
應用資料探勘技術建構顧客流失預測模型

林金賢
國立中興大學

謝欣樺
國立中興大學

論文編號：IJCS 2020032

收稿 2020 年 7 月 15 日→第一次修正 2020 年 9 月 21 日→正式接受→2021 年 1 月 8 日。

隨著 ICT 技術的進步，利用資料探勘技術從大量數據中挖掘所隱藏的知識對企業越趨重要，尤其在顧客關係管理中的流失顧客預測。雖然決策樹的易讀性以及羅吉斯迴歸捕捉變數間函數關係的優勢使得這兩種方法在流失預測的文獻中常被使用，然而結合此兩種方法優點的羅吉斯葉在流失預測中的適用性並沒有太多的討論。本研究鑒於決策樹的預測結果可視為是一種監督式的分群結果，再輔以羅吉斯迴歸來捕捉每一群中客戶流失的原因，非常契合精準行銷的先分群再建模的概念，因此主張以羅吉斯葉來建構流失預測模型。實證結果顯示羅吉斯葉確實有較好的檢定力與較低的錯誤歸類成本，而決策樹規則與羅吉斯迴歸的顯著變數也都可以提供管理者重要的管理意涵。

關鍵字：資料探勘、顧客關係管理、顧客流失、決策樹、羅吉斯迴歸

Applying Data Mining Techniques to Construct Customer Churn Prediction Model

Chin-Shien Lin

National Chung Hsing Univesity

Hsin-Hua Hsieh

National Chung Hsing Univesity

Paper No. : IJCS 2020032

Received July 15, 2020→First Revised September 21, 2020→Accepted January 18, 2021.

With the advancement of ICT technology, how to use Data mining techniques to discover the potential knowledge from big data is becoming more and more important for enterprises, especially for the churn prediction in the field of customer relationship management (CRM). Although the easy to read rules of Decision Trees and the advantages of Logistic regression in capturing the functional relationships among variables had made these two algorithms widely used in the literature on churn prediction, the applicability of Logit Leaf Model combining the benefit of these two algorithms had not yet been discussed too much. In view of the fact that the prediction result of the Decision Tree at the first stage can be regarded as a supervised clustering result, and it can further be supplemented by Logistic regression to find the causes of customer churn in each cluster at the second stage, which fits the concept of precision marketing, this study advocates using Logit Leaf Model to construct customer churn prediction models. The empirical results showed that Logit Leaf Model had the higher power and the lower misclassification costs among the algorithms, and the rules of the Decision Tree and the significant variables of Logistic regression can also provide decision makers important managerial implications.

Keywords : *Data Mining, Customer Relationship Management, Customer Churn, Decision Tree, Logistic Regression*

緒論

在資訊化的時代下，如何將大量數據中所隱藏的資訊轉化成有價值且可利用的決策依據，對一個公司在變動環境下的永續發展扮演一個很重要的角色。而在顧客關係管理的訴求中，異於傳統單一行銷作法的邏輯，企圖在所擁有的顧客中，精準的挖掘出不同屬性客戶的不同需求，在正確的時間、地點與價格下給予正確的產品，需要的就是資料探勘工具從大量數據中挖掘出所隱含知識的資料分析能力，因此資料探勘技術與顧客關係管理內容已經變成不可分割的一體兩面。顧客關係管理的範疇包括尋找新客戶、發展新客戶、客戶保留、挖掘新需求，其中降低顧客流失率 5% 便能提升獲利能力 25% 至 125%，而且需要投入高達 16 倍的保留客戶成本，才能促使新客戶帶來與既有客戶相同的利潤(Ogino, 2017)，因而如何識別即將流失的客戶已成為不同產業中相同的痛點，也已有許多學者針對此議題進行探討，包括電信業 (Amin, Al-Obeidat, Shah, Adnan, Loo and Anwar, 2019; Coussement, Lessmann and Verstraeten, 2017)、金融業(Kaya, Dong, Suhara, Balcisoy, Bozkaya and Pentland, 2018)、能源業 (Moeyersoms and Martens, 2015)、零售業 (Jahromi, Stakhovych and Ewing, 2014)、服務業 (Barfar, Padmanabhan and Hevner, 2017)等，顯而易見的如何辨識即將流失的客戶已經被視為是顧客關係管理中最嚴峻的挑戰之一 (Miguéis, Van den Poel, Camanho and Falcão e Cunha, 2012)。

另外一方面資料探勘是從大型數據庫中找出有用或有意義的知識的過程，將原始數據轉換成有用的知識型態(Sharma and Sachdeva, 2017)。對於企業而言，資料探勘技術可用來找出或檢測隱藏的顧客特徵和行為，或提供改善行銷，提高顧客滿意度，以及市場分析等決策的依據(Qabbaah, Sammour and Vanhoof, 2017)，甚至於可以用來預測既有客戶轉移至競爭對手的可能性(Giraud-Carrier and Povel, 2003)等。顧客關係管理各個階段的問題可以運用不同的資料探勘技術來分析，包含關聯、分類、集群、預測、迴歸和可視化(Ngai, Xiu and Chau, 2009)。其中最常被用來預測顧客流失的就是分類的技術(Ahmed, 2004; Giraud-Carrier and Povel, 2003)，藉由客戶特徵，找出可能流失的客戶清單，協助企業制訂合適的保留客戶策略，使企業可以花費比較少的成本來留住客戶(Ahmed, 2004)。

而在眾多的分類工具中，最常被用來作流失預測的工具包含有決策樹與羅吉斯迴歸(Koh and Gerry, 2002)。雖然決策樹的易讀性以及羅吉斯迴歸捕捉變數間函數關係的優勢使得這兩種方法在流失預測的文獻中常被使用，然而結合此兩種方法優點的羅吉斯葉較少討論其在流失預測中的適用性。由於變數間的函數關係往往同時兼具直線與非線性特性，如何有效捕捉變數間的關係變得非常重要也不容易。本研究鑒於羅吉斯葉的模型建構過程中(先做決策樹再做羅吉斯迴歸)，決策樹的預測結果可視為是一種監督式的分群結果，再輔以羅吉斯迴歸來捕捉每一群中客戶流失的原因，非常契合精準行銷的先分群再建模的概念，因此主張以羅吉斯葉來建構流失預測模型，認為預測績效將會比單一決策樹或羅吉斯迴歸會有更好的表現。除了提供一個更有效的預測工具之外，決策樹以及羅吉斯迴歸結果的簡單易讀特性也可以提供決策者重要的管理意涵。

此外，顧客流失議題多以 B2C 為對象，較少針對 B2B 進行探討(Figalst, Elsner, Bosch and Olsson, 2019; Jahromi et al., 2014)，且過去探討 B2B 的客戶端類型多以中小型規模以上企業為主，主要差別在於此類型顧客需要透過多位決策者並以企業整體利益為優先，歷經長時且理性的決策過程，不同於 B2C 屬於主觀性高之消費行為(Brown, Zablah, Bellenger and Donthu, 2012; Swani, Brown and Milne, 2014)。由於本研究的資料來自於一家成品農藥加工廠，並以其顧客(農藥行)為對象進行研究，而目前全台約有超過 6000 家農藥行大多屬小型規模，組織類型較為扁平，在營運過程中，大多由其所有者直接進行決策，主觀性高，很類似 B2C 的消費者行為。在同時兼具有 B2B 與 B2C 之交易特性下，本研究希望可以經由驗證的過程了解農藥行之流失因素究竟主因為何，除了可以檢測工具的預測有效性之外，對於其影響因素也可以有進一步的了解，進而找到因應之道。

基於上述，本研究將針對兼具 B2B 與 B2C 性質的環境用藥製造業探討顧客流失議題，與過去研究不同的地方如下：第一，解釋變數包括 B2C 文獻中常用到的客戶基本資料及交易明細，以及 B2B 文獻中供應商評選所考慮的準則，包括客戶服務、品質績效、交期績效、彈性績效與價格，再加上利用原始資料計算出的衍生變數(如顧客行為變數與時空與選擇行為變數)，來探討影響農藥行顧客流失的關鍵因素。第二，鑒於決策樹與羅吉斯迴歸工具本身的侷限性，本研究嘗試

結合此兩種工具的優點，利用決策樹對顧客先作監督式的分群，再針對所得到的子葉(群組)配置羅吉斯迴歸建立顧客流失預測模型，除了希望可以準確的預測績效提供給企業提早發現即將流失的顧客之外，也希望透過決策樹與羅吉斯迴歸的易讀性可以知道影響該公司顧客流失的重要因素，在顧客流失前，作出合適的保留策略。本研究的進行方式如下，第二章為文獻探討，第三章為研究方法介紹資料的來源與如何取得，第四章為資料分析與結果討論，第五章為結論與未來研究建議。

文獻探討

顧客流失

顧客關係管理起源於 20 世紀初，主要在幫助企業取得知識，改善客戶的獲取、保留和獲利能力(Sabbeh, 2018)，其策略包括獲取和部署客戶相關的知識，滿足客戶所需，以使公司能更有效地銷售其更多產品或服務，極大化顧客對組織的價值。而保留客戶為提升客戶忠誠度並減少其流失，是顧客關係管理的主要目標，因為其成效比獲得新客戶之策略高出許多，尤其流失率是企業力求最小化的重要指標，故客戶流失預測是保留客戶必要的一部分(Vadakattu, Panda, Narayan and Godhia, 2015)。

顧客流失通常定義為顧客在一段期間內中止訂購產品或服務，並訂購另一公司的產品或服務(Alamsyah and Salma, 2018)。過去研究主要從兩個不同的角度解決顧客流失問題，分別為提升模型預測能力和了解流失因素，前者為研究人員專注於改善客戶流失預測模型，開發和提出更複雜的模型，以提高預測效能(Verbeke, Dejaeger, Martens, Hur and Baesens, 2012)；後者為研究人員將客戶流失預測視為取決於客戶的個人選擇的管理問題，想了解導致客戶流失的原因，並找出關鍵因素為何，例如客戶滿意度(Gustafsson, Johnson and Roos, 2005; Hansen, Samuelsen and Sallis, 2013)，以便提出適當且有效地保留客戶策略。基於實際的需求，本研究所提出之客戶流失預測模型希望可以兼具良好的預測準確度並可以提供明確的資訊以利作出顧客保留的決策。

本研究分別針對 B2C 與 B2B 整理數篇過去曾探討顧客流失的文獻，並列舉應用的產業、變數、工具及研究貢獻，首先以 B2C 而言(見表 1)，Coussement et al.

(2017)透過預處理縮減變數維度，將高度相關自變數擇一放入預測模型，發現變數篩選對模型預測能力的重要性。Kaya et al. (2018)運用客戶交易行為的時間和位置資訊，計算時空與選擇(Spatio-temporal and choice, STC)行為衍生變數，分別為多樣性、忠誠度及規律性，並使用隨機森林建構了一個動態行為模型，發覺比起傳統變數(如人口統計變數)，時空與選擇行為變數可以更加準確地預測客戶流失。以 B2B 而言(見表 2)，Jahromi et al. (2014)使用 RFM 變數對 B2B 的非合約類型客戶建構預測流失模型，以提升演算法為預測能力最佳的模型，並找出利潤極大化之保留客戶策略，證實以預測模型為導向所進行之保留策略，優於常用的管理啟發式方法。Gordini and Veglio (2017)以義大利消費品電商之企業客戶為研究對象，以顧客基本資料、交易行為、網頁紀錄等預測變數，並運用類神經網路、羅吉斯迴歸、支援向量機、新的演算法基於曲線下面積(Area under the cost curve, AUC)參數選擇之支援向量機(SVM based on AUC parameter selection, SVM AUC)，開發針對 B2B 電商客戶流失之預測模型，結果以基於 AUC 參數選擇之支援向量機為預測能力最佳之演算法。

透過以往文獻可以發現，首先，B2B 與 B2C 之預測模型所採用的變數內容與演算法並沒有太大差異，但多數學者強調預測 B2B 流失帶來的效益優於 B2C(Figalistic et al., 2019; Gordini and Veglio, 2017)，雖然預測 B2B 仍屬少數。再者，預測流失變數以顧客基本資料變數、顧客行為變數為主，部分研究會再針對顧客行為數據計算相關衍生變數，如 Kaya et al. (2018)所提出的時空與選擇行為變數，或過去文獻提及解釋未來客戶行為表現最強變數之一的 RFM 變數(Bose and Chen, 2009; Mitrovic, Singh, Baesens, Lemahieu and De Weerd, 2017)。第三，預測模型隨著研究對象不同，預測能力也有所差異，因此如何編寫新的或組合不同的演算法發展出良好的顧客流失模型也是一重要的研究方向。最後，變數選擇對於提升模型預測能力亦為相當重要，有利於發現顧客流失的原因以助於做出更適當的保留策略。

表 1 B2C 顧客流失預測之相關研究整理

學者	產業	變數	工具
Coussement et al. (2017)	電信業	顧客基本資料變數(如性別、年齡)、顧客行為(如撥出/撥入通話時間(分鐘)、與客服的聯繫次數、資費方案類型)	羅吉斯迴歸、自助聚合(Bagging)、貝氏網路(Bayesian network)、單純貝氏(Naïve bayes, NB)、決策樹、類神經網路(Artificial neural network, ANN)、隨機森林(Random forest, RF)、支援向量機、隨機梯度推進(Stochastic Gradient Boosting)
Kaya et al. (2018)	金融業	顧客基本資料變數(性別、年齡、婚姻狀況、學歷、工作等等)、時空與選擇行為變數(多樣性、忠誠度及規律性)	隨機森林

表 2 B2B 顧客流失預測之相關研究整理

學者	產業	變數	工具
Jahromi et al. (2014)	快速消費品零售業	顧客行為(總交易次數、最後一次交易日、第一次交易日、下半年與上半年交易金額的變化)	決策樹、成本敏感決策樹、羅吉斯迴歸、提升(Boosting)
Gordini and Veglio (2017)	快速消費品電商	顧客基本資料(企業註冊者之人口統計變數、註冊日)、顧客行為(頻率、最近一次購買日、首次登入至今時間差、金額、產品種類、交易失敗次數)、網頁紀錄(使用日期、主機、方法、頁面、請求狀態、登入日期與頁面)	基於 AUC 參數選擇之支援向量機、支援向量機、類神經網路、羅吉斯迴歸

供應商評選準則

過去已有相當多的文獻探討挑選供應商的重要標準(Choi and Hartley, 1996; Dickson, 1966)，強調即時滿足客戶需求的能力，並快速應對需求的變化(Hsu, Kannan, Keong Leong and Tan, 2006)，齊德彰、洪敘峰(2011)統整 14 篇探討供應商評選準則的重要文獻篩選出五個重要考量因素分別為服務品質、品質績效、交期績效、價格、以及彈性績效，以下分別說明本研究以此五個準則作為探討農藥行評估供應商滿意度的構面。

首先為服務品質，服務通常發生在提供服務的員工和客戶間，或所提供的商品或系統，作為解決客戶問題的方法(J. A. Fitzsimmons and M. J. Fitzsimmons, 2000)，因此能否即時為客戶解決困難並使其感到滿意，對客戶選擇供應商作為其合作對象是重要的考量之一。第二為品質績效，品質是供應商選擇過程中的關鍵標

準之一，產品品質的良莠將對整個組織產生重大影響，因為將直接影響採購者向客戶銷售的最終產品，不僅冒著產品被退貨(財務損失)以及終止現有和將來的商業合約的風險，也可能會連帶損害企業的無形資產(如企業形象和聲譽)(McCann and Fingleton, 1996)，因此供應商的品質績效是相當重要的。第三為交期績效，交期為供應商的後勤能力以及從接單至送達客戶端(即訂單履行)之執行關鍵活動和流程，可能影響企業客戶的成本、上市速度或最終用戶如何看待其價值主張，供應商及時交付貨物有可能帶動最終用戶對企業客戶的需求(Stalk, 1988)。第四為彈性績效，彈性是供應商選擇過程中的關鍵標準營運策略模型的重要目標之一(Schroeder, 2000)，彈性反映企業有效適應或回應在客戶眼中增加價值的變化能力(Upton, 1995)，供應商能否快速因應市場變化，對採購商而言是提升其競爭優勢的重要因素。最後為價格，過去研究顯示價格是選擇供應商決定的主要因素(Zeng, Yang, Li and Fam, 2011)，

換言之，客戶可能會因價格低廉，而與某些供應商合作，故價格是供應商選擇的主要考慮因素之一。

時空與選擇行為變數

時空與選擇行為變數由 Kaya et al. (2018) 所提出，分別為多樣性、忠誠度及規律性，不同於過去傳統預測流失變數，如顧客基本資料變數(性別、年齡等)及顧客行為資料變數(交易頻率、交易金額等)，Kaya et al. (2018) 考量顧客交易的時間及地點特性，計算客戶進行交易之時間與地點的集中度、複雜度(多樣性)與規律程度，並應用於金融業 B2C 之情境，進而發現此種變數對預測流失的高度準確性。另外，Kaya et al. (2018) 所提出之忠誠度有別於過去文獻定義主要分為態度忠誠及行為忠誠，前者為客戶產生心理層次上對產品、服務或供應商的整體依戀；後者為客戶持續與同一供應商交易，凝聚與供應商間的關係(Hallowell, 1996)，Kaya et al. (2018) 所提出之忠誠度為交易時間及位置之集中程度，為避免混淆，本研究將 Kaya et al. (2018) 的忠誠度命名為集中度以利區別。此外，本研究之客戶端類型兼具 B2B 與 B2C 之交易特質，具有 B2C 之交易靈活性，因此本研究認為交易行為特性與變化對於預測農藥行流失與否具重要影響，而由於本研究之交易型態多為透過電話訂購配送到客戶端，較不適用於位置資訊的計算，且考量交易週期較長，故本研究以月的上、中、下旬作為交易時間點種類，計算短期(12 個月)多樣性、長期(30 個月)多樣性、短期集中度、長期集中度、以及規律性。

羅吉斯葉

決策樹與羅吉斯迴歸是常見的顧客流失預測工具，主要原因是因為其兼具良好的預測能力及易於解讀的特性(Verbeke et al., 2012)。決策樹是一個被廣泛運用的分類與預測技術，支援類別變數和連續變數，基於此而成為最常見的流失預測模型之一(Sabbeh, 2018)，決策樹屬於監督式學習(Supervised learning)，運用樹狀分支的結構圖來表示分類資料的路徑與規則，圖示結果易於理解(Sheu, 2007)。創建決策樹的訓練過程稱為歸納法，將訓練集(Training set)的資料分割至後續的分類子集合中，且大多數決策樹演算法會歷經兩個階段，分別是成長(分割)階段與修剪階段。決策樹生長階段為迭代過程，第一次迭代從包含所有資料的根節點開始分割，後續迭代在含有數據子集的衍生節點上進行，每次分割都會分析變數並選擇最佳分割點。而決策樹

通常有多種停止分割的規則，包括極大化樹的分支或極小化節點分割所考量的變數種類。而決策樹容易發生過度配適的現象或分割規則不適當，故可使用修剪(Pruning)將不合適的節點或子葉刪除，建立極大化的決策樹並修剪，使其為更適用的常用技術(Bounsaythip and Rinta-Runsala, 2001)。而羅吉斯迴歸是以二元變數為依變數的一種廣義線性模型(Zumel and Mount, 2014)，目的是建立以多個自變數預測一個依變數的迴歸方程式(Heeringa, Weat and Berglund, 2010)，且當線性迴歸無法捕獲自變數與依變數間的關係，更好的替代方法是使用非線性函數，將預測值轉換為 0 至 1 的範圍，且用於對二元依變數建立模型，以羅吉斯和常態機率(Probit)的鏈結函數最常見。而當事件發生($Y=1$)的機率為 $P(Y=1 | x) = \pi(x)$ 時，羅吉斯函數如下所示：

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

其中 $\beta_0, \beta_1, \dots, \beta_k$ 是對數模型的迴歸係數，而 x_1, x_2, \dots, x_k 是自變數，依變數則為勝算比(Odds)取對數，落在區間 $(-\infty, \infty)$ 內的值，勝算比 $(\pi(x))/(1 - \pi(x))$ 則可以取 0 到 ∞ 之間的任何值。

決策樹與羅吉斯迴歸雖然都是相當有效的預測工具，但也各有缺點，決策樹適合處理變數間的交互作用，但是不易處理變數間的線性關係；羅吉斯迴歸則適合處理變數間的線性關係，但無法檢測與容許變數間的交互作用。結合上述兩種的演算法工具，De Caigny, Coussement and De Bock (2018) 認為羅吉斯葉模型(Logit Leaf Model)可以改善其缺點且能保有既有的優點，進行步驟如圖 1 所示，第一步驟，透過決策樹分群，將同質性高的樣本歸類於同一個子葉；第二步驟則針對各個子葉的樣本進行羅吉斯迴歸，透過此二階段演算法，可以提高決策樹與羅吉斯迴歸的預測能力，且決策樹的分群利於後續的資料分析，亦能了解影響各群樣本的重要變數。

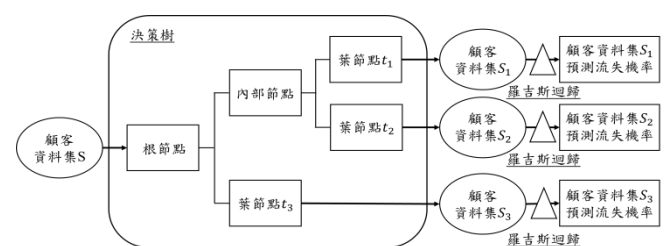


圖 1 羅吉斯葉概念介紹

資料來源：De Caigny et al. (2018)

由於資料集來源包括客戶基本資料、交易明細、供應商評選所考慮準則的滿意度、以及時間相關的顧客行為衍生變數，單一使用決策樹所得到的答案可能因為分類依據包含不同類型的變數(B2B 與 B2C)，原本很好解釋的優點反而變得很不容易解釋，而單一使用羅吉斯迴歸就類似於過去行銷哲學，對所有的顧客都使用同一種行銷手法，忽視其可能存在的市場區隔。基於上述的了解，本研究認為羅吉斯葉可以提供一個更好的解決之道，利用決策樹做顧客分群，然後利用羅吉斯迴歸來探究流失與不流失的影響因素，除了保有兩種工具的優點外，也可以同時兼顧預測準確度與易讀性。

研究方法

研究架構

本研究之架構如圖 2 所示，自變數包括原始資料(客戶基本資料與交易資料)、RFM 行為變數、時空與選擇行為變數、以及供應商評選準則的滿意度。

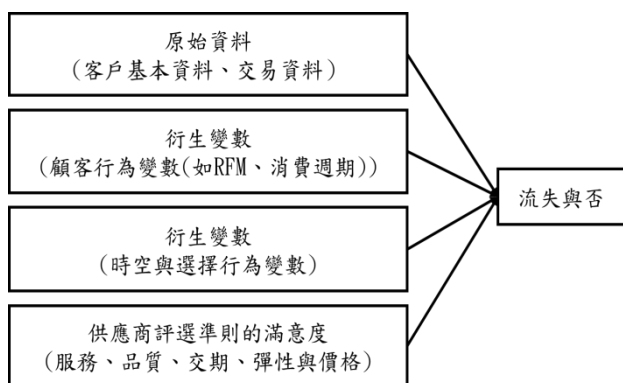


圖 2 研究架構圖

資料來源與分析步驟

該公司原本的資料庫包含有客戶基本資料與交易資料。另外為了蒐集每一個客戶對該公司的感受，本研究採用李克特七點量表，參考過去供應商評選準則的滿意度問卷選擇具代表性的問項，並依據本研究產業特性加以修改共包含五個構面，分別為服務品質(Zeng, Yang, Li and Fam, 2011)、品質績效、交期績效、彈性績效(Maestrini, Luzzini, Caniato and Ronchi, 2018)、以及價格(Wu and Weng, 2010)。資料回收後，進行預處理、淨化資料，計算衍生變數，如 RFM、時空與選擇行為變數等。再將資料集切割成訓練集、驗證集與測試集，並運用決策樹、羅吉斯迴歸與羅吉斯葉分別建

立預測模型，最後透過檢定力與錯誤歸類成本評估找出顧客流失的最佳預測模型。

研究變數

本研究利用原始交易資料的兩個交易區間來計算自變數及依變數如圖 3 所示，以 2014 年 1 月 1 日至 2016 年 7 月 1 日合計共 30 個月的交易資料計算自變數，並以 2016 年 7 月 1 日至 2018 年 1 月 1 日合計共 18 個月的交易資料來計算依變數決定客戶是否流失；以同樣的方式用 2015 年 1 月 1 日至 2017 年 7 月 1 日的交易資料來計算自變數，再用 2017 年 7 月 1 日至 2019 年 1 月 1 日的交易資料來計算依變數來決定客戶是否流失。



圖 3 變數計算區間示意圖

本研究之依變數為流失與否的二元變數，客戶流失的定義會因產業特性改變而有所不同，如金融業的流失客戶定義為註銷所有帳戶的人(Van den Poel and Larivière, 2004)，而本研究將流失客戶定義為交易期間最後 18 個月沒有交易的顧客，反之，交易期間最後 18 個月內有交易的顧客則定義為未流失客戶。雖然資料集的變數內容有時間的重疊性，然而由於函數的關係是在找尋一段期間的採購行為來預測另一段期間的採購行為(流失與否)，模型的建構是將縱斷面的資料轉化成橫斷面的資料來作預測，資料集的分析單位是不同的顧客在一段期間的行為整理出來的一筆資料與 18 個月之後的流失與否，而非同一顧客不同時間的資料點，因此序列相關的問題並不存在。自變數的定義詳見附表 1 並分別敘述如下：

顧客基本資料變數

除了在資料庫中的顧客基本資料外，本研究亦使用財政部公開資料取得各家公司資本額及成立時間，運用 Google earth 專業版將地址的文字資料轉換為圖層(圖 4)，再將圖層匯入 QGIS 將每筆資料轉換為經緯度，最後使用 EXCEL 2 earth 的地圖計算功能，透過經緯度計算出農藥行方圓 500 公尺內同為個案公司客戶之農藥行家數，作為商闖內競爭程度的衡量指標(販售同質性商品)。

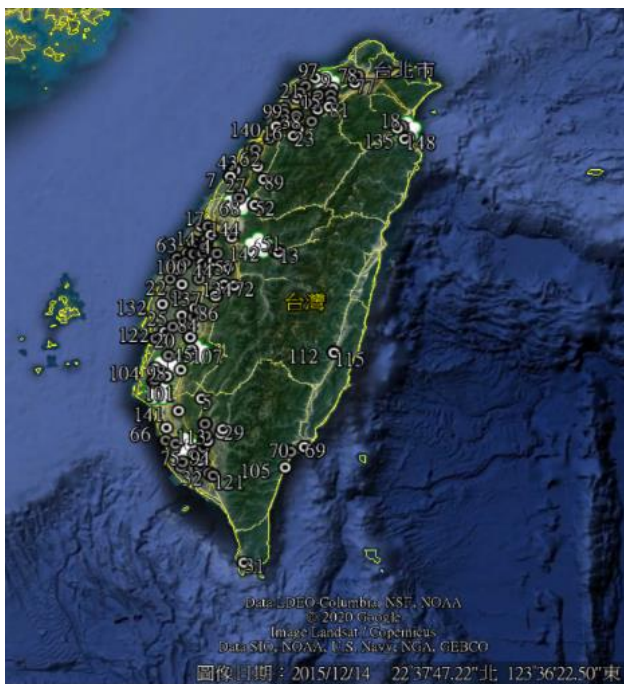


圖 4 農藥行地址轉換圖層示意圖

顧客行為變數

藉由顧客與個案公司在資料期間內的交易資料來計算交易行為變數，以區間最後一日為基準點計算最近一次交易日距離基準點的天數為 Recency；以總交易次數來計算交易頻率(Frequency)；以總交易金額與交易總次數之比率來計算平均交易金額(Monetary)，將上述數值分別轉換為五個級數來計算 RFM 總和，每一位客戶的最小值為 3，最大值為 15。並計算交易期間最後一年內之總交易次數，以及交易期間最後一年內之平均交易金額。

時空與選擇行為變數

為提升預測準確度，本研究參考 Kaya et al. (2018) 對時空選擇變數的定義來計算新的變數。時間的多樣性代表客戶在不同時間點進行交易的傾向，值越大代表客戶的交易時間點越分散。令 p_{ij} 為客戶 i 在時間點 j 進行交易的機率； N 為有進行交易的時間點數量； M 為時間點總數量，多樣性 D_i 計算公式如下：

$$D_i = -\sum_{j=1}^N p_{ij} \log_M p_{ij}$$

集中度則代表顧客在特定時間進行交易的程度，值越高則代表集中程度越高。令 f 為交易次數最多的時間點，客戶 i 的集中度 L_i 計算公式如下：

$$L_i = p_{if}$$

規律性代表短長期交易的多樣性和集中度的相似程度，越接近 1 表示規律性越高，短期與長期的交易

行為無太大的變化。令 D_i^S 和 D_i^L 為短期及長期的多樣性， L_i^S 和 L_i^L 為短期及長期的集中度，客戶 i 的規律性 R_i 計算公式如下：

$$R_i = 1 - \sqrt{((D_i^S - D_i^L)^2 + (L_i^S - L_i^L)^2)/2}$$

供應商評選滿意度

供應商評選滿意度的五個構面包括服務品質(4 題)、品質績效(2 題)、交期績效(2 題)、彈性績效(4 題)及價格(2 題)，包含 4 個題項之構面，用探索性因素分析進行維度縮減；只有兩個題項之構面，則將各題項先以標準化處理後再平均進行維度縮減。

模型評估

當流失客戶被預測分類為流失客戶時稱真陽性(True Positive, TP)，被分類為未流失客戶時則稱偽陽性(False Positive, FP)；當未流失客戶被分類為流失客戶時稱偽陰性(False Negative, FN)，被分類為未流失客戶時則稱真陰性(True Negative, TN)。而一般的預測模型評估指標包括準確性(accuracy)、精度(precision)、召回率(recall)、以及特異性(specificity)如表 3 所列。而相較於精確預測未流失客戶，正確預測流失客戶更為重要(Datta and Das, 2015; Weiss et al., 2007)，對企業而言，若將流失客戶歸類為未流失客戶，可能會損失該客戶原先能帶給企業的效益，我們稱之型二錯誤歸類成本 C_{FN} ，反之，若將不會流失客戶歸類為流失客戶，企業對其採取多餘的客戶保留措施，浪費不必要的成本，稱之型一錯誤歸類成本 C_{FP} 。而過去研究指出需要投入高達 16 倍保留客戶的成本，才能使新客戶帶來的利潤與既有客戶相同(Ogino, 2017)，在不失一般推論的原則下，本研究令 $C_{FN}=16$ 與 $C_{FP}=1$ 計算錯誤歸類成本，作為本研究的主要評估指標之一。

表 3 評估績效指標

指標	計算方式
錯誤歸類成本	$C_{FN} \times FN + C_{FP} \times FP$
準確性(Accuracy)	$(TP + TN)/(TP + FP + TN + FN)$
精度(Precision)	$TP/(TP + FP)$
召回率(Recall)	$TP/(TP + FN)$
特異性(Specificity)	$TN/(FP + TN)$

研究結果

資料樣本特性

本研究的資料來源為位於中部的一家農藥及環境用藥製造公司從 2014 年至 2018 年的顧客基本及交易資料，以客戶類型為農藥行作為研究對象，交易的資料筆數約 3260 筆，共 151 家農藥行，並針對這些農藥行郵寄與電訪發放問卷，另將已歇業及資料不全之樣本剔除後，共回收 139 份有效問卷，有效問卷回收率為 92.05%。另外，將資料集分為 2014 年至 2017 年的 A 資料集(136 家農藥行，流失比例 15.44%)、2015 年至 2018 年的 B 資料集(129 家農藥行，流失比例 16.27%) 分別計算自變數及依變數再將其合併，並以 60%、20%、20% 的比例，分割訓練集、驗證集、測試集，分割後流失比例，訓練集為 15.29%、驗證集為 16.67%、測試集為 16.67%。

農藥行基本資料敘述性統計

表 4 為農藥行基本資料之敘述性統計，顯示地區比例以中部最多為 41.01%；南部次多為 33.09%。而成立年數比例則以 41-50 年最多為 30.94%；以 11-20 年次多為 20.14%。最後，資本額比例以 1 萬以下最多為 81.29%，以 1-10 萬次多為 10.79%。

表 4 農藥行基本資料之敘述性統計

變數	項目	次數	百分比	累計百分比
地區	北部地區	30	21.58%	21.58%
	中部地區	57	41.01%	62.59%
	南部地區	46	33.09%	95.68%
	東部地區	6	4.32%	100.00%
成立年數	10 年以下	19	13.67%	13.67%
	11-20 年	28	20.14%	33.81%
	21-30 年	14	10.07%	43.88%
	31-40 年	22	15.83%	59.71%
	41-50 年	43	30.94%	90.65%
	51 年以上	13	9.35%	100.00%
資本額	1 萬以下	113	81.29%	81.29%
	1-10 萬	15	10.79%	92.09%
	10-100 萬	9	6.47%	98.56%
	100 萬以上	2	1.44%	100.00%

供應商評選滿意度問卷之統計分析

供應商評選滿意度五個構面之而服務品質、品質績效、交期績效、彈性績效及價格此五個構面的 Cronbach's α 值分別為 0.847、0.719、0.831、0.865 及 0.722，信度係數均在 0.7 以上，表示具有良好之信度。此外，本研究進行適合性評估以確保問卷資料集適用於因素分析，服務品質(KMO=0.544; 卡方值為 363.203(p< 0.01))與彈性績效(KMO=0.829; 卡方值為 256.564 (p< 0.01))之 KMO 值均大於 0.5，符合 Wania, Kühn and Klotz (2006)採用可接受之 KMO 值標準，且 Bartlett 球形檢定結果皆為顯著，表示母體之間相關矩陣中有共同因素存在。本研究使用主成份分析法，利用相關性矩陣為輸入矩陣來萃取因素；使用特徵值大於 1 的準則作為萃取因素個數的準則。探索性因素分析結果顯示服務品質與彈性績效之單一因素能解釋 69.119%與 71.483%的總變異量。

資料探勘模型分析結果

資料探勘模型分析結果，依序為決策樹、羅吉斯迴歸、羅吉斯葉、以及各模型之分析比較。

決策樹

本研究的決策樹使用 CART 演算法建構，以四大類共 21 個變數作為輸入變數；流失與否作為目標變數，結果如圖 5 所示，規則一為當客戶總交易次數小於兩次時，預測顧客流失的機率訓練集為 61.90%；驗證集為 83.33%，顯示當總交易次數低於兩次時，客戶容易成為流失客戶；規則二為當客戶總交易次數為兩次以上時，預測顧客流失的機率，訓練集為 8.09%；驗證集為 8.33%，顯示當總交易次數超過兩次時，客戶較不容易成為流失客戶。模型評估績效指標如表 5 所示，透過在同一模型匯入驗證集與測試集對照評分後，可以發現訓練集、驗證集、測試集的結果相近，代表此分析結果並無發生過度配適情況。

表 5 決策樹模型績效評估指標

評估指標	訓練集	驗證集	測試集
錯誤歸類成本	184	65	69
準確性	87.90%	90.74%	83.33%
精度	61.90%	83.33%	50.00%
召回率	54.17%	55.56%	55.56%
特异性	93.98%	97.78%	88.89%

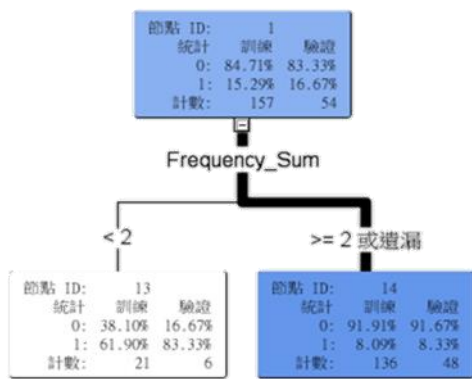


圖 5 決策樹狀圖

羅吉斯迴歸

羅吉斯迴歸模型以逐步羅吉斯迴歸來作分析，將輸入顯著水準設為 1；停留顯著水準設為 0.1，以四大類共 21 個輸入變數；流失與否作為目標變數，實證結果如表 6 所示，有兩個自變數顯著，分別為「RFM 總和」、「長期多樣性(旬)」，羅吉斯迴歸模型評估績效指標如表 7 所示。透過在同一模型匯入驗證集與測試集對照評分後，發現訓練集、驗證集、測試集的結果相近，代表分析結果並無發生過度配適情況。羅吉斯迴歸模型如下：

$$\ln\left(\frac{\pi}{1-\pi}\right) = 1.5727 - 0.2838 \times \text{RFM 總和} - 2.2298 \times \text{長期多樣性(旬)}$$

表 6 羅吉斯迴歸預測結果

變數	係數	標準誤	P 值	勝算比
截距	1.5727	0.7794	0.0436**	
RFM 總和	-0.2838	0.1168	0.0151**	0.753
長期多樣性(旬)	-2.2298	0.7114	0.0017***	0.108

註：*P<0.1; **P<0.05; ***P<0.01

表 7 羅吉斯迴歸模型績效評估指標

評估指標	訓練集	驗證集	測試集
錯誤歸類成本	276	81	66
準確性	86.62%	88.89%	88.89%
精度	63.64%	80.00%	71.43%
召回率	29.17%	44.44%	55.56%
特異性	96.99%	97.78%	95.56%

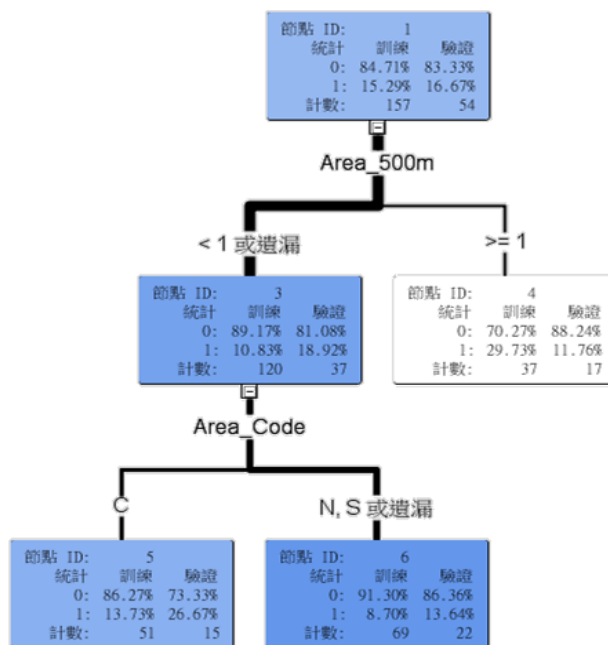
註：*P<0.1; **P<0.05; ***P<0.01

羅吉斯葉

本研究首先以六個顧客基本資料的變數作為決策樹的輸入變數，再將決策樹的子葉加入羅吉斯迴歸模

型內分析，以顧客行為、時空與選擇行為、供應商評選滿意度等三大類共 15 個變數作為羅吉斯迴歸的輸入變數，並皆以流失與否作為目標變數，使用逐步羅吉斯迴歸分析，將輸入顯著水準設為 1；停留顯著水準設為 0.1。決策樹及羅吉斯迴歸結果將分別列示如下：

第一階段決策樹結果如圖 6 所示。規則一為當客戶方圓 500 公尺內的農藥行家數為一家以上時，預測顧客流失機率訓練集為 29.73%；驗證集為 11.76%；規則二為當客戶方圓 500 公尺內的農藥行家數低於一家且位在中部地區時，預測顧客流失的機率訓練集為 13.73%；驗證集為 26.67%；規則三為當客戶方圓 500 公尺內的農藥行家數低於一家且不位在中部地區時，預測顧客流失的機率訓練集為 8.70%；驗證集為



13.64%。

圖 6 羅吉斯葉之決策樹樹狀圖

第二階段首先對決策樹規則一(當客戶方圓 500 公尺內的農藥行家數為一家以上時，預測顧客流失的機率訓練集為 29.73%；驗證集為 11.76%)進行逐步羅吉斯迴歸分析，結果如表 8 所示，15 個自變數中「長期多樣性(旬)」顯著。羅吉斯迴歸模型如下：

$$\ln\left(\frac{\pi}{1-\pi}\right) = 1.9770 - 5.3062 \times \text{長期多樣性(旬)}$$

表 8 決策樹規則一之羅吉斯迴歸預測結果

變數	係數	標準誤	P 值	勝算比
截距	1.9770	0.9541	0.0383**	
長期多樣性(旬)	-5.3062	1.5511	0.0006***	0.005

針對第一群，客戶的方圓 500 公尺內有一家以上的同業(農藥行)，所配置的羅吉斯迴歸顯示「長期多樣性(旬)」為預測流失客戶之重要變數，長期多樣性(旬)每增加一單位，客戶成為流失客戶的機率會比成為非流失客戶的機率減少 99.5%，顯示當農藥行之商闖內有一家以上的同業競爭時，長期多樣性越低之客戶較容易成為流失客戶。

再對決策樹規則二(當客戶方圓 500 公尺內的農藥行家數低於一家且位在中部地區時，預測顧客流失的機率訓練集為 13.73%；驗證集為 26.67%)進行逐步羅吉斯迴歸分析，結果如表 9 所示，15 個自變數中「最後一年總交易次數」為顯著。羅吉斯迴歸模型如下：

$$\ln\left(\frac{\pi}{1-\pi}\right) = -2.6543 - 4.7020 \times \text{最後一年總交易次數} \\ -0.0014 \times \text{平均交易金額} + 2.2341 \\ \times \text{彈性績效} + 11.8517 \times \text{長期集中度(旬)}$$

表 9 決策樹規則二之羅吉斯迴歸預測結果

變數	係數	標準誤	P 值	勝算比
截距	-2.6543	5.4537	0.6265	0.070
最後一年總交易次數	-4.7020	2.5308	0.0632*	0.009
平均交易金額	-0.00141	0.0009	0.1186	0.999
彈性績效	2.2341	1.3744	0.1041	9.338
長期集中度(旬)	11.8517	9.3878	0.2068	999.000

註：*P<0.1; **P<0.05; ***P<0.01

針對第二群，客戶方圓 500 公尺內的農藥行家數低於一家且位在中部地區時，「最後一年總交易次數」為預測流失客戶之重要變數，最後一年總交易次數每增加一單位，客戶成為流失客戶的機率會比成為非流失客戶的機率減少 99.1%，顯示當農藥行之商內無同業競爭且位於中部地區時，最後一年總交易次數越低之客戶較容易成為流失客戶。

最後對決策樹規則三(當客戶方圓 500 公尺內的農藥行家數低於一家且不位在中部地區時，決策樹預測顧客流失的機率訓練集為 8.70%；驗證集為 13.64%)進行逐步羅吉斯迴歸分析，結果如表 10 所示，15 個自變數中「總交易次數」、「平均交易金額」、「RFM 總和」為顯著。羅吉斯迴歸模型如下：

$$\ln\left(\frac{\pi}{1-\pi}\right) = 3.4783 + 0.1589 \times \text{總交易次數} \\ + 0.00027 \times \text{平均交易金額} \\ - 1.26863 \times \text{RFM 總和}$$

表 10 決策樹規則三之羅吉斯迴歸預測結果

變數	係數	標準誤	P 值	勝算比
截距	3.4783	1.9208	0.0702*	
總交易次數	0.1589	0.0807	0.0491**	1.172
平均交易金額	0.00027	0.00013	0.0407**	1.000
RFM 總和	-1.26863	0.4388	0.0034***	0.276

註：*P<0.1; **P<0.05; ***P<0.01

針對第三群，客戶方圓 500 公尺內的農藥行家數低於一家且不在中部地區時，「總交易次數」、「平均交易金額」、「RFM 總和」為預測流失客戶之重要變數，總交易次數每增加一單位，客戶成為流失客戶的機率會比成為非流失客戶的機率增加 17.2%；RFM 總和每增加一單位，客戶成為流失客戶的機率會比成為非流失客戶的機率減少 72.4%，顯示當農藥行之商闖內無同業競爭且位於中部地區時，總交易次數越高、RFM 總和越低之客戶較容易成為流失客戶，此外，平均交易金額每增加一單位，客戶成為流失客戶的機率會與成為非流失客戶的機率相等。

單一演算法的羅吉斯模型下，總交易次數越多則客戶越不容易成為流失客戶，但在羅吉斯葉的模型下，於決策樹規則三，客戶方圓 500 公尺內的農藥行家數低於一家且不位在中部地區時，總交易次數越高反而較容易成為流失客戶，非中部地區之客戶相較於中部地區的客戶而言，基於地理位置的限制，交易成本較高，而 t 檢定結果顯示，其供應商評選滿意度的服務品質(t=-4.34, p<0.01)、品質績效(t=-1.8, p<0.05)、彈性績效(t=-5.36, p<0.01)與價格(t=-1.87, p<0.1)的分數，非中部地區客戶皆顯著低於中部地區客戶，而單一演算法羅吉斯迴歸的結果顯示供應商評選滿意度，並非影響流失與否之關鍵因素，且由 t 檢定的結果得知，同樣位於非中部地區之流失客戶與未流失客戶於此四構面皆無顯著差異，表示非中部地區的客戶普遍認為個案公司於服務品質(t=1.38, p>0.1)、品質績效(t=1.37, p>0.1)、彈性績效(t=1.71, p>0.1)、價格(t=1.60, p>0.1)上表現不佳，但由於其產業屬於特許行業，限制多、門檻高，對客戶而言，轉換供應商的成本較高，因此大部分的客戶仍選擇繼續與個案公司往來(e.g., Lam et al., 2004; Russo & Confente, 2017)，但可能隨著交易次數的增加，考量交易成本與認為其無法改善現況，且近幾年供應商競爭廠商家數漸漸增加，有了更多選擇，故轉而選擇與其他供應商合作。

整合決策樹規則一、二、三之羅吉斯迴歸之模型評估績效指標如表 11 所示，對照評分後，發現訓練集、驗證集、測試集的結果相近，可發現分析結果並無過度配適情況。

表 11 羅吉斯葉模型績效評估指標

評估指標	訓練集	驗證集	測試集
錯誤歸類成本	132	50	52
準確性	92.36%	90.74%	87.04%
精度	80.00%	75.00%	60.00%
召回率	66.67%	66.67%	66.67%
特異性	96.99%	95.56%	91.11%

模型比較

模型的預測績效評比可以使用不同的指標，本研究列出不同指標下各個方法的表現結果如表 12 所示。羅吉斯葉的召回率(訓練集 66.67%、驗證集 66.67%、測試集 66.67%)在三個模型中最高；羅吉斯葉的錯誤歸類成本(訓練集 132、驗證集 50、測試集 52)在三個模型中最低。然而，羅吉斯葉的表現只在召回率與錯誤歸類成本下表現得比較好，其餘三個指標並不是最好。進一步檢視此結果可以發現，此三個指標的內涵並沒有考慮到錯誤歸類成本。由於流失預測的最終目的是找出潛在的流失客戶，提前作流失管理以預防客戶真實流失，而錯估流失客戶為正常客戶的錯誤歸類成本遠大於將正常客戶歸類為流失客戶，從效益的觀點來看，錯誤歸類成本的考量會較純粹的準確度、精度、特異性與召回率要來得重要。Zhu, Baesens and Broucke (2017) 指出目標函數的設定(績效指標挑選)對於預測工具的績效有很大的影響，而且也與資料集的特性有關係。上述結果顯示在錯誤歸類成本、準確度、精度、召回率與特異性五種不同的評估指標下，顯示出針對此資料集召回率和錯誤歸類成本此兩個指標與羅吉斯葉有最好的對應，更加說明了評估指標與工具間對應之重要性。

表 12 模型間的績效比較

評估指標	模型	訓練集	驗證集	測試集
錯誤歸類成本	決策樹	184	65	69
	羅吉斯迴歸	276	81	66
	羅吉斯葉	132	50	52
召回率	決策樹	54.17%	55.56%	55.56%
	羅吉斯迴歸	29.17%	44.44%	55.56%
	羅吉斯葉	66.67%	66.67%	66.67%

表 12 模型間的績效比較(續)

評估指標	模型	訓練集	驗證集	測試集
準確度	決策樹	87.90%	90.74%	83.33%
	羅吉斯迴歸	86.62%	88.89%	88.89%
	羅吉斯葉	92.36%	90.74%	87.04%
精度	決策樹	61.90%	83.33%	50.00%
	羅吉斯迴歸	63.64%	80.00%	71.43%
	羅吉斯葉	80.00%	75.00%	60.00%
特異性	決策樹	93.98%	97.78%	88.89%
	羅吉斯迴歸	96.99%	97.78%	95.56%
	羅吉斯葉	96.99%	95.56%	91.11%

結論建議及管理意涵

研究結論

對於該環境用藥製造商而言，其農藥行客戶同時具備 B2B 與 B2C 的特性，因而本研究在預測其農藥行客戶流失行為時，除了使用每家農藥行的基本資料，包括迄今的成立時間、資本額、地理位置方圓五百里之內的競爭者個數外，也包含了過去 B2B 常用的供應商評選準則，以及 B2C 常用的顧客行為變數 RFM，及使用交易資料產生的時空與選擇行為衍生變數。另外，鑒於預測工具的特性不同，本研究嘗試在配置模型的不同階段挑選不同的變數，來驗證工具的結合可以提升各別工具的預測績效。本研究分別使用單一演算法決策樹、羅吉斯迴歸、與組合式演算法羅吉斯葉，以錯誤歸類成本與召回率為評估指標來做模型比較，實證結果顯示羅吉斯葉之召回率最高(訓練集 66.67%、驗證集 66.67%、測試集 66.67%)且錯誤歸類成本最低(訓練集 132、驗證集 50、測試集 52)，顯示羅吉斯葉之預測能力優於單一演算法，並且提供之流失客戶訊息亦較單一演算法更細膩，利於後續制定相關保留策略。整體來看，顧客基本資料變數在單一演算法中，對預測客戶流失並無顯著效果，但在羅吉斯葉中卻扮演著重要的客戶分群變數，透過決策樹利用背景條件作分群(子葉)，找出同質性高的客戶，再透過羅吉斯迴歸利用 RFM 以及時空行為衍生變數提升客戶流失預測能力並挖掘出更加精準的客戶特性。

理論意涵

農藥行的流失現象比較適用 **B2C** 的行為變數來解釋

顧客行為變數、時空與選擇行為變數無論在單一演算法或羅吉斯葉皆對預測客戶流失皆具顯著效果，表示客戶的交易行為是影響流失與否之關鍵影響因素；而供應商評選滿意度對客戶流失的預測模型並不顯著，代表描述農藥行客戶是否流失，**B2B** 的解釋變數並無法增加預測能力，亦即供應商評選準則上之表現並非影響農藥行流失與否之關鍵因素。

方圓五百里的競爭者個數為農藥行是否流失的重要預測變數

過去文獻之顧客基本資料變數較少考量客戶的地理位置資訊，而本研究運用地理資訊系統計算經緯度並考量區域的概念，利用經緯度計算出客戶方圓 500 公尺內的同業數量，運用於羅吉斯葉，發現此變數為決策樹分群的一個重要變數。

管理意涵

為了防止客戶流失，廠商可以在收集相關資料後，首先以顧客基本資料作為決策樹之輸入變數，將客戶分群，再以顧客行為、時空與選擇行為及供應商評選滿意度作為羅吉斯迴歸之輸入變數建構客戶流失預測模型，在確認沒有過度配適的情況下完成模型建構。然後將現有的客戶資料集，輸入該預測模型，依照流失機率大小排序，找出最有可能流失的客戶進行客戶保留措施。

研究限制與未來研究建議

由於本研究樣本來自中部地區某中小企業所提供資料，並且以特定客戶類型為研究對象，樣本數較少，建議未來研究者以較大樣本數訓練模型，提升模型預測能力。另外，本研究採用電話訪問方式蒐集問卷，受訪者可能會受到其他因素干擾，造成訪問效果不佳，且研究對象有時隔一年以上未購買產品之客戶，故受訪者對於企業所提供的產品或服務認知，可能會有因時間落差而造成的誤差。另外，兩階段的羅吉斯葉分析方法，哪些變數適合在第一階段當決策樹的輸入變數，哪些變數適合在第二階段當羅吉斯迴歸的輸入變數可以作為未來的研究方向。

參考文獻

- 齊德彰、洪敘峰，2011。供應商評估模式建構之新概念－SEM 與 RST 之結合，*資訊管理學報*，第十八卷第二期，17-42。
- Ahmed, S. R., 2004. Applications of Data Mining in Retail Business. *Proceedings of International Conference on Information Technology: Coding and Computing*, Las Vegas, NV, USA.
- Alamsyah, A. and Salma, N., 2018. A Comparative Study of Employee Churn Prediction Model. *2018 4th International Conference on Science and Technology (ICST)*, 1, 1-4.
- Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J. and Anwar, S., 2019. Customer Churn Prediction in Telecommunication Industry Using Data Certainty. *Journal of Business Research*, 94, 290-301.
- Barfar, A., Padmanabhan, B. and Hevner, A., 2017. Applying Behavioral Economics in Predictive Analytics for B2B Churn: Findings from Service Quality Data. *Decision Support Systems*, 101, 115-127.
- Bose, I. and Chen, X., 2009. Quantitative Models for Direct Marketing: A Review from Systems Perspective. *European Journal of Operational Research*, 195(1), 1-16.
- Bounsaythip, C. and Rinta-Runsala, E., 2001. Overview of Data Mining for Customer Behavior Modeling. *VTT Information Technology Research Report*, (1), 1-53.
- Brown, B. P., Zablah, A. R., Bellenger, D. N. and Donthu, N., 2012. What Factors Influence Buying Center Brand Sensitivity? *Industrial Marketing Management*, 41(3), 508-520.
- Choi, T. Y. and Hartley, J. L., 1996. An exploration of supplier selection practices across the supply chain. *Journal of Operations Management*, 14(4), 333-343.
- Coussement, K., Lessmann, S. and Verstraeten, G., 2017. A Comparative Analysis of Data Preparation Algorithms for Customer Churn Prediction: A Case Study in the Telecommunication Industry. *Decision Support Systems*, 95, 27-36.

- Datta, S. and Das, S., 2015. Near-Bayesian Support Vector Machines for Imbalanced Data Classification with Equal or Unequal Misclassification Costs. *Neural Networks*, 70(2015), 39-52.
- De Caigny, A., Coussement, K. and De Bock, K. W., 2018. A New Hybrid Classification Algorithm for Customer Churn Prediction Based on Logistic Regression and Decision Trees. *European Journal of Operational Research*, 269(2), 760-772.
- Dickson, G. W., 1966. An Analysis of Vendor Selection Systems and Decisions. *Journal of Purchasing*, 2(1), 5-17.
- Figalíst, I., Elsner, C., Bosch, J. and Olsson, H. H., 2019. Customer Churn Prediction in B2B Contexts. *The International Conference on Software Business*, Jyväskylä, Finland.
- Fitzsimmons, J. A. and Fitzsimmons, M. J., 2000. *Service Management: Operations Strategy and Information(3 rd.)*. New York: Mc Graw-Hill.
- Giraud-Carrier, C. and Povel, O., 2003. Characterising Data Mining Software. *Intelligent Data Analysis*, 7(3), 181-192.
- Gordini, N. and Veglio, V., 2017. Customers Churn Prediction and Marketing Retention Strategies. An Application of Support Vector Machines Based on the Auc Parameter-selection Technique in B2B E-commerce Industry. *Industrial Marketing Management*, 62, 100-107.
- Gustafsson, A., Johnson, M. D. and Roos, I., 2005. The Effects of Customer Satisfaction, Relationship Commitment Dimensions, and Triggers on Customer Retention. *Journal of Marketing*, 69(4), 210-218.
- Hallowell, R., 1996. The Relationships of Customer Satisfaction, Customer Loyalty, and Profitability: An Empirical Study. *International Journal of Service Industry Management*, 7(4), 27-42.
- Hansen, H., Samuelson, B. M. and Sallis, J. E., 2013. The Moderating Effects of Need for Cognition on Drivers of Customer Loyalty. *European Journal of Marketing*, 47(8), 1157-1176.
- Heeringa, S., Weat, B. and Berglund, P., 2010. *Applied Survey Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC.
- Hsu, C. C., Kannan, V. R., Keong Leong, G. and Tan, K. C., 2006. Supplier Selection Construct: Instrument Development and Validation. *The International Journal of Logistics Management*, 17(2), 213-239.
- Jahromi, A. T., Stakhovych, S. and Ewing, M., 2014. Managing B2B Customer Churn, Retention and Profitability. *Industrial Marketing Management*, 43(7), 1258-1268.
- Kaya, E., Dong, X., Suhara, Y., Balcisoy, S., Bozkaya, B. and Pentland, A. S., 2018. Behavioral Attributes and Financial Churn Prediction. *EPJ Data Science*, 7(1), 1-18.
- Koh, H. C. and Gerry, C. K. L., 2002. Data Mining And Customer Relationship Marketing In The Banking Industry. *Singapore Management Review*, 24(2), 1-27.
- Lam, S. Y., Shankar, V., Erramilli, M. K. and Murthy, B., 2004. Customer Value, Satisfaction, Loyalty, and Switching Costs: An Illustration from a Business-to-Business Service Context. *Journal of the Academy of Marketing Science*, 32(3), 293-311.
- Maestrini, V., Luzzini, D., Caniato, F. and Ronchi, S., 2018. Effects of Monitoring and Incentives on Supplier Performance: An Agency Theory Perspective. *International Journal of Production Economics*, 203, 322-332.
- McCann, P. and Fingleton, B., 1996. The Regional Agglomeration Impact of Just-I-Time Input Linkages: Evidence from the Scottish Electronics Industry. *Scottish Journal of Political Economy*, 43(5), 493-518.
- Miguéis, V. L., Van den Poel, D., Camanho, A. S. and Falcão e Cunha, J., 2012. Modeling Partial Customer Churn: On the Value of First Product-category Purchase Sequences. *Expert Systems with Applications*, 39(12), 11250-11256.
- Mitrovic, S., Singh, G., Baesens, B., Lemahieu, W., De Weerd, J., 2017. Scalable RFM-enriched Representation Learning for Churn Prediction,

- Proceedings of 2017 International Conference on Data Science and Advanced Analytics*, Tokyo, Japan, 79-88.
- Moeyersoms, J. and Martens, D., 2015. Including High-cardinality Attributes in Predictive Models: A Case Study in Churn Prediction in The Energy Sector. *Decision Support Systems*, 72, 72-81.
- Ngai, E. W. T., Xiu, L. and Chau, D. C. K., 2009. Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification. *Expert Systems with Applications*, 36(2), 2592-2602.
- Ogino, S., 2017. The Standout E-Commerce Customer Loyalty Platform Stats of 2017. Retrieved from <https://www.annexcloud.com/blog/standout-customer-loyalty-stats-2017/>
- Qabbaah, H., Sammour, G. and Vanhoof, K., 2017. Data Mining Techniques to Improve the Response Rate of E-mail campaigns and Customer Loyalty. *The International Conference of Technology Innovation, Management and Entrepreneurship*, Amman.
- Russo, I. and Confente, I., 2017. *Customer Loyalty and Supply Chain Management: Business-to-Business Customer Loyalty Analysis*. England, UK: Routledge.
- Sabbeh, S. F., 2018. Machine-Learning Techniques for Customer Retention: A Comparative Study. *International Journal of Advanced Computer Science and Applications*, 9(2), 273-281.
- Schroeder, R. G., 2000. *Operations Management: Contemporary Concepts and Cases*. New York, NY.: McGraw-Hill/Irwin.
- Sharma, R. R. and Sachdeva, R., 2017. Evaluating Prediction of Customer Churn Behavior Based on Artificial Bee Colony Algorithm. *International Journal Of Engineering and Computer Science*, 6(1), 20017-20021.
- Sheu, J.-B., 2007. A Hybrid Fuzzy-optimization Approach to Customer Grouping-based Logistics Distribution Operations. *Applied Mathematical Modelling*, 31(6), 1048-1066.
- Stalk, G., 1988. Time-the Next Source Of Competitive Advantage. *Harvard Business Review*, July-August, 41-51.
- Swani, K., Brown, B. P. and Milne, G. R., 2014. Should Tweets Differ for B2B and B2C? An Analysis of Fortune 500 Companies' Twitter Communications. *Industrial Marketing Management*, 43(5), 873-881.
- Upton, D. M., 1995. What Really Makes Factories Flexible? *Harvard Business Review*, July-August, 74-84.
- Vadakattu, R., Panda, B., Narayan, S. and Godhia, H., 2015. Enterprise Subscription Churn Prediction. 2015 *IEEE International Conference on Big Data (Big Data)*, 1317-1321.
- Van den Poel, D. and Larivière, B., 2004. Customer Attrition Analysis for Financial Services Using Proportional Hazard Models. *European Journal of Operational Research*, 157(1), 196-217.
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J. and Baesens, B., 2012. New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach. *European Journal of Operational Research*, 218(1), 211-229.
- Wania, A., Kühn, I. and Klotz, S., 2006. Plant Richness Patterns in Agricultural and Urban Landscapes in Central Germany- Spatial Gradients of Species Richness. *Landscape and Urban Planning*, 75(1-2), 97-110.
- Weiss, G. M., McCarthy, K. and Zabar, B., 2007. Cost-Sensitive Learning vs. Sampling: Which is Best for Handling Unbalanced Classes with Unequal Error Costs? *The Proceedings of the International Conference on Data Mining*, Las Vegas, Nevada, USA.
- Wu, M.-Y. and Weng, Y.-C., 2010. A Study of Supplier Selection Factors for High-tech Industries in the Supply Chain. *Total Quality Management & Business Excellence*, 21(4), 391-413.
- Zeng, F., Yang, Z., Li, Y. and Fam, K.-S., 2011. Small Business Industrial Buyers' Price Sensitivity: Do Service Quality Dimensions Matter in Business Markets? *Industrial Marketing Management*, 40(3), 395-404.

Zhu, B., Baesens, B. and vanden Broucke, S. K., 2017.

An Empirical Comparison of Techniques for the Class Imbalance Problem in Churn Prediction.

Information Sciences, 408, 84-99.

Zumel, N. and Mount, J., 2014. *Practical Data Science*

with R. Shelter Island, NY: Manning Publications

Co

附錄

附表 1 自變數計算說明

變數種類	變數名稱	變數代碼	說明	類型
顧客 基本資料	資本額	CapitalAmount	用客戶名稱及地址查詢財政部稅籍登記資料平台。	連續
	成立年數	YearsFounded	以財政部資料平台公開之成立時間，以區間最後一日為基準點推算。	連續
	地區	AreaCode	以台灣之北、中、南、東所隸屬之縣市定義。	類別
	方圓 500 公尺農藥行數	Area_500m	方圓 500 公尺內同為企業客戶的農藥行家數。	連續
	經度	XCOO	利用地理資訊軟體將地址轉換為經緯度。	連續
	緯度	YCOO	利用地理資訊軟體將地址轉換為經緯度。	連續
顧客行為	總交易次數	Frequency_Sum	客戶於交易期間內交易的總次數	連續
	平均交易金額	Monetary_Ave	總交易金額除以交易期間天數。	連續
	最後一年總交易次數	Frequency_3	交易期間最後一年為期計算之總交易次數。	連續
	最後一年平均交易金額	Monetary_Ave_3	交易期間最後一年為期計算之平均交易金額。	連續
	RFM 總和	RFM_Sum	R、F、M 之加總。	連續
時空與 選擇行為變數	短期多樣性(旬)	STC_Month_DS	使用 Kaya et al. (2018)的時間多樣性，以旬作為時間種類，以交易期間最後 12 個月為計算範圍。	連續
	長期多樣性(旬)	STC_Month_DL	使用 Kaya et al. (2018)的時間多樣性，以旬作為時間種類，以交易期 30 個月為計算範圍。	連續
	短期集中度(旬)	STC_Month_LS	使用 Kaya et al. (2018)的時間集中度，以旬作為時間種類，以交易期間最後 12 個月為計算範圍。	連續
	長期集中度(旬)	STC_Month_LL	使用 Kaya et al. (2018)的時間集中度，以旬作為時間種類，以交易期 30 個月為計算範圍。	連續
	規律性(旬)	STC_Month_R	使用 Kaya et al. (2018)的時間規律性。	連續
供應商評選 滿意度	服務品質	PC_Service	對四個問項進行探索性因素分析。	連續
	品質績效	StandMean_Quality	分別對兩個問項進行標準化後平均。	連續
	交期績效	StandMean_Delivery	分別對兩個問項進行標準化後平均。	連續
	彈性績效	PC_Flexibility	對四個問項進行探索性因素分析。	連續
	價格	StandMean_Price	分別對兩個問項進行標準化後平均。	連續

林金賢為國立中興大學企業管理學系教授，美國華盛頓州立大學管理博士。研究領域為決策科學、策略配適與資料探勘等。學術論文曾發表於 Health Care Management Review、Computer and Operations Research、Networks、會計評論、管理與系統、管理學報、以及國內外重要學術期刊與研討會等。

Chin-Shien Lin is a professor of Department of Business Administration, National Chung Hsing University. He completed his Ph.D. Degree at Department of Management and Systems, Washington State University. His research areas include decision science, strategic fit and data mining. His papers have been published at Health Care Management Review, Computer and Operations Research, Networks, International Journal of Accounting Studies, Journal of Management & Systems, Journal of Management, and some other prestigious Journals and conference proceedings.

謝欣樺為國立中興大學企業管理學系碩士班畢業生。主要研究領域為顧客關係管理與資料探勘。

Hsin-Hua Hsieh is a graduate student of the Department of Business Administration, National Chung Hsing University. Her main research areas include Customer Relationship Management and Data Mining.