



嶺東科技大學
LING TUNG UNIVERSITY

資訊管理系

顧客群分析-以荷蘭銀行營銷為例

指導教授：張顯榮 教授

組員名單：林奕宏 A28C179

薛嘉琪 A28C185

中華民國一〇六年五月

嶺東科技大學

資訊管理系

顧客群分析－荷蘭銀行營銷為例

中華民國一〇六年五月



嶺東科技大學
LING TUNG UNIVERSITY

資訊管理系專題口試委員審定書

顧客群分析-以荷蘭銀行營銷為例

指導教授：_____張顯榮_____教授

組員名單：_____林奕宏_____ A28C179

_____薛嘉琪_____ A28C185

指導教授：_____

口試委員：_____

中華民國 106 年 5 月 11 日

謝 誌

本專題報告得以順利完成，首先要感謝恩師張顯榮老師細心引導我們，耐心的協助我們，克服研究過程中所面臨的困難，給予我們最大的協助，使本專題得以順利完成。

研究報告口試期間，感謝林孟源老師、黃國華老師不辭辛勞細心審閱，不僅給予我們指導，並且提供寶貴的建議，使我們的專題內容以更臻完善，在此由衷的感謝。

最後，感謝系上諸位老師在各學科領域的熱心指導，增進商業管理知識範疇，在此一併致上最高謝意。

謹誌

中華民國106年5月於嶺東

摘 要

近年來，隨著產業日趨成熟，市場趨於飽和時，策略對企業就愈重要，策略的選擇更是關鍵，所面臨的競爭也日益激烈，顧客可能在彈指之間，投向競爭者的懷抱；於是，顧客關係管理彷彿成了企業前進的唯一標竿，受到當前企業界與資訊科技業界所熱烈討論，相關的解決方案也紛紛出爐。經過歷年累積的顧客數據，透過正確的分析，可以提供企業在制定策略時，作為重要的參考資訊，讓企業可以針對不同的顧客群使用不同的行銷策略。本研究專題從UCI(University of California, Irvine)機器學習儲存庫取得荷蘭銀行營銷數據集資料，使用粗集理論方法作為資料分析工具，從資料集中萃取出核心屬性集，建立找出目標顧客群的預測規則，提供管理者簡單易用的規則，以利制定有效的營銷宣傳策略。

關鍵字：機器學習、目標顧客、粗集理論、行銷策略

目錄

摘要	1
目錄	2
圖目錄	3
表目錄	4
第一章 緒論.....	6
第一節 研究動機	7
第二節 研究目的	8
第二章 文獻探討.....	9
第一節 粗集理論	9
第二節 目標顧客	17
第三節 策略.....	20
第三章 研究方法.....	22
第一節 研究模型	22
第二節 資料來源	23
第三節 演算步驟	23
第四章 實證分析與研究結果.....	33
第一節 實證演算與預測結果.....	34
第二節 研究結果	38
第五章 結論.....	42
參考文獻	45

圖目錄

照片 1 上、下近似值的示意圖.....	13
照片 2 研究流程圖.....	22
照片 3 荷蘭銀行營銷預測之實證演算流程圖.....	35

表目錄

表格 1 國內學者粗集理論之相關研究	11
表格 2 資訊系統	16
表格 3 可辨識矩陣	16
表格 4 統整學者策略定義	20
表格 5 本研究荷蘭銀行營銷之研究變數彙總表	24
表格 6 部分觀察值之原始資料	26
表格 7 部分觀察值之原始資料	26
表格 8 粗集理論應用軟體所產生之篩選屬性集 (第一組) ...	27
表格 9 以粗集理論應用軟體所產生之篩選屬性集 (第二組)	27
表格 10 以粗集理論應用軟體所產生之篩選屬性集 (第三組)	27
表格 11 以篩選屬性產生的部份決策規則表(第一組)	29
表格 12 以篩選屬性產生的部份決策規則表(第二組)	30
表格 13 以篩選屬性產生的部份決策規則表(第三組)	31
表格 14 主要研究變數說明	33
表格 15 研究變數篩選訓練執行結果(第一組)	36
表格 16 研究變數篩選測試執行結果(第一組)	36
表格 17 研究變數篩選訓練執行結果(第二組)	37
表格 18 研究變數篩選測試執行結果(第二組)	37
表格 19 研究變數篩選訓練執行結果(第二組)	38
表格 20 研究變數篩選測試執行結果(第二組)	38

表格 21 本研究預測結果整體正確率之比較表	39
表格 22 第三組資料集各項變數重復出現在各預測規則的組合次數統計	40
表格 23 第三組資料集年齡變數在各數離散類別出現次數統計分析表	40
表格 24 第三組資料集最後一次聯繫月份變數在各數離散類別出現次數統計分析表	41
表格 25 第三組資料集最後聯繫天數變數在各數離散類別出現次數統計分析表	41
表格 26 第三組資料集上次通訊時間變數在各數離散類別出現次數統計分析表	41

第一章 緒論

企業要永續經營，必須保留提供高利潤的顧客，並且可以持續的開拓新的目標顧客。隨著產業日趨成熟，市場趨於飽和時，企業所面臨的競爭也日益激烈，因此策略對企業就愈重要，策略的訂定會影響到顧客決策，不當的政策可能就會導致顧客投向競爭者的懷抱。

顧客關係管理對於企業制定策略提供有利的參考資訊，受到當前企業界與資訊科技業界所熱烈討論，業界與學界紛紛提出不同的解決方案，Prahalad(2002)認為顧客和企業的關係在今日與過去不同。過去以企業為中心，提供高效率取向的價值創造觀點，一項是消費者為逆來順受的目標市場，公司生產什麼，消費者就照單全收，現在拜網際網路知賜，市場改變了，消費者以自己重視的角度看事情。企業必須學習與顧客共同創造價值，在價值創造過程中與消費者建立夥伴關係，因為共同創造已成為競爭首要條件。

資料探勘方法可從歷年的顧客數據中探討出隱含其中的知識，這隱含的知識可以提供企業在訂定策略時做為有利的參考資訊，讓企業可以針對不同的顧客群採用不同的行銷策略。本研究專題使用粗集理論方法做為顧客群分析的工具，利用UCI(University of California, Irvine)機器學習儲存庫取得之荷蘭銀行營銷數據集，以探討出營銷顧客群的隱含知識，產生找出目標顧客群規則，供管理決策人員進行營銷宣傳策略。

第一節 研究動機

顧客群分析是所有企業持續保有營運競爭力的一個環節，顧客群分析可讓企業有效的找出其目標顧客在哪裡，然後再針對這些顧客去進行有效的營運宣傳策略，以獲得更大更多的顧客群，以增加企業的營業績效，並從中獲取更多的營運利潤。企業要永續經營及繼續保有競爭力的最有效方法就是要如何保留有高利潤的顧客群及如何尋找新的目標顧客，因此要如何去確認哪些顧客群可以為企業帶來最高利潤，哪些是企業未來的目標顧客。所以藉由顧客群分析方法以有效掌握企業的目標顧客，本實務專題使用粗集理論從企業營運資料集中找出目標顧客的隱含規則。

粗集理論是資料探勘中的一種方法，其可分析大量資料也能針對複雜的資訊做處理，將每個欄位資料視為一符號（symbol），避免如傳統統計分析時，因資料數值大小不同而獲得簡化與有規則的資訊且它利用近似值的概念，可以表達模糊或不確定性很高的資料，其並可依據自身資訊去處理不精確、不確定與不完全的數據，有別於傳統統計分析只能找出所有影響因子，尚可從影響因子中分類出核心影響因子與非核心影響因子，將粗集理論導入到不確定性多屬性決策中時，其能以較少的屬性因子對資料進行預測，且可獲得較佳之正確率。

本研究以粗集理論應用於企業顧客群分析，以有效找出企業目標顧客，並依據資料數據實證所產生的規則提供企業對目標顧客進行有效的營運宣傳策略。

第二節 研究目的

營運績效是企業的管理營運的指標，要有好的營運績效就需要有好的顧客群，為企業帶來最大的利潤。因此企業必須具有保留優良的顧客及開發更多可為企業帶來利潤的未來目標顧客，所以從歷史營運資料集中探討出目標顧客的隱涵知識及規則 也可提供一間企業獲利、成長、償債、規模等不同構面的績效表現。

本研究專題使用粗集理論資料探勘方法，以銀行顧客群為研究目標，產生顧客群分類規則，找出有利於企業營銷成長的目標顧客，作為企業行銷時選擇何種顧客群之參考。

粗集方法有以下幾個優點：因為不需要任何預備的或額外的關於數據之資訊或假設，同時適用於質性(qualitative) 及量化(quantitative)的因子變數分析，能由經驗數據中擷取隱藏於其中的現象(fact)，並且以符合自然語法的方式表達這些決策規則，所擷取出的決策規則，代表由數據資料庫中，經過消除多餘或無用的訊息後，包含於數據資料中知識(knowledge)，與其它方法相較，模型的決策規則容易理解；對使用者而言，不需額外對最終模型的參數(parameter)進行解讀，因此針對企業家所將面臨的問題，進而提出解決辦法來進行策略分析。(賴家瑞、溫坤禮，2005)。

第二章 文獻探討

本章節主要詳述粗集理論及目標顧客相關概念，並以文獻探討的結果進行有條理的排列與歸納，作為本研究專題後續研究的基礎架構。

第一節 粗集理論

1982 年，波蘭學者 Zdzislaw Pawlak 提出了粗集理論，是一種處理不精確、不確定和不完全數據的新的數學方法。它可以通過對數據的分析和推理來發現隱含的知識、揭示潛在的規律。在粗集理論中，知識被認為是一種分類能力。其核心是利用等價關係來對對象集合進行劃分。

粗集理論提出了知識的約簡方法，是在保留基本知識(信息)，同時保證對象的分類能力不變的基礎上，消除重覆、冗餘的屬性和屬性值，實現對知識的壓縮和再提煉。其操作步驟：(1)通過對條件屬性的約簡，即從決策表中消去某些列；(2)消去重覆的行和屬性的冗餘值。(智庫·百科，2014)。

(一)粗集理論的基本概念

1. 知識的涵義：

在應用上，粗集理論在許多領域得到了應用，比如說：臨床醫療診斷以及電力系統和其他工業過程故障診斷，去判斷出電力系統的故障原因有那些，判斷出的結果就是「知識」。

2. 不可分辨關係與基本集：

當兩個對象由相同的屬性來描述時，這兩個對象在該系統中被歸於同一類，它們的關係稱之為不可分辨關係。

3. 集合的下近似，上近似及邊界區：

把用於分類的知識嵌入集合內，作為集合組成的一部分，一個對象 A 是

否屬於集合 X 需根據現有的知識來判斷，可分為三種情況：(1)對象 A 肯定集合 X 、(2)對象 A 肯定不屬於集合 X 、(3)對象 A 可能屬於也可能不屬於集合 X 。(百度·百科，2012)

(二)粗糙集理論的研究方向

1. 理論

- (1) 利用抽象代數來研究粗糙集代數空間這種特殊的代數結構。
- (2) 利用拓撲學描述粗糙空間。
- (3) 研究粗糙集理論和其他軟計算方法或者人工智能的方法相接合，例如和模糊理論、神經網絡、支持向量機、遺傳算法等。
- (4) 針對經典粗糙集理論框架的局限性，拓寬粗糙集理論的框架，將建立在等價關係的經典粗糙集理論拓展到相似關係甚至一般關係上的粗糙集理論。

2. 應用領域

- (1) 臨床醫療診斷
- (2) 電力系統和其他工業過程故障診斷
- (3) 預測與控制
- (4) 模式識別與分類
- (5) 機器學習和數據挖掘
- (6) 圖像處理

3. 演算法

一方面研究了粗糙集理論屬性約簡演算法和規則提取啟發式演算法，例如基於屬性重要性、基於信息度量的啟發式演算法，另一方面研究和其他智能人工智慧演算法的結合，比如：和類神經網絡的結合，利用粗糙集理論進行數據預處理，以提高類神經網絡收斂速度；和支持向量機 SVM 結合；和遺傳演算法結合。(百度·百科，2014)

粗集理論相關研究

表 1 國內學者粗集理論之相關研究

作者	時間	研究題目	研究方法	研究結論
朱筱嵐	2005	應用模糊與粗集合理論於多準則決策之研究	模糊集合理論與粗集合理論	以模糊決策理論和粗集合理論的特性來改進傳統決策審核的缺點，藉由迷糊集合、模糊數、解模糊化及粗集合的運算過程，將資料轉換成有用的資訊，以獲得最客觀的審核結果。
余尚武、賴佩君	2007	灰色系統、模糊理論與約略集合理論於權變投資組合保險策略之應用	灰關聯分析法及模糊-約略集合法	灰關聯分析法適合短期投資，而模糊約略集合法較適合長期投資。
陳裕文	2010	考量綜合指標於財務危機預警模式之研究	倒傳遞類神經網路、羅吉斯迴歸	倒傳遞類神經網路的效果優於羅吉斯迴歸
葉忠興	2012	企業財務危機診斷模式之購建	約略集合、支援向量機	以約略集合最好，其次為支援向量機。
鍾健平、李靜芳、陳耀茂、吳東洋	2012	以商品企劃七工具及約略集合分析企劃計程車新型服務	商品企劃七工具及約略集合分析	由約略集合分析可得知受訪者選取偏好屬性水準之正確人數，可提供業者更精準之分析調查結果。
古雅婷	2013	應用人工智慧於太陽能網印製程參數與品質特性探勘之研究	倒傳遞類神經網路、可變精確度粗集合(VPRSM)萃取規則、決策樹萃取規則	利 VPRSM 以及 Decision Tree 方法提供現場人員瞭解網印製程產品的情況，藉此對網印製程產品作預先建議，解決太陽能電池矽片於網印製程中隱藏的不良危機、預防製程惡化，並且達到適化生產的目的。

資料來源：李佳育 戴思蘋 張舒雅 葉玉蘭 柯彥吟，2013、古雅婷，2013

(三) 粗集理論演算步驟

1. 資訊系統(Information Systems)

知識表示可藉由資訊系統來完成，資訊系統的基本成分為被研究物件(object)的集合，關於這些物件的知識則是由其準則和他們的準則來描述。資訊系統定義如下：

$$IS=(U, A)$$

U: 論域(Universe)，是一個非空集合，表示資料表中所有物件的有限集合。

$$U = \{x_1, x_2, \dots, x_m\}$$

A : 準則(Attribute)的集合

存在一個準則 $a \in A$ ，則可定義資訊函數(Information function)為：

$$f_a : U \rightarrow V_a$$

其中 V_a 為準則 a 值的集合，稱作 a 的準則域(Domain of attributes a)。

2. 不可辨別關係(Indiscernible Relation)

對於一個準則 $B \in A$ ，則不可識別關係 Ind_B 定義如下：對於兩個物件 x_i 和 x_j ，對於每一個 $b \in B$ 中，假設 $b(x_i) = b(x_j)$ ，則 x_i 和 x_j 為準則 B 的不可識別關係。

在 Ind_B 中相同的等級(Equivalence Class)代表物件中最小可辨別的群，稱作初等集合(Elementary Set)；對於任一個物件 x_i ，在其中 Ind_B 的初等集合可表示成 $\{x_i\}_{Ind_B}$ 。

3. 上近似值和下近似值(Lower and Upper Approximations)

粗集合方法對資料分析有兩個基本的概念，分別為上近似值與下近似值；上近似值為毫無疑問屬於集合，圖 1 為集合中上、下近似值的示意圖。
(李雅菁，2006)

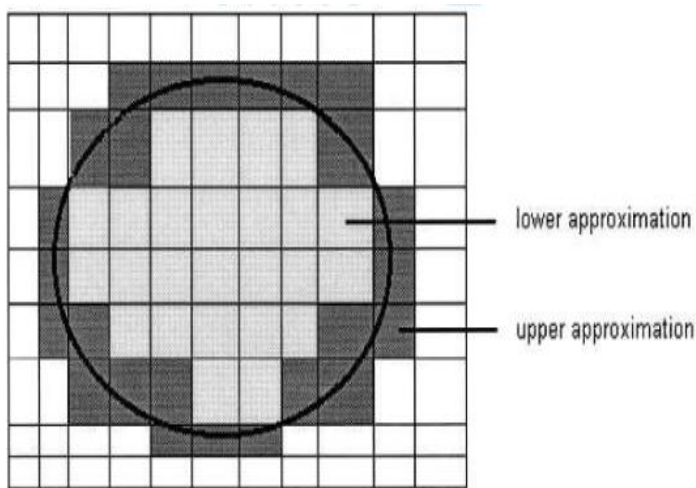


圖 1 上、下近似值的示意圖

關於上、下近似值的定義如下：

上近似(upper approximation):所包含的是所有的決策對象可能屬於的集合範圍，下近似與上近似之間所包含的不同區域則形成模糊邊界概念。
下近似(lower approximation):是指所有的決策對象確定有屬於的集合範圍。

假設 X 為論域 U 元素中的子集合 $X \subseteq U$

下近似值為 $\underline{BX} = \{x_i \in U \mid \exists x_i \in X\}$

上近似值為 $\overline{BX} = \{x_i \in U \mid \exists x_i \in X\}$

X 在 U 中的邊界(Boundary)為 $BNX = \overline{BX} \setminus \underline{BX}$

4. 準則的獨立關係

為了確定準則集合的獨立與否，我們可檢測任一個準則的移除是否增減單位集合的數目。

假設 $Ind(A) = Ind(A \setminus a_i)$ ，則準則 a_i 為多餘可省略的(Dispensable)；否則

準則 a_i 在 A 中是獨立的(Indispensable)。(朱筱嵐，2005)

5. 近似正確率與近似品質

粗集合能夠計算近似度分類的正確率，利用等價類集合 B 產生不可區分關係 $\text{IND}(B)$ 的等價類，由等價類集合 B 定義 X 的正確係數如下：

$$\alpha_B(X) = \frac{\text{card}(\underline{BX})}{\text{card}(\overline{BX})} = \frac{|\underline{BX}|}{|\overline{BX}|}$$

其中 $X \subseteq U$ ， $\text{card}(\cdot)$ 是指在 X 的近似集合當中所包含對象的數目，由上頭的式子來看，每個等價類近似的正確係數是 $0 \leq \alpha_B(X) \leq 1$ ，所以若 $\alpha_B(X) = 1$ ，表示 X 可以完全用等價類集合 B 來描述，也就是 X 在現有的知識域 U 中是可以被定義；除此之外，若 $\alpha_B(X) < 1$ ，表示集合 X 無法用等價類集合 B 來明確說明，集合 X 只是與集合 B 內等價類的某些元素相關，所以集合 X 與集合 B 的等價類關係還是模糊的。

在屬性集合中，假設 ψ 為論域 U 的分類， $X_i \subseteq U$ 為等價類，等價類的基本集合分類 $\psi = \{X_1, X_2, X_3, \dots, X_n\}$ ，而 ψ 在屬性集合 $B \times A$ 中的下近似與上近似分別定義為：

$$\underline{B}\psi = \{\underline{BX}_1, \underline{BX}_2, \underline{BX}_3, \dots, \underline{BX}_n\}$$

$$\overline{B}\psi = \{\overline{BX}_1, \overline{BX}_2, \overline{BX}_3, \dots, \overline{BX}_n\}$$

近似分類的品質，指在決策資訊系統中，由屬性集合 B 正確分類的決策對象數目與所有決策對象數目的比例 γ ：

$$\gamma_B(\psi) = \frac{\sum \text{card}(\underline{BX}_i)}{\text{card}(U)}$$

6. 屬性篩選與核心屬性

在決策資訊系統，不同屬性可能具有不同的重要性，為了找出某些屬性的重要性，粗集理論利用資訊表刪減掉某些屬性，然後觀察在刪除該屬

性後所分類的結果會有什麼變化，在面對大量資料時通常會考慮到，能否刪除決策表中某些資料的屬性特徵，並且還能保持初始資料的分類性質，因為在龐大的資料中，有些資料的屬性可能不會影響對資料分析的結果。所以如果能找出這些不必要的屬性並且刪除，不但不會造成分析上誤差，還能夠提昇資料規則判斷的效率。這種找尋屬性更小的集合，且集合的分類能力與初始集合的分類能力或相同的過程，就稱為篩選(reduction)。

在粗集理論中尋找核心屬性(Core)與屬性篩選(Reduct)是相當重要的一項工作，屬性篩選是降低決策資訊系統屬性項目的部分，而核心屬性則是所有的屬性篩選中屬性項目相同的部分。經過屬性篩選的過程，多餘的屬性被刪除並不會造成分類能力的降低。假設 U 是一個論域， $A、B$ 是 U 上的等價關係，若 $A、B$ 在 U 上導出的所有等價類集合 a 為屬性集合 B 的子集合， $a \times B$ ，而 $B \equiv a$ 與 B 具有相同的區分能力，則稱關係 a 在 B 中是多餘的(redundant)，否則，稱 a 在 B 中是必要的。在等價關係中去除不必要的關係，並不會改變等價關係族的區分能力。

可辨識矩陣包含了決策表中所有的多性資訊，此為一 $n \times n$ 的二維方陣，其中 n 表示集合群的數目，而每個集合內都具有與其他集合不同的元素，利用這些相異的元素項目，將可建構出一個對角線為零的可辨識對稱矩陣。

一般來說，資料通常採用知識表示系統和決策系統對問題進行描述，粗集理論的方法和模型就是建立在二維表基礎上。由於可辨識矩陣是對稱矩陣，在實際的計算上只需算出可辨識矩陣的下三角部分即可。為了更了解可辨識矩陣的意義，利用一個簡單的資訊系統說明，如表 2 所示：

表 2 資訊系統

U	a	b	c	d
x_1	0	0	0	0
x_2	0	2	1	1
x_3	0	1	0	0
x_4	1	2	1	2
x_5	1	0	0	1
x_6	1	2	1	2

在表 3 的資訊系統中，利用可辨識矩陣以及可辨識函數求屬性篩選與核心屬性，整理出可辨識矩陣如 3 所示：

表 3 可辨識矩陣

	1	2	3	4	5	6
1	-					
2	b, c, d					
3	b	b, c, d				
4	a, b, c, d	a, d	a, b, c, d			
5	a, d	a, b, c	a, b, d	b, c, d		
6	a, b, c, d	a, d	a, b, c, d	-	b, c, d	-

再利用可辨函數從可辨識矩陣中獲得資訊系統的屬性篩選與核心屬性，該資訊系統有兩個屬性篩選 $\{a, b\}$ 與 $\{b, d\}$ ，核心屬性是 $\{b\}$ 。

$$\begin{aligned}
\text{而 } \rho(a, b, c, d) &= \frac{1}{b} \frac{1}{c} \frac{1}{d} \frac{1}{a} \frac{1}{b} \frac{1}{c} \frac{1}{d} \frac{1}{a} \frac{1}{b} \frac{1}{c} \frac{1}{d} \frac{1}{a} \frac{1}{b} \frac{1}{c} \frac{1}{d} \\
&= \frac{1}{b} \frac{1}{c} \frac{1}{d} \frac{1}{a} \\
&= \frac{1}{abcd}
\end{aligned}$$

第二節 目標顧客

目標顧客是指企業的產品或者服務的目標對象，是企業產品的直接購買者或使用者。目標顧客要解決的根本問題是，企業準備向哪些市場區間傳遞價值。

企業與市場營銷渠道中的各種力量保持密切關係的目的就是為了有效地向其目標顧客提供產品和服務。顧客的需求正是企業營銷努力的起點和核心。因此，認真分析目標顧客需求的特點和變化趨勢是企業極其重要的基礎工作。

市場營銷學根據購買者和購買目的來對企業的目標顧客進行分類。包括：

- 1、消費者市場：消費者市場由為了個人消費而購買的個人和家庭構成。
- 2、生產者市場：生產者市場是由為了加工生產來獲取利潤而購買的個人和企業構成。
- 3、中間商市場：中間商市場由為了轉賣來獲取利潤而購買的批發商和零售商構成。
- 4、政府市場：政府市場由為了履行政府職責而進行購買的各級政府機構構成。
- 5、國際市場：國際市場由國外的購買者構成，包括國外的消費者、生產者、中間商和政府機構。

每種市場類型在消費需求和消費方式上都具有鮮明的特色。企業的目的

標顧客可以是以上五種市場中的一種或幾種。也就是說，一個企業的營銷對象可以不僅包括廣大的消費者，也包括各類組織機構。企業必須分別瞭解不同類型目標市場的需求特點和購買行為。

本研究專題所探討的目標顧客是屬於消費者市場顧客；消費者市場又稱生活資料市場、最終產品市場。它是指生產經營者從事消費品經營，滿足人們生活消費需要的經濟活動領域，或指消費者為滿足生活消費需要而購買商品的場所。

在網路行銷時代，行銷人的集體焦慮不減反增，因為行銷預算永遠追不上有限流量的轉換率，讓投資報酬率低於 1 的賠本生意，真實的在網路行銷上演。追根究柢，行銷黑洞最根本的肇因是錯投廣告，把廣告投放給錯誤的目標顧客，這就是「有去無回」的行銷成本，那麼，投資報酬率低於 1 也就不足為奇了。

拋開「包山包海」的行銷包裹思維，行銷不是非得漫天撒網、鋪天蓋地式的廣投，新行銷人該學的是「釣竿式的精準投放」，用最少預算、最短時間，找到最精準的客群，這就是大數據改變行銷決策的顛覆式創新，低成本高成效的「行銷前測」，就是大數據行銷的開路先鋒。

低成本高成效的行銷前測

如果你是新創團隊，或是即將推出一個新產品的品牌電商主，如何知道目標顧客在哪裡？來看看一個群眾募資的成功例子，群眾募資結合跨境電商，讓全球都是市場版圖，「Im brief」團隊只花了 300 美元做行銷前測，讓目標顧客們在短短 1 個月內，從全世界裡找到最精準的目標顧客。

「Im brief」是一款智能型公事包，擁有 GPS 防盜、防水、指紋辨識的多項功能，開發團隊原本設想的目標市場是美國矽谷，猜測這類前端又高科技的產品，當地人接受度高。

數據推翻了直覺，行銷前測的結果告訴目標顧客們，完全不是這麼一回事，對「Im brief」最感興趣的市場竟然是在西非，於是開發團隊將行銷預算集中投注單一市場，不僅獲得廣大迴響，也成功找到代理商與投資

人。目標顧客們是怎麼做到的呢？

STEP1：競品分析

如何找到對產品有興趣的潛在目標顧客？第一步要做競品分析，在網購、募資平台找到相似產品，歸納出個別主打的產品性能，以及每個產品採用的媒體宣傳管道，盤點出自身產品的獨特優勢與劣勢。

STEP2：A/B TEST

當你的產品擁有 GPS 定位的防盜功能，還有防水、指紋辨識等多種特性，你該主打哪一項特質？最好的方法是，將 3 個主要特性分別製作 3 種文案，同時投放給目標顧客，藉由點擊率、停留時間的比較，找出網友最感興趣的文案類型。後續再將行銷預算投放在單一特點。

大數據分析團隊 QSearch 行銷企劃郭毅驊分析，科技產品若強調性能，難以找到差異化亮點，因此，最好的文案是「情境化產品」，引導消費者想像「什麼時候會用到它」。

想像一個業務員在午後大雨前往拜訪顧客，公事包俐落外型不怕潮濕，更不怕見顧客失禮，這就是強調防水功能的文案寫法，如果想主打 GPS 定位功能，就可以在文案中強調機密文件，需要防盜的情境功能。

STEP3：監督結果

將擬定的數個文案以小額預算進行行銷前測，透過網路廣告曝光，分別投放給同一特質的目標顧客，開始回收監控結果，比較各文案的點擊率、討論串、分享數、停留時間，找出最佳的文案組合，更好的做法還可將消費者喜歡的特點，延伸製作宣傳影片，強化多媒體曝光效果。

STEP4：決策執行-人類學家的解析

「數據推翻直覺」是大數據分析最可貴的驚喜，行銷決策可以有所本的讓「錢花在刀口上」，但別忘了，人類學家才是告訴你「why」的關鍵人物。

以「Im brief」為例，最有潛力的市場竟然是在西非，而不是設想中

的矽谷，郭毅驊細究當地新聞才解開謎團，因為當地治安不好，防盜公事包的 GPS 定位、指紋辨識功能，完全貼切當地上班族的需求。亂槍打鳥的行銷時代已經過去，小量廣告前測讓「釣竿取代魚網」，用最少預算找到精準目標顧客。以數據為基底的行銷決策，改變了以往行銷人用直覺做判斷的盲點。當數據遇上人類學家，發揮了最好的行銷效能，也是行銷人最能大顯身手的時機。(今天就要的競爭力，2016)

第三節 策略

策略原為軍事用語，自 1950 年代起，策略才開始出現在企業管理的文獻中，當時由於企業管理對於策略的概念還不清楚，故直接引用自軍事用語的意義，將策略定義為 企業運用其所有的資源與技術，達成基本目標的一種科學與藝術。基本上，策略是引導組織行為的一套決策準則，自從策略被引用於管理界之後，經過數十年的演進，策略已成為管理學界相當重要的研究領域，表 4 統整學者策略定義。

表 4 統整學者策略定義

學者	策略定義
Ansoff(1965)	策略是一個廣泛的概念策略提供企業經營方向，並引導企業發掘機會的方針。
Newman&Logan(1971)	策略是確認企業範疇與決定達成目的方式。企業策略首在確認企業所要針對的產品一市場範疇，使組織獲得相對優勢；其次，策略須決定企業如何由目前狀態達到期望的結果，其具體步驟如何，以及如何衡量最後成果。
Kotler(1976)	策略是一個全盤性的概略設計。企業為了達到其所設立的目標，需要一個全盤性計畫，策略就是一個融合行銷、財務、與製造等所擬定之作戰計畫。
Glueck(1976)	策略是企業為了因應環境挑戰所設計的一套統一的、全面的及整合性的計畫，以進一步達成組織的基本目標。
Porter(1980)	企業的競爭策略是企業為了在產業中取得較佳的地位所採取的攻擊性或防禦性行動
Hofer&Schendel(1979)	策略是企業為了達成目標，而對目前及未來在資源部屬及環境互動上所採行得型態。
吳思華(1998)	策略至少顯示下列四方面的意義：評估並界定企業的生存利基、建立並維持企業不敗的競爭

	優勢、達成企業目標的系列重大活動、形成內部資源分配過程的指導原則。
司徒達賢(2001)	策略是企業經營得形貌，以及在不同時點間，這些形貌改變的軌跡。企業形貌包括經營範圍與競爭優勢等重要而足以描述經營特色與組織定位的項目。

資料來源：施育堯. 2006.

第三章 研究方法

本研究之研究範圍，係以 UCI(University of California, Irvine)資料庫荷蘭銀行營銷為研究對象，選取自西元 2002 至 2012 年間之公開資訊資料，且符合選樣標準的公司作分析。本研究以台灣上市公司之企業實務績效評估資料作為研究變數，對於個別總體環境因素的影響則不在本研究的討論範圍之內。

本章內容主要介紹本研究之研究模型，說明本研究實證分析之運作流程、樣本選取、研究變數及演算步驟。

第一節 研究模型

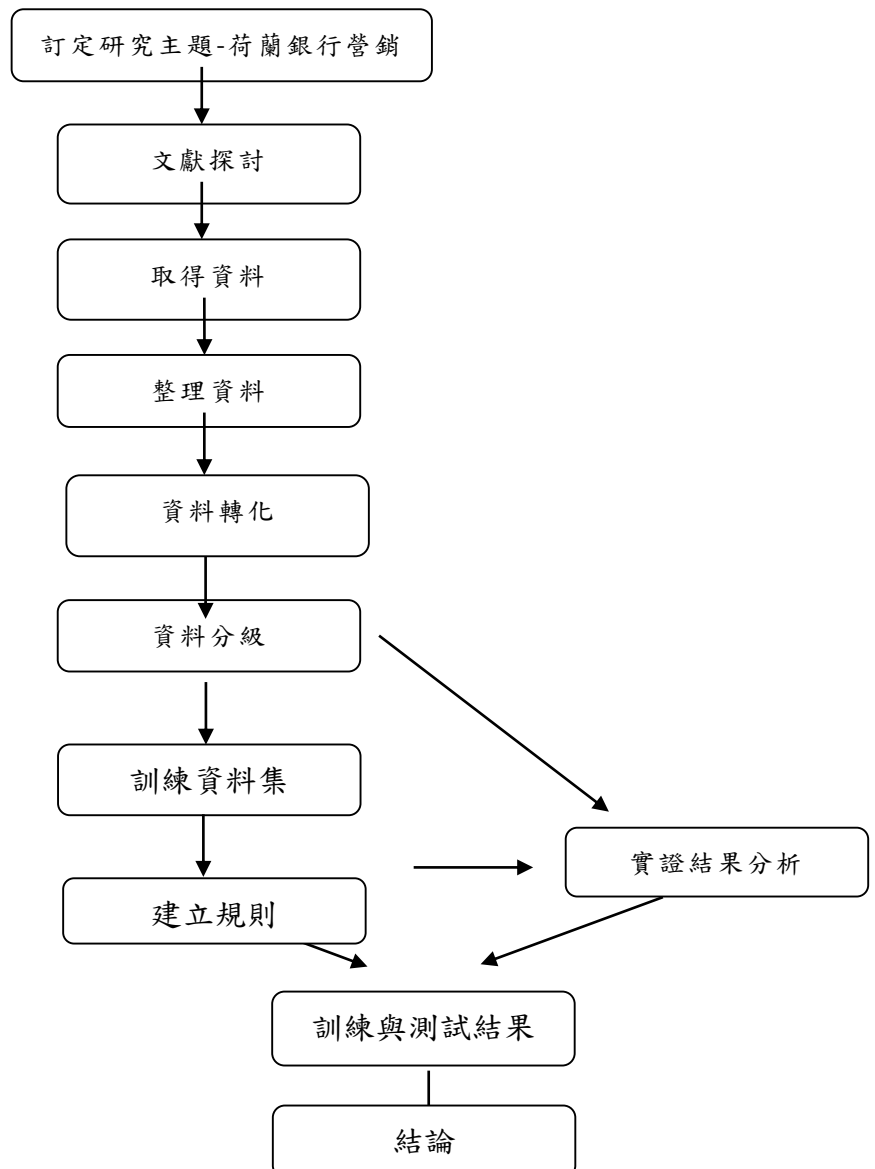


圖 2 研究流程圖

第二節 資料來源

一、資料來源

我們從 UCI(University of California, Irvine)的機器學習庫(Machine learning repository)取得實證資料集，此數據集是公開可用於研究。細節描述[Moro et al., 2014]，一個數據驅動的方法預測荷蘭銀行電話銷售的成功，決策支持系統

二、樣本選取

本研究之研究樣本選自UCI機器學習庫裡提供的荷蘭銀行營銷數據，在選取樣本時，本研究將不完整的比率資料和遺漏值刪除，以利進行研究分析時較易判斷，以提供投資者及管理者作為決策的參考依據。

第三節 演算步驟

本研究應用粗集探勘系統分析軟體 Rough set exploration system (RSES) 為銀行營銷評估之主要資料探勘技術。該軟體為波蘭華沙大學根據粗集合理論開發出來的工具。其透過法則推演技術以擷取大量資料庫中不精確、不確定與模糊之法則資訊。粗集理論的演算步驟包含五部份：資料收集、前置處理、屬性選取、建立規則、訓練與測試。

一、資料收集

本研究選定荷蘭銀行營銷為主之公開行銷成敗資料作為主要探討對象，對其進行行銷策略評估，作為銀行業者之行銷決策；本研究之資料來源，因為銀行顧客資料為個資問題，所以較難取得，主要係取自 UCI 機器學習庫裡提供的荷蘭銀行營銷之公開行銷資料，。表 5 為本研究荷蘭銀行營銷衡量指標之研究變數彙總表。

表 5 本研究荷蘭銀行營銷之研究變數彙總表

研究變數	名稱	說明	資料型態
Age	年齡	這個活動對哪些年齡層成功率高 18~95 歲	整數
Job	工作	這活動對哪種工作的人成功率高	文字
marital	婚姻	已婚或未婚對活動成敗有無幫助	文字
education	教育	教育程度對於成敗有無差別	文字
default	信用	這個活動信用好壞對成功率差別	文字
balance	年平均收入	這個活動對哪階層收入成功率高	整數
housing	房子	有無房子對活動成敗差別	文字
loan	貸款	這個活動對有貸款或沒貸款成功率高	文字
contact	聯繫通訊類型	使用哪種通訊類型較容易成功	文字
day	最後聯繫天數	了解活動開始前多久需聯繫顧客	整數
month	最後一次聯繫月份	在哪個月分聯繫顧客較成功率高	文字
duration	上次通訊時間	與顧客聯繫時間長短較成功率高	整數
campaign	活動期間和此客戶端執行的聯繫人數	活動期間與顧客聯繫人數幾人次對活動成功率高	整數
pdays	客戶上次與之前的廣告系列聯繫後通過的	活動推銷在幾天後較成功率高	整數

研究變數	名稱	說明	資料型態
	天數		
previous	活動之前與顧客聯絡人數	活動之前與顧客聯繫人數幾人次對活動成功率高	整數
poutcome	以前的營銷活動的結果	以前活動成敗的人哪種成功率高	文字
y	決策變數	YES: 成功、NO: 失敗	文字

將所收集之各項研究變數資料，依照其對銀行優惠存款活動之含意，將其評估標準列示如表 5。如「工作」研究變數指標，工作職級越者高薪資所得越高，期傾向於存款的機率就越高，亦代表本優惠存款活動成功率越高；又如「房子」研究變數指標，沒有房子的顧客可能比較有儲存購屋款得動機，所以對本優惠存款活動成功率也會比較高，其他研究變數含義以此類推。

二、前置處理：(分為 3 步驟)

將所蒐集到的資料作前置處理，包含將資料作分類、去除極端值的資料、補足遺漏的資料，並且適當定義所有資料的型態及屬性。主要分為以下三步驟：

- (一)、遺漏值：當比率資料某些欄位可能有遺失內容，必須將比率資料整筆刪除，或者用其他數值取代。
- (二)、錯誤內容：當比率資料不符合所需內容時，為了避免錯誤發生，必須將整筆資料進行刪除。
- (三)、類別化：按照分析的屬性加以分類。

表 6 為部分之原始數據，而表 7 為部分資料前置處理類別化的結果，其是依照分析的屬性加以分類而得。本研究將顧客的各項數值研究變數，依其涵蓋範圍離善化為區分為 5 類 (age)、6 類 (balance)(day)(duration)，並分別以 1、2、3、4、5…來表示。決策值分別為成功(YES)或不成功(NO)。

表 6 部分觀察值之原始資料

年齡	平均收入	最後聯繫天數	上次通訊時間
age	balance	day	duration
58	2143	5	261
44	29	5	151
33	2	5	76
47	1506	5	92
33	1	5	198
35	231	5	139
28	447	5	217
42	2	5	380
58	121	5	50
43	593	5	55

資料來源:本研究整理

表 7 部分觀察值之原始資料

年齡	平均收入	最後聯繫天數	上次通訊時間
age	balance	day	duration
3	1	1	1
1	1	1	1
1	1	1	1
2	1	1	1
3	1	1	1
2	1	1	1
3	1	1	1
3	1	1	1
1	1	1	1
1	1	1	1

資料來源:本研究整理

三、屬性選取：

約簡屬性(Reduct)的觀念就是將原始表格中的一個或多個屬性刪除後，其剩下的屬性仍然可達到原先屬性的分類效果，而被刪除的屬性稱為不必要的屬性，而剩餘的屬性稱為必要的屬性。下圖即為粗理論執行屬性約簡後所得之約簡屬性集；表 8 為第一組；表 9 為第二組；表 10 為第三組。

表 8 粗集理論應用軟體所產生之篩選屬性集（第一組）

(1-1)	個數	正向區域	涵概率	篩選屬性集
1	14	0.992	1	{ age , job , marital , education , default , balance , housing , loan , contact , day , month , duration , campaign , pdays }

資料來源:本研究整理

表 9 以粗集理論應用軟體所產生之篩選屬性集（第二組）

(1-6)	個數	正向區域	涵概率	篩選屬性集
1	15	0.989	1	{ age , job , marital , education , default , balance , housing , loan , contact , day , month , duration , campaign , pdays , poutcome }
2	15	0.989	1	{ age , job , marital , education , default , balance , housing , loan , contact , day , month , duration , campaign , pdays , previous }

資料來源:本研究整理

表 10 以粗集理論應用軟體所產生之篩選屬性集（第三組）

(1-10)	個數	正向區域	涵概率	篩選屬性集
1	14	0.908	1	{ age , job , marital , education , default , balance , housing , loan , contact , day , month , duration , camp , pdays , previous }

資料來源:本研究整理

四、建立規則：

本研究應用粗集理論所提供之資料集分割功能將資料分成兩部份，訓練資料和測試資料，然後以訓練資料集建立決策規則。決策規則就是指一種有條理的方式來描述經屬性篩選過後，由最小屬性集所另外形成的決策表。決策表的分析步驟與資訊系統大致相同，主要有：(1)建構決策表的單位集合，(2)計算決策表中屬性的核心與簡化，(3)計算決策表中屬性值的核心與簡化。

本研究依篩選屬性集作屬性選取，選取多個屬性和 1 個決策屬性，再以粗集之基因演算(Generic Algorithm)方法來訓練產生決策規則(表 11~表 13)。如表 11 所示，第一個欄位(1-25):是在說明 RSES 對訓練資料集進行訓練後，共產生 25 條決策規則；第二欄位(Macth):是指符合支持該對應的決策規則的資料筆數，其說明資料集中有幾筆資料支持該對應的決策規則。例如在規則 1 中，共有 17 筆資料支持該決策規則；最後一欄(Decision rules):決策規則即指當研究變數資料中的條件屬性的排列組合符合決策規則時，則可推導預測出該顧客資料對於優惠存款營銷活動之成功或失敗跡象。例如規則 1，當婚姻為已婚(marital=married)、教育為高中(education=secondary)、聯繫通訊類型為不知道(contact=unknown)、上次連繫持續時間為 1(duration=1)及最後連繫天數為 1(pdays=-1)，則預測決策值為失敗(NO 表示該營銷宣傳不成功)，該例說明在資料集中共有 390 筆資料符合這項決策規則。

表 11 以篩選屬性產生的部份決策規則表(第一組)

規則 (1- 25)	符 合 (Match)	決策規則 (Decisionrules)
1	390	(marital=married)&(education=secondary)&(contact=unknown)&(duration=1)&(pdays=-1)=>(class=no[390]) 390
2	390	(marital=married)&(education=secondary)&(contact=unknown)&(duration=1)&(previous=0)=>(class=no[390]) 390
3	390	(marital=married)&(education=secondary)&(contact=unknown)&(duration=1)&(poutcome=unknown)=>(class=no[390]) 390
4	324	(age=2)&(education=secondary)&(contact=unknown)&(duration=1)=>(class=no[324]) 324
5	272	(marital=married)&(contact=unknown)&(month=jun)&(duration=1)=>(class=no[272]) 272
6	271	(age=3)&(contact=unknown)&(duration=1)&(pdays=-1)=>(class=no[271]) 271
7	271	(age=3)&(contact=unknown)&(duration=1)&(previous=0)=>(class=no[271]) 271
8	271	(age=3)&(contact=unknown)&(duration=1)&(poutcome=unknown)=>(class=no[271]) 271
9	271	(contact=unknown)&(day=1)&(duration=1)=>(class=no[264]) 264
10	264	(age=1)&(marital=married)&(housing=yes)&(duration=1)&(poutcome=unknown)=>(class=no[264]) 264
11	264	(age=1)&(marital=married)&(housing=yes)&(duration=1)&(previous=0)=>(class=no[264]) 264
12	264	(age=1)&(marital=married)&(housing=yes)&(duration=1)&(pdays=-1)=>(class=no[264]) 264
13	244	(marital=married)&(housing=yes)&(loan=no)&(duration=1)&(campaign=2)&(previous=0)=>(class=no[244]) 244
14	244	(marital=married)&(housing=yes)&(loan=no)&(duration=1)&(campaign=2)&(pdays=-1)=>(class=no[244]) 244
15	244	(marital=married)&(housing=yes)&(loan=no)&(duration=1)&(campaign=2)&(poutcome=unknown)=>(class=no[244]) 244
16	243	(marital=married)&(education=secondary)&(contact=unknown)&(month=may)&(duration=1)=>(class=no[243]) 243
17	227	(marital=married)&(housing=no)&(contact=unknown)&(duration=1)=>(class=no[227]) 227
18	200	(balance=1)&(contact=unknown)&(day=2)&(duration=1)=>(class=no[200]) 200
19	194	(job=blue-collar)&(education=secondary)&(contact=unknown)&(duration=1)=>(class=no[194]) 194
20	193	(marital=married)&(contact=unknown)&(duration=1)&(campaign=2)=>(class=no[193]) 193
21	182	(age=2)&(marital=single)&(balance=1)&(housing=yes)&(duration=1)&(previous=0)=>(class=no[182]) 182
22	182	(age=2)&(marital=single)&(balance=1)&(housing=yes)&(duration=1)&(poutcome=unknown)=>(class=no[182]) 182
23	182	(age=2)&(marital=single)&(balance=1)&(housing=yes)&(duration=1)&(pdays=-1)=>(class=no[182]) 182
24	178	(loan=yes)&(contact=unknown)&(duration=1)=>(class=no[178]) 178
25	176	(job=management)&(balance=1)&(contact=unknown)&(duration=1)=>(class=no[176]) 176

資料來源:本研究整理

表 12 以篩選屬性產生的部份決策規則表(第二組)

規則 (1-30)	符 合 (Match)	決策規則 (Decision rules)
1	423	(education=secondary)&(contact=unknown)&(month=may)&(duration=1)=>(class=no[423]) 423
2	261	(marital=married)&(balance=1)&(contact=unknown)&(duration=1)&(campaign=1)=>(class=no[261]) 261
3	255	(age=3)&(balance=1)&(contact=unknown)&(duration=1)=>(class=no[255]) 255
4	247	(contact=unknown)&(day=3)&(duration=1)=>(class=no[247]) 247
5	241	(age=2)&(contact=unknown)&(month=jun)&(duration=1)=>(class=no[241]) 241
6	236	(education=tertiary)&(default=no)&(contact=unknown)&(duration=1)=>(class=no[236]) 236
7	235	(balance=1)&(contact=unknown)&(day=4)&(duration=1)=>(class=no[235]) 235
8	233	(housing=no)&(contact=unknown)&(month=jun)&(duration=1)=>(class=no[233]) 233
9	214	(marital=married)&(housing=yes)&(contact=unknown)&(duration=1)&(campaign=1)=>(class=no[214]) 214
10	212	(contact=unknown)&(month=may)&(duration=1)&(campaign=2)=>(class=no[212]) 212
11	209	(education=secondary)&(housing=yes)&(day=4)&(duration=1)&(pdays=-1)=>(class=no[209]) 209
12	207	(loan=yes)&(contact=unknown)&(duration=1)=>(class=no[207]) 207
13	193	(marital=married)&(balance=1)&(housing=no)&(contact=unknown)&(duration=1)=>(class=no[193]) 193
.....		
20	169	(job=management)&(default=no)&(contact=unknown)&(duration=1)=>(class=no[169]) 169
21	169	(marital=married)&(balance=1)&(housing=yes)&(month=may)&(campaign=2)&(pdays=-1)=>(class=no[169]) 169
22	168	(marital=married)&(housing=yes)&(month=may)&(duration=1)&(campaign=2)&(pdays=-1)=>(class=no[168]) 168
23	163	(contact=unknown)&(duration=1)&(campaign=3)=>(class=no[163]) 163
24	162	(age=2)&(balance=1)&(month=may)&(duration=1)&(campaign=2)=>(class=no[162]) 162
25	159	(job=blue-collar)&(contact=unknown)&(duration=1)&(campaign=1)=>(class=no[159]) 159
26	159	(age=3)&(housing=yes)&(contact=unknown)&(duration=1)=>(class=no[159]) 159
27	158	(age=2)&(housing=yes)&(month=may)&(duration=1)&(campaign=2)=>(class=no[158]) 158
28	157	(age=2)&(education=secondary)&(balance=1)&(day=4)&(duration=1)&(pdays=-1)=>(class=no[157]) 157
29	155	(marital=divorced)&(contact=unknown)&(duration=1)=>(class=no[155]) 155
30	155	(age=2)&(marital=married)&(contact=unknown)&(duration=1)&(campaign=1)=>(class=no[155]) 155

資料來源:本研究整理

表 13 以篩選屬性產生的部份決策規則表(第三組)

規則 (1- 25)	符 合 (Match)	決策規則 (Decisionrules)
1	95	(marital=single)&(education=secondary)&(default=no)&(contact=unknown) &(month=may)&(duration=1)&(pdays=-1)=>(y=no[95]) 95
2	95	(marital=single)&(education=secondary)&(default=no)&(contact=unknown) &(month=may)&(duration=1)&(previous=0)=>(y=no[95]) 95
3	84	(education=secondary)&(contact=unknown)&(day=3)&(month=may)&(duration=1)=>(y=no[84]) 84
4	77	(job=management)&(education=tertiary)&(contact=unknown)&(month=may)&(duration=1)=>(y=no[77]) 77
5	65	(marital=single)&(default=no)&(contact=unknown)&(day=1)&(duration=1)=>(y=no[65]) 65
6	64	(marital=married)&(education=tertiary)&(contact=unknown)&(month=may)&(duration=1)=>(y=no[64]) 64
7	61	(age=2)&(marital=married)&(balance=1)&(contact=unknown)&(day=4)&(duration=1)=>(y=no[61]) 61
8	60	(age=2)&(job=management)&(contact=unknown)&(month=may)&(duration=1)=>(y=no[60]) 60
9	60	(age=2)&(balance=1)&(day=4)&(month=may)&(duration=1)&(pdays=-1)=>(y=no[60]) 60
10	60	(age=2)&(balance=1)&(day=4)&(month=may)&(duration=1)&(previous=0)=>(y=no[60]) 60
11	59	(contact=unknown)&(day=1)&(month=may)&(duration=1)=>(y=no[59]) 59
12	59	(job=services)&(education=secondary)&(contact=unknown)&(month=may)&(duration=1)=>(y=no[59]) 59
13	58	(job=blue-collar)&(day=5)&(month=may)&(duration=1)&(pdays=-1)=>(y=no[58]) 58
14	58	(job=blue-collar)&(day=5)&(month=may)&(duration=1)&(previous=0)=>(y=no[58]) 58
15	57	(age=2)&(housing=yes)&(loan=yes)&(contact=unknown)&(duration=1)=>(y=no[57]) 57
16	57	(age=2)&(balance=1)&(contact=unknown)&(day=4)&(month=may)&(duration=1)=>(y=no[57]) 57
17	57	(job=blue-collar)&(contact=unknown)&(day=5)&(month=may)&(duration=1)=>(y=no[57]) 57
18	55	(age=2)&(education=secondary)&(contact=unknown)&(day=3)&(duration=1)=>(y=no[55]) 55
19	55	(education=secondary)&(contact=cellular)&(month=may)&(duration=2)&(pdays=-1)=>(y=yes[55]) 55
20	55	(education=secondary)&(contact=cellular)&(month=may)&(duration=2)&(previous=0)=>(y=yes[55]) 55
21	54	(housing=no)&(loan=no)&(contact=unknown)&(day=4)&(month=jun)&(duration=1)=>(y=no[54]) 54
22	53	(marital=single)&(education=tertiary)&(default=no)&(balance=1)&(housing=yes)&(contact=unknown) &(duration=1)=>(y=no[53]) 53
23	52	(marital=single)&(education=tertiary)&(contact=unknown)&(month=may)&(duration=1)=>(y=no[52]) 52
24	52	(age=2)&(marital=married)&(loan=yes)&(contact=unknown)&(duration=1)=>(y=no[52]) 52
25	51	(job=blue-collar)&(housing=yes)&(day=5)&(month=may)&(duration=1)=>(y=no[51]) 51

資料來源:本研究整理

五、測試：

預測技術乃依據某一特定對象屬性，觀察其過去的行為或歷史資料，藉以推估其未來的趨勢會是如何。例如由營銷宣傳活動過去之結果記錄預測其未來之投資決策，最後依測試所得正確率以評定預測模式之優劣。本研究將利用步驟四所建立的決策規則測試預測模型之預測正確率。

第四章 實證分析與研究結果

本章內容包括實證分析與研究結果兩大部份。在實證結果分析，將選取的樣本資料，應用粗集理論應用軟體(RSES)，來進行訓練與測試本研究營行銀消模型之整體預測正確率。本研究資料集係由個 17 屬性所組成，其中 16 個屬性為條件屬性(表 14)，1 個為決策屬性，該資料集經由粗集應用軟體萃取出具顯著性的約簡屬性集後，將分割成兩個資料集（訓練與測試資料集），首先以粗集理論應用軟體內之基因演算法進行訓練以產生決策規則，然後再以所產生之決策規則作營行銀消狀況之預測，以獲得本研究模型之測試預測結果。

表格 14 主要研究變數說明

研究變數	研究變數	說明
年齡	Age	了解此活動主要的消費群落在哪個年齡層
工作	Job	工作類型分類：管理員、未知、失業、管理、女傭、企業家、學生、藍領、個體經營、退休、技術員、服務
婚姻	marital	婚姻狀況分類：已婚、離婚、單身；注：“離婚”是指離婚或喪偶
教育	education	分類：未知、次要、小學、三級
信用	default	默認信用？（二進制：是、否）
年平均收入	balance	平均年平均，歐元（數字）
房子	housing	有住房貸款？（二進制：是、否）
貸款	loan	有個人貸款？（二進制：是、否）
聯繫通訊類型	contact	聯繫人通訊類型分類：未知、電話、蜂窩
最後聯繫天數	day	最後一個聯繫日（數字）
最後一次聯繫月份	month	最後一個聯繫月份分類：一月、二月、三月、...、十一月、十二月
持續時間	duration	上一個聯繫人持續時間（以秒為單位）
客戶上次與之前的廣告系列聯繫後通過的	pdays	客戶上次與之前的廣告系列聯繫後通過的天數（數字，-1 表示客戶端以前未聯繫

研究變數	研究變數	說明
天數		過)
活動之前與顧客聯絡人數	previous	在此廣告系列和此客戶端之前執行的聯繫人數量(數字)

資料來源:本研究整理

第一節 實證演算與預測結果

本研究資料樣本取得為 UCI 機器學習資料庫；經過資料前處理篩選後，劃分為 2 組樣本資料集，分割為訓練資料集及測試資料集，以訓練資料集建構決策規則，再以測試資料集驗證模型預測準確率。

本研究證實方法是應用 RSES 軟體進行行銷策略預測之實證分析，該軟體操作介面是以圖形化方式進行，如圖 3 所示。以第一個資料集為例，經屬性約簡後，分割成訓練資料集 4523 筆資料與測試資料集 4523 筆資料，分別載入並開始進行實證演算程序。利用 RSES 軟體之基因演算法來產生(屬性約簡)決策規則，再以所產生之決策規則做營行銀消評估，然後再以測試資料集對本實證研究模型作營行銀消評估之驗證。圖 3 是以綜合屬性資料集來作約簡處理，並依約簡屬性集作屬性選取。

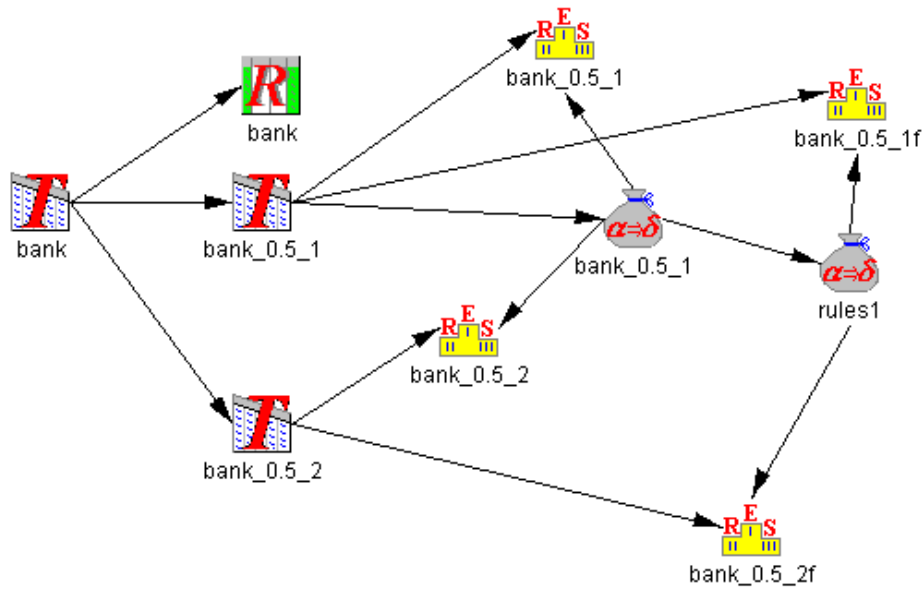


圖 3 荷蘭銀行營銷預測之實證演算流程圖

為了驗證粗集理論方法所建構模型之預測能力，本研究首先載入資料集，以粗集理論屬性約簡功能作屬性約簡，產生之約簡屬性集，然後將資料集分割成訓練資料集與測試資料集，以訓練資料集產生決策規則，最後以測試資料集驗證預測模型之正確率。

一、第 1 組

依訓練資料集所產生之 10119 條決策規則進行訓練預測，然後再使用測試資料集進行行銷策略評估，訓練預測結果列示於表 15；測試預測結果列示於表 16。表 15 為訓練預測結果，該資料集共有 4523 筆，應用 RSES 進行預測，訓練資料集之整體預測正確率為 99.8%。表 16 之測試資料集共有 4523 筆，應用 RSES 進行預測，測試資料集之整體預測正確率為 88.6%。

表 15 研究變數篩選訓練執行結果(第一組)

實際值		no	yes	樣本數量	預測正確率	樣本涵概率
	No	3999	11	4010	0.997	1
	yes	0	513	513	1	1
	正 確 率	1	1			
樣本總數量：4523						
整體預測正確率：0.998						
總樣本涵概率：1						
資料來源:本研究整理						

表 16 研究變數篩選測試執行結果(第一組)

實際值		no	yes	樣本數量	預測正確率	樣本涵概率
	No	3891	64	3968	0.984	0.997
	yes	449	94	555	0.173	0.978
	正 確 率	1	1			
樣本總數量：4523						
整體預測正確率：0.886						
總樣本涵概率：0.994						
資料來源:本研究整理						

二、第 2 組

依訓練資料集所產生之 7495 條決策規則進行訓練預測，然後再使用測試資料集進行行銷策略評估，訓練預測結果列示於表 17；測試預測結果列示於表 18。表 17 為訓練預測結果，該資料集共有 4515 筆，應用 RSES 進行預測，訓練資料集之整體預測正確率為 100%。表 18 之測試資料集共有 4516 筆，應用 RSES 進行預測，測試資料集之整體預測正確率為 88.4%。

表 17 研究變數篩選訓練執行結果(第二組)

實際值		no	yes	樣本數量	預測正確率	樣本涵概率
	No	3939	0	3975	1	0.991
	yes	0	143	540	1	0.265
	正確率	1	1			
樣本總數量：4515						
整體預測正確率：1						
總樣本涵概率：0.904						

資料來源:本研究整理

表 18 研究變數篩選測試執行結果(第二組)

實際值		no	yes	樣本數量	預測正確率	樣本涵概率
	No	3882	40	3967	0.99	0.989
	yes	475	59	549	0.11	0.973
	正確率	0.89	0.6			
樣本總數量：4516						
整體預測正確率：0.884						
總樣本涵概率：0.987						

資料來源:本研究整理

二、第 3 組

依訓練資料集所產生之 6504 條決策規則進行訓練預測，然後再使用測試資料集進行行銷策略評估，訓練預測結果列示於表 19;測試預測結果列示於表 20。表 19 為訓練預測結果，該資料集共有 6027 筆，應用 RSES 進行預測，訓練資料集之整體預測正確率為 99.7%。表 20 之測試資料集共有 6028 筆，應用 RSES 進行預測，測試資料集之整體預測正確率為 82.7%。

表 19 研究變數篩選訓練執行結果(第二組)

實際值		no	yes	樣本數量	預測正確率	樣本涵概率
	No	2563	0	2986	1	0.858
	yes	17	2707	3041	0.994	0.896
	正確率	0.99	1			
樣本總數量：6027						
整體預測正確率：0.997						
總樣本涵概率：0.877						

資料來源:本研究整理

表 20 研究變數篩選測試執行結果(第二組)

實際值		no	yes	樣本數量	預測正確率	樣本涵概率
	No	2293	531	3024	0.812	0.934
	yes	449	2406	3004	0.843	0.95
	正確率	0.84	0.82			
樣本總數量：6028						
整體預測正確率：0.827						
總樣本涵概率：0.942						

資料來源:本研究整理

第二節 研究結果

本研究將預測結果之整體正確率列示在表 21，彙總呈現三組資料集所得之營銷策略預測結果。本研究訓練結果:兩種組合方法之第一組、第二組整體訓練結果正確率第一組為 99.8%，第二組為 100%，第三組為 99.7%。而預測測試正確率結果則分別為:第一組之整體正確率為 88.6%，第二組之整體正確率 88.4%，第三組之整體正確率 82.7%。經三組測試預測正確率比較顯示:第一組優於第二組和第三組，其預測結果整體正確率均在可接受範圍。

表 21 本研究預測結果整體正確率之比較表

屬性	訓練結果			測試結果		
	第一組	第二組	第三組	第一組	第二組	第三組
15 項研究變數	99.8%	100%	99.7%	88.6%	88.4%	82.7%

資料來源:本研究整理

由於第三組成功與失敗總筆數較接近(1 比 1)關係，故本研究即以第三組來做更詳細的分析，依約簡屬性集所產生的決策規則，統計分析各屬性出現次數的權重比率，年齡 (age) 屬性出現次數的權重比率佔整體決策規則的 18.2%；工作 (job) 佔整體的 6.9%；婚姻(marital)佔整體的 8.2%；教育(education)佔整體的 8.3%；信用(default)佔整體的 0.2%；年平均收入(balance)佔整體的 3.5%；房子(housing)佔整體的 7.8%；貸款 (loan)佔整體的 3.8%；聯繫通訊類型(contact)佔整體的 6.2%；最後連繫天數(day)佔整體的 9.8%；最後一次聯繫月份(month)佔整體的 10.6%；上次通訊時間(duration)佔整體的 9.3%；客戶上次與之前的廣告系列聯繫後通過的天數(pdays)佔整體的 3.2%；活動之前與顧客聯絡人數(previous)佔整體的 3.3%。從表 22 權重比率統計分析中發現前四高權重分別為年齡(18.2%)、最後一次聯繫月份(10.6%)、最後聯繫天數(9.8%)及上次通訊時間(9.3%)。

表 22 第三組資料集各項變數重復出現在各預測規則的組合次數統計

變數	研究變數	符合 (match)	比率
年齡	Age	1895	18.2%
最後一次聯繫月份	month	1102	10.6%
最後聯繫天數	day	1025	9.8%
上次通訊時間	duration	973	9.3%
教育	education	868	8.3%
婚姻	marital	853	8.2%
房子	housing	818	7.8%
工作	Job	717	6.9%
聯繫通訊類型	contact	653	6.2%
貸款	loan	395	3.8%
年平均收入	balance	365	3.5%
活動之前與顧客聯絡人數	previous	346	3.3%
客戶上次與之前的廣告系列聯繫後通過的天數	pdays	345	3.2%
信用	default	24	0.2%
總計		10379	100%

資料來源:本研究整理

表 23 第三組資料集年齡變數在各數離散類別出現次數統計分析表

年齡(age)					
類別	第一類	第二類	第三類	第四類	第五類
符合 (match)	285	1319	228	63	0
成功(YES)總計	1895				

資料來源:本研究整理

表 24 第三組資料集最後一次聯繫月份變數在各數離散類別出現次數統計分析表

最後一次聯繫月份(month)												
類別	jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
符合 (match)	9	79	46	122	202	127	105	97	108	138	59	10
成功(YES)總計	1102											

資料來源:本研究整理

表 25 第三組資料集最後聯繫天數變數在各數離散類別出現次數統計分析表

最後聯繫天數(day)						
類別	第一類	第二類	第三類	第四類	第五類	第六類
符合 (match)	230	215	243	145	191	1
成功(YES)總計	1025					

資料來源:本研究整理

表 26 第三組資料集上次通訊時間變數在各數離散類別出現次數統計分析表

上次通訊時間(duration)				
類別	1	2	3	4
符合 (match)	62	732	173	6
成功(YES)總計	973			

資料來源:本研究整理

第五章 結論

本研究以 UCI 取得荷蘭銀行優惠存款營銷活動資料結合粗集理論之應用，以建立營銷策略之預測，透過本研究荷蘭銀行營銷策略預測的結果，提供各荷蘭銀行營銷活動負責人員作營銷策略之參考以提升營銷活動成功效率之衡量。本章根據前述各章節之研究結果，進行統整及歸納，分析出具體的研究結論。

基於分辨實體、狀態、抽象的概念、過程等分類對象，如何對不精確的結果而進行分類數據的能力為粗集理論的特點，也反映了人們習慣使用約略的概念來進行處理不明確問題的習慣。粗集理論其有效性已在許多科學與工程領域的成功應用中得到證實，但截至目前學者甚少應用粗集方法來對企業營銷宣傳活動做一完整的評估。

依據本研究所提出之應用粗集理論預測營銷活動成敗預測模型，第一步驟對收集的 17 項研究變數進行資料整理，整理出 17 項年齡 (age)；工作 (job)；婚姻 (marital)；教育 (education)；信用 (default)；年平均收入 (balance)；房子 (housing)；貸款 (loan)；聯繫通訊類型 (contact)；最後連繫天數 (day)；最後一次聯繫月份 (month)；上次通訊時間 (duration)；客戶上次與之前的廣告系列聯繫後通過的天數 (pdays)；活動之前與顧客聯絡人數 (previous) 為主要研究變數作為本研究實證用屬性，先將數值研究變數作離散化類別，然後以粗集理論的屬性約簡功能粹取出約簡屬性集，再依約簡屬性集選出實證用之資料集，如此可減少資料集之資料維度與複雜性，然後將資料集分割成訓練集及測試集，以訓練資料集建立決策規則進行荷蘭銀行營銷預測，其訓練預測結果及測試預測結果如下所述：

本研究利用三組資料集作實證驗證，每組資料集依約簡屬性集分割成訓練與測試用資料集，然後再利用訓練資料集產生預規則，並進行預測驗證。三組訓練資料集之預測結果整體正確率均為 99.9%。測試資料集預測結果則分別為第一組之整體正確率 88.6%，第二組之整體正確率為 88.4% 及第三組之整體正確率為 82.7%。由於第三組資料集內之成功與失敗記錄

較接近 1:1 的組合，故本研究依第三組資料集所產生的決策規則，依屬性出現次數高低的權重比率作統計分析，分別為年齡（age； 18.2%）；工作（job； 6.9%）；婚姻（marital； 8.2%）；教育（education； 8.3%）；信用（default； 0.2%）；年平均收入（balance； 3.5%）；房子（housing； 7.8 %）；貸款（loan； 3.8%）；聯絡通訊類型（contact； 6.2%）；最後聯繫天數（day； 9.8%）；最後一次聯繫月份（month； 10.6%）；上次通訊時間（duration； 9.3%）；客戶上次與之前的廣告系列聯繫後通過的天數（pdays； 3.2%）；活動之前與顧客聯絡人數（previous； 3.3%），可得知荷蘭銀行營銷活動成敗為第三組主要的驗證依據。因此本研究所產生的預測規則應可完全套用並可正確的反映對荷蘭銀行未來營銷活動的偵測。經統計各屬性於預測規則中出現的次數權重分析結果，表 22 顯示前四高的主要預測屬性分別為年齡(18.2%)、最後一次聯繫月份(10.6%)、最後聯繫天數(9.8%)及上次通訊時間(9.3%)。重複出現次數最高變數為年齡，本研究將年齡變數離散類別化分成五類，出現在年齡各類別的統計次數(表 23)分別為第一類 285 筆、第二類 1319 筆、第三類 228 筆、第四類 63 筆、第五類 0 筆；由此可知如年齡的數據資料類別第二類時，即表示營銷策略對此顧客宣傳成功率較高。依此類推最後一次聯繫月份落在 4, 5, 6, 7, 9, 10，最後聯繫天數落在一至三類及上次上次通訊時間為第二類時營銷策略對這些顧客宣傳的成功率也較高。

(一)依表 23 從本研究實證分析的結果可發現下列情況，首先能夠知道最重要的顧客屬性為「年齡」，研究結果分析發現年齡層為第二類別(33-48 歲)客戶群營銷宣傳成功率最高。

(二)依表 24 第二個重要的屬性為「聯繫月份」，聯繫顧客是每家銀行都會做的，但是根據本研究專題能知道要做行銷時最好通知顧客的月份是在幾月，若是最後聯繫月份是在五月(MAY)成功次數是較多。

(三)依表 25 最後聯繫天數，若是最後聯繫天數是在本研究專題的第三類成功次數較高

(四)依表 26 最後通訊時間，若是最後通訊時間在 2 的話成功次數較高
依據表 23-26 各屬性重複出現權重統計分析，若有荷蘭銀行要做銀行

營銷，可以參考本研究實證分析資訊，應能更容易及有效的找出哪些顧客群對公司行銷成功率較高，以上數據是本研究專題分析出來的前四個影響成功率較高的變數，因此荷蘭銀行要做行銷活動尋找的顧客群首先塞選以上四個條件，對此銀行行銷成功率會有提升的效果，所以本研究實證分析知識可提供營銷宣傳活動經理人簡單易懂的規則找出目標顧客，降低營銷活動成本，並提昇營銷活動成功率。

參考文獻

1. 古雅婷 (2013)。應用人工智慧於太陽能網印製程參數與品質特性探勘之研究。
2. 國立雲林科技大學工業工程與管理研究所碩士論文。
3. 朱筱嵐 (2005)。應用模糊與粗集合理論於多準則決策之研究。義守大學資訊管理研究所碩士論文。
4. 百度. 百科, 2014。http://baike.baidu.com/view/223951.htm。
5. 余尚武、賴佩君 (2007)。灰色系統、模糊理論與約略集理論於權變投資組合保險策略之應用。中華管理評論國際學報, 10 (4)。
6. 李佳育 戴思蘋 張舒雅 葉玉蘭 柯彥吟 (2013)。粗集理論屬性篩選於企業財務危機之預測。嶺東科技大學會計資訊系實務專題。
7. 李雅菁 (2006)。資料探勘分析 IC 封裝業客訴案件之研究。逢甲大學工業工程與系統管理碩士班碩士論文。
8. 許智宇 (2010)。整合 KMV 模型、約略集合及隨機森林應用於企業信用評等之研究。國立臺北科技大學碩士學位論文。
9. 陳裕文 (2010)。考量綜合指標於財務危機預警模式之研究。朝陽科技大學財務金融系碩士論文。
10. 黃俊霖 (2004)。應用約略集理論與模糊理論於台股指數期貨漲跌幅之預測。國立台灣科技大學資訊管理系碩士學位論文。
11. 葉忠興 (2012)。企業財務危機診斷模式之構建。人文暨社會科學期刊, 8(1)。
12. 賴家瑞、溫坤禮 (2005)。粗糙集方法於授信審核之應用。計量管理期刊, 第 2 卷第 1 期, 69-78。
13. 鍾健平、李靜芳、陳耀茂、吳東洋 (2012)。以商品企劃七工具及約略集合分析企劃計程車新型服務。長庚人文社會學報, 5(1), 197。
14. 今天就要的競爭力, 2016。行銷難題: 目標族群在哪? http://www.30.com.tw/article_content_2_9064.html
15. [Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014
16. 施育堯(2006)啤酒行銷策略之研究-以中國上海地區合資公司為例。
17. Prahalad, C. K., & Hammond, A. (2002). Serving the world's poor, profitably. Harvard business review, 80(9), 48-59.
18. UCI(University of California, Irvine)的機器學習庫(Machine learning repository)取得實證資料集