易，在顧客價值指標上以取得之資料為基礎作分析，然若能加入「被保人所得」、「被保人貸款狀況」、「被保人已育子女數」等三個面向，將使整個顧客價值層級之分群架構更為完整，對於各層級所萃取之法則貢獻度亦能提升。
4. 本研究針對保戶與保單資料分析之結果提出策略建議，然未能探討策略實際施行之狀況與保戶反應，未來可針對後續行銷反應的狀況做一分析研究。
5. 本研究主乃運用保戶人口基本變數與保戶投保行為以為析變數，進行資料採礦的工作，然亦有學者認為心理層面與生活型態更必須納入考量，才能真正準確掌握顧客價值關鍵，因此，如何在實務上取得顧客心理層面與生活型態的分析資料，可於未來進行探討。

參考文獻

中文部分

- 【1】 大前研一，策略家的智慧，長河出版社，1983年10月。
- 【2】 比爾.蓋茲，數位神經系統，樂為良譯，商業周刊出版，1999年。
- 【3】 安迅資訊公司，”整合企業經營策略與顧客關係管理”，電子化企業經理人報告，1999年11月，pp20~25。
- 【4】 江茂榮，自組織映射圖類神經網路建構壽險業客戶投保關係管理模式，碩士論文，國立高雄第一科技大學風險管理與保險所，2002。
- 【5】 李宏谷，應用自組織映射圖網路於顧客關係管理之分群研究，碩士論文，私立中華大學科技管理研究所，2002。
- 【6】 李昇暉，「以資料採礦深化顧客關係管理」，電子化企業經理人報告，第7期，遠擎管理顧問公司，2000年3月，pp.37~pp.42。
- 【7】 李家旭，「應用資料採礦技術於保險公司附加保單之增售」，碩士論文，國立政治大學統計研究所，2002年。
- 【8】 李珮榕，以資料採礦技術分析大台北地區保單貸款，碩士論文。國立政治大學/統計研究所，2002。
- 【9】 沈聰益，「人格五因素模式預測保險業務員銷售績效的效度—NEO-PI-R 量表之跨文化檢驗與人際特質架構之實證探討」，博士論文，國立交通大學經營管理研究所，2002年。
- 【10】 林育宏，建構可供資料採擷之壽險客戶關係管理模式，碩士論文，國立高雄第一科技大學風險管理與保險所，2002。
- 【11】 林明宏，壽險保單早期失效之預測—類神經網路之應用，碩士論文，私立朝陽科技大學保險金融管理系，2002。
- 【12】 吳旭智、賴淑貞譯，Data Mining 資料採礦理論與實務顧客關係管理的技巧與科學，Michael J.A. Berry & Gordon S. Linoff，維科圖書有限公司，2001，台北
- 【13】 洪英瑞，「網際網路之顧客服務策略-以運用網路下單之證券商為例」，碩士論文，台灣大學工業工程學，台北，2001年。
- 【14】 涂靜儀，結合自組織映射圖類神經網路與基因演算法建構壽險業顧客關係管理之知識採擷模式，碩士論文，高雄第一科技大學風險管理與保險所碩士論文，2001年。
- 【15】 張家鳳，「利用資料挖掘技術建構保險業之差異化行銷模型」，碩士論文，私立世

新大學資訊管理學系，2004 年。

- 【16】張愷芬，「資料採礦－保險業商品市場分析」，碩士論文，私立淡江資訊工程學系，2003 年。
- 【17】彭文正譯，資料採礦－顧客關係管理暨電子行銷之應用，Berry and Linoff，數博網股份有限公司，1998 年。
- 【18】陳佳楨，「資料探勘應用於就診行為與醫師排班之研究-以埔里基督教醫院為例」，碩士論文，國立暨南國際大學資訊管理研究所，2003 年。
- 【19】陳佳鈴，「應用資料探勘於客戶關係管理之分群研究-以壽險業為例」，碩士論文，私立中華大學科技管理研究所，2004 年。
- 【20】陳曉開，電子化企業經理人報告，第三期。
- 【21】陳麒文，健康休閒俱樂部顧客流失分析模式之研究，碩士論文，私立輔仁大學體育研究所，2002。
- 【22】陸承志，”如何運用網站經營顧客關係”，DATA 數位行銷，2，2001，71-72。
- 【23】黃品豪，「網路下單對投資行為與投資績效之影響」，碩士論文，國立中央大學資訊管理學系研究所，2002 年。
- 【24】詹凱傑，「應用資料採礦技術於顧客流失分析之研究－以國產汽車維修業為例」，碩士論文，私立東海大學工業工程與經營資訊學系，2004 年。
- 【25】壽險商業同業公會、台灣經濟研究院產經資料庫整理，2005 年 9 月。
- 【26】潘筱瑜譯，簡明行銷聖經，Joseph H. Boyett & Jimmie T. Boyett，商周出版，2004 年。
- 【27】廖峻毅，「兼顧顧客價值與企業利潤之服務策略」，碩士論文，靜宜大學管理科學研究所，台中，1996 年。
- 【28】廖雅郁，「應用資料探採於我國西藥行銷之研究」，碩士論文，國立交通大學經營管理研究所，新竹，2003 年。
- 【29】趙民德，關於數據採礦中取樣概念的一些探討，成功大學資料採礦研討會，2001 年。
- 【30】劉玉萍，”整理 NCR 公司助理副總裁 Ronald S. Swift 於遠擎智慧企業部舉辦 2000 「一對一行銷」研討會中之專題演講：運用「一對一行銷」執行顧客關係管理已提升企業利潤”，電子企業：經理人報告，第 11 期，7 月，2000 年，頁 7-10。
- 【31】劉典嚴，「仰賴 TQM 解開顧客價值的三角習題」，品質管制月刊，第 37 期，2001 年 2 月，pp.57~pp.60。
- 【32】樓玉玲，「以資料發掘分析政大通識課程」，碩士論文，國立政治大學資訊管理學系，台北，1998 年。
- 【33】賴信良，「資料挖掘在教育上的應用－以國小學童『體適能測驗』為例」，碩士論

文，國立台北師範學院數理教育研究所，2002 年。

- 【34】賴耐志，「應用資料採礦於市場區隔分析」，碩士論文，國立台北師範大學商業自動化與管理研究所，2002 年。
- 【35】羅國輝，壽險業務人員運用顧客關係管理系統（CRM）滿意度與績效之研究，碩士論文，私立逢甲大學保險所，2002。
- 【36】「1999 年度台灣『顧客關係管理』運用現狀調查報告」，電子商務導航，第二卷，第 13 期。

英文部分

- 【1】 Agrawal, R. & R. Srikant, "Fast Algorithms for Mining Association Rules", *Proc. Of the 20th Int'l Conference on Very Large Databases*, Santiago, Chile, Sep. 1994.
- 【2】 Agrawal, R., T. Imielinski, and Swami, A., "Mining association Rules between Sets of Items in Large Databases", *ACM SIGMOD*, Washington, DC, U.S., May. 1993.
- 【3】 Alex, S. "Winning CRM strategies". *ABA Banking Journal*, October, 1999, pp 54-66.
- 【4】 Bauer, C. L., "A Direct Mail Customer Purchase Model", *Journal of Direct Marketing*, 2nd ed., 1988, pp.16-pp.24.
- 【5】 Berry, M. J. A., & G. S. Linoff, *Data Mining Techniques: for Marketing, Sales, and Customer Support.*, New York: Wiley Computer, 1997.
- 【6】 Chien-Hsing Wu, Shu-Chen Kao, Yann-Yean Su, Chuan-Chun Wu, "Targeting customers via discovery knowledge for the insurance industry", *Expert Systems With Applications*, vol.29, 2005, pp.291-299.
- 【7】 Engel, J. F., R. D. Blackwell, & Miniard, P. W., *Consumer Behavior*, 8th ed., no.237, *Dryden Press Fort Worth, Texas*, 1995.
- 【8】 Fu, Y, "Data mining task, technique and applications," *IEEE POTENTIALS*, October/November 1997.
- 【9】 Gale, B. T., *Managing Customer Value*, Free Press, New York, 1983.
- 【10】 H. Curt, "The Devil's in the Detail: Techniques, Tools, and Application for Database mining and Knowledge Discovery-Part 1", *Intelligent Software Strategies*, vol. 6, no.9, 1995, pp.1-pp.15.
- 【11】 Han, J., and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2001.
- 【12】 Hughes, M. Arthur, *Strategic Database Marketing*, Chicago: Probus Publishing, 1994.
- 【13】 IBM, *IBM Intelligent Miner For Data*, 1998.
- 【14】 Inmon, W.H., *Building the data warehouse*, John Wiley & Sons, New York, 2nd edition, 1996.
- 【15】 J. Miglautsch, Thoughts on RFM scoring, *Journal of Database Marketing* 8(1), 2000, pp. 67-72.
- 【16】 Jackie Kandell, "CRM, ERM, one-to-one Decoding Relationship Management Theory and technology", *Trusts & Estates*, April 2000, pp.49-53.
- 【17】 Kalakota, Ravi and Marcia Robinson, *e-Business : Roadmap for success*, 1999, pp.177.
- 【18】 Kahan, Ron, "Using Database Marketing Techniques to Enhance Your One-to-One Marketing Initiatives", *Journal of Consumer Marketing*, Vol 15(5), 1998, pp.491-493.

- 【19】 Kleissner, C., “Data Mining for the Enterprise.” Thirty-First Annual Hawaii International Conference on System Sciences, 1998, pp.295-pp.304.
- 【20】 Kotler, P., S. H. Ang, Leong, S. M. & C. T. Tan , *Marketing Management: An Asian Perspective*, Prentice-Hall, Singapore, 1999
- 【21】 Magson, N., Database Workshop : Determining and Measuring Customer Value, *Journal of Database Marketing*, Vol. 6, No. 1, 1998, pp.24-pp.23.
- 【22】 Megiddo, N. and R. Strkan, “Discovering Predictive Association Rules,” IBM Almaden Research Center, KDD-98 Poster Sessions, 1998.
- 【23】 Naumann Earl, *Creating Customer Value*, 1st ed., Thomson Executive Press Inc, U.S., 1995.
- 【24】 Paul Ratcliff, “Managing Deeper Relationship”, *Mortgage Banking*, March 2000, pp. 94-100.
- 【25】 Peppers, D., & Rogers, M. *The Future of One to One: Building Business Relationships One Customer at Time*.1993, New York: Currency Doubleday.
- 【26】 Pieter, Z. Dolf, *Data Mining*, Addison Wesley, 1996.
- 【27】 R. Kahan, Using database marketing techniques to enhance your one-to-onte marketing initiatives, *Journal of Consumer Marketing* 15(5), 1998, pp.491-493.
- 【28】 Solomon, H., *Battle Waged Over State of CRM*, Computing Canada, 2000.
- 【29】 Sung, Ho Ha and Chan Park, Sang, "Application of data mining tools to hotel data mart on the Inteanet for database markrting," *Expert Systems With Application*, 15th ed., 1998, pp.1-pp.31.
- 【30】 Suh, E.H., K.C. Noh, and Suh, C.K.,”Customer list segmentation using the combined response model.” *Expert Systems With Application*, Vol. 17, 1999, pp.89-97.
- 【31】 Tzu-Tsung Wong, Kuo-Lung Tseng, “ Mining negative contrast setts from data with discrete attributes.”, *Expert Systems With Applications*, vol. 29, 2005, pp. 401-407.
- 【32】 Wang, Y.S. & H. Shao,” Data warehouse technology in process industry. ,” *The 3rd World Congress on Intelligent Control and Automation*, 2000, Vol.2037-2041, No.3.
- 【33】 Witten, I.H. & E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*.Morgan Kaufmann Publishers, San Francisco, 2000.
- 【34】 Woodruff, Robert B., "Customer Value: The Next Source for Competitive Advantage, " *Journal of the Academy of Marketing Science*, 1997, Vol. 25, No. 2, pp.139-pp.153.
- 【35】 Yeo, A. C., Smith, K. A., Willis, R. J. & Brooks, M. (2002), “A mathematical programming approach to optimize insurance premium pricing with a data mining framework”, *The Journal of the Operational Research Society*, 53(11), pp. 1197-1203
- 【36】 <http://www.lia-roc.org.tw/index03/statisc.htm>