

企業客戶語音業務流失預測之研究

The Research on the Prediction of Enterprise Customer Churn on Voice Services

黃三益 *San-Yih Hwang*

國立中山大學資訊管理系

Department of Information Management,

National Sun Yat-sen University

林彥君 *Yen-Chun Lin*

國立中山大學資訊管理研究所

Department of Information Management,

National Sun Yat-sen University

賴佳瑜* *Chia-Yu Lai*

國立屏東科技大學資訊管理系

Department of Management Information Systems,

National Pingtung University of Science and Technology

本文引用格式建議：黃三益、林彥君、賴佳瑜，2022，企業客戶語音業務流失預測之研究，中山管理評論，30 卷 2 期：293~323。

Suggested Citation: Hwang, S. Y., Lin, Y. C., and Lai, C. Y., 2022, "The Research on the Prediction of Enterprise Customer Churn on Voice Services," **Sun Yat-sen Management Review**, Vol. 30, No. 2, 293-323.

* 通訊作者:賴佳瑜。地址:屏東縣內埔鄉老埤村學府路 1 號資訊管理學系; 聯絡電話: 08-7703202 # 7907; Email: chiayulai@mail.npust.edu.tw

摘要

網際網路技術成熟及行動網路速度提升，使得電信業者提供之語音服務被網路服務取代。以營收貢獻來說，企業客戶遠大於一般消費客戶，且若企業客戶部分業務一旦流失至競業，其他相關業務皆可能轉至競業，期能保留目前使用語音服務的客戶，維持現有客戶之語音服務營收。本研究主題為企業客戶語音流失預測，為找出可能流失客戶，運用機器學習之方法建立預測模型，並考量企業客戶特性，納入企業客戶特有變數，以提升預測準確率。根據真實的電信客戶資料的研究結果顯示，企業客戶特有變數在前十大重要變數裡佔了三個；此外，依據預測結果對可能流失之客戶進行挽留，透過設定最適合閾值後，使營收損失最小化，相較於單純不作為，則可有效降低營收損失。

關鍵詞：電信業客戶流失、客戶流失預測、客戶固守、機器學習、行銷預測

Abstract

Due to the maturity of IP network technology and the provision of 4G mobile high speed networks, Internet-based services have become more popular. Nevertheless, the revenue of voice services for telecom operators has been substantially reduced. Yet the construction cost of broadband network and mobile phone base station remain the same. As a result, the profit of telecom operators has been drastically reduced. In addition, reports from the NCC shows that Taiwan's telecommunications market has been saturated. Therefore, customer retention and customer churn management become important issues for telecom operators.

In this work, we engage in the study of predicting enterprise customer churns in telecommunication industry because enterprise customers contribute more revenues to telecom service providers. Various variables, including the enterprise customers' unique variables, have been identified, and the Xgboost algorithm is used to establish the prediction model. Our experimental results based on the real telecom customer data show that three enterprise customers' unique variables are among the top 10 most important variables. In addition, our proposed prediction model is able to increase AUC and recall rate by 3% and 5.4% respectively, when compared to the prediction model that simply incorporates variables identified by previous work. We

further try to minimize revenue loss by setting the most appropriate threshold. The experimental results show that by setting the threshold at 0.72 and applying customer retention strategy to the predicted customers, we are able to reduce the revenue loss by 525 units per customer.

Keywords: Customer Churn in Telecom, Customer Retention, Machine Learning, Churn Prediction, Marketing

壹、緒論

近年來因網際網路與資訊科技的迅速發展，使得客戶通信行為產生巨大的轉變，傳統的通訊方式(如撥打電話)已無法滿足客戶需求，客戶追求更即時、方便、可無限使用及帶有附加價值的通訊方式，傳統通訊方式逐漸式微，且因行動網路速度大幅提升，及應用於網際網路的服務日益普及，如 YouTube、Skype、Line 等，昔日人際間的連絡以市內電話或行動電話為主，現今則為透過高速網路加免費社群通訊軟體，即可達成溝通互動的目的，例如視訊或語音聊天、傳送文字訊息或圖片、多人聊天及開會等，客戶每月僅須支付定額網路月租費，即可無限使用。

因此，語音服務被網路應用程式取代已是無可逆轉的趨勢，然而以電信業者的立場並不樂見此現象，因從語音服務轉換至網路服務為減價轉換，語音服務是以量計價，隨著客戶使用量增加，其消費總額亦會增加，而網路服務是定額計價，客戶消費額度並不因其使用量而變動，電信業者可獲取之營收相對減少，然而投注在寬頻網路及行動電話基地台建設的成本卻是不變，因而使得電信業者獲利降低；再者，NCC（2016）公布之我國電信業者營運實績統計數據顯示，國內電信業者的用戶總數不再成長，其中市內電話及行動電話用戶數卻呈現衰減的趨勢，台灣電信市場顯然已飽和。種種現象突顯了電信業者的困境，中華電信、台灣大哥大、遠傳電信等三大台灣電信業者及其餘小型業者困在紅海市場內激烈競爭，如何擺脫這樣的困境，並且順利持續成長獲利，是所有電信業者目前最關注的重要議題。

電信客戶依申租身份分為兩類，一類為消費客戶，另一類則為企業客戶；消費客戶租用電信業務特性是申租數量少，喜愛行動電話勝於市內電話，消費

客戶使用行動電話連絡、社群互動及娛樂；企業客戶係指公司行號、學校、政府機構...等，租用多種電信業務且數量眾多，其租用市內電話及行動電話主要是提供給員工連絡溝通用。從國內某家大型電信公司資料來看，其語音服務年收入衰退約 14%，其中消費客戶衰退 18%，企業客戶衰退 5%，由此可知相對於消費客戶，企業客戶仍是比較仰賴使用語音服務。而與消費客戶相比，企業客戶貢獻度更高，現今電信業者多以整合性方案網綁企業客戶，希望能固守企業客戶，倘若企業客戶的部份業務流失至競業，往後之所有業務皆可能陸續轉至競業，因此本研究主題為電信業者企業客戶流失預測，並專注在語音服務，希望從語音服務流失預測，及早發現客戶流失意圖，進而鞏固客戶，達成客戶保留的目的。

過去電信業務流失預測之相關研究對象多為一般消費客戶，最近一項根據 Kaggle 公開資料所進行的美國電信業客戶流失之研究顯示 (Kapoor, 2017)，電信設備的老舊程度、客戶每月消費額、客戶任期長度、近三個月的使用量，以及手機的價錢為前五大因素。本研究則是著重在企業客戶流失預測，因此在蒐集變數的過程中，本研究在考量企業客戶特性後，彙集消費客戶預測之重要變數，並依其合適程度進行篩選，最後留下 15 個變數作為基準變數，並加上企業客戶特有變數 23 個。研究結果顯示，納入企業客戶特有變數可有效提升預測的準確度；此外依據預測結果對於可能流失客戶要進行挽留，透過設定最適合閾值後，可以使營收損失最小化，比單純不作為更可有效降低營收損失。

本文接下來進行文獻探討，羅列電信業客戶流失預測之相關研究；第三節說明研究方法，包括資料來源及定義、資料前置處理、羅吉斯迴歸分析、整體學習的 Xgboost 分析及整體學習的各種演算法建立預測模型；第四節實驗結果說明本研究之預測結果，並計算預測流失客戶且採取固守客戶方案的最適合閾值(Threshold)，以其達到營收損失最小化；第五節針對實驗結果提出結論及未來研究建議。

貳、文獻探討

面對日益競爭的全球市場和商業環境，客戶流失管理為企業經營是首要議題，及早偵測客戶流失的可能性，才能使客戶流失率降至最低。企業若可減少客戶流失率，不僅能大幅減少招攬新客戶的成本，並可增加企業總體盈利。根據 Reichheld & Sasser (1990) 研究顯示企業若能降低 5% 的顧客流失率，除了

可提升顧客的忠誠度，更可使企業獲利率由 25% 增加至 85%。Kotler et al. (2014) 亦指出，若可有效保留現有的客戶，則其獲利率將為吸引新客戶之十六倍。可知顧客流失對於企業獲利的影響之大，客戶的去留將會決定企業的存活的關鍵因素。因而現今許多企業盡力於提升顧客保留率，同時也降低顧客流失率，以利增進企業營收成長率。然而，企業為了在競爭激烈的環境中生存與增加獲利，花費大量成本與行銷，推出了許多新方案來吸引新客戶，卻往往忽略現有顧客群，以致企業多只針對已流失的顧客進行事後抽樣與調查，對已流失的客戶的消費方式與行為作事後分析與研究，但事後的分析卻往往早已失去時效性。因此，過去十年來對客戶管理與建置流失客戶系統預測模型已有許多的研究，可用來預測客戶可能的流失行為，及早發現客戶可能流失的原因，提供給企業加以運用，則可有效避免客戶流失的行為產生，減低客戶流失所產生的損失 (Neslin et al., 2006; Verbeke et al., 2011; Peppers & Rogers, 2016)。

一、客戶流失定義

客戶流失(Customer churn)可稱為客戶轉換(Customer turnover)，或是客戶偏離(Customer defection)的行為，廣義而言是客戶與公司終止關係，客戶不再使用原有的服務提供者的服務，或是終止原有的服務的行為。依據不同行業別，對於客戶流失的定義也不同，而亦有研究指出，顧客流失的行為是顧客的忠誠度下降所帶來的行為。客戶流失管理是企業重要的生存議題，企業所追求的目標即是最大化顧客保留率，並極小化顧客流失，以獲取最大利益。客戶流失率為企業經營之警訊，若流失率上升，則表示有更多客戶欲終止關係，企業應調整市場策略或改變客戶服務流程。以電信業者而言，根據 Berson & Smith (2002) 研究定義電信產業中的客戶流失表示現有客戶從原有業者移轉至其他電信業者的行為，而衍生出客戶的流失管理為如何維繫重要客戶的管理。許多研究指出電信業者的一般客戶流失率相對率非常高，可高達46% (Neslin et al., 2006)。因此如何管理電信業的客戶流失與探討相關流失原因現已有許多相關研究文獻 (Berson & Smith, 2002; Verbeke et al., 2012; Premkumar & Rajan, 2017)。例如可透過顧客的通話習性，主要包含平均通話時間、平均通話次數、通話對象等來了解流失客戶的因素。例如，Mattison (2006) 從不同訪談者、調查、目標群及其他不同形式的研究之發現彙整流失原因，並將其分為兩大類：非自主性流失及自主性流失。非自主性流失包含詐欺(Fraud)、不繳款(non-payment)、與低使用率(Underutilization)。自主性流失又分為兩類：(一)深思熟慮型(Deliberate)，包

含科技、經濟、品質、社會／心理、方便等因素，(二)附帶型(Incidental)，包含財務、地點等因素。本研究針對企業客戶的自主性流失，另外因企業客戶自主性流失包含轉換至其他競業及終止租用服務等兩種現象，受限於可取得資料無法明確區分這兩種現象，因此在本研究合併視為同一現象，不做區隔。

二、電信業流失客戶預測之相關研究

由於客戶的去留將決定企業之生存，客戶流失問題是一個非線性且複雜的分類問題，因此許多文獻已提出各種方法來進行相關研究，客戶流失預測除了可提前預測潛在客戶的流失，防止客戶流失產生的巨大損失，另可加以探討影響顧客流失之相關變數外，並擬定客戶流失因應與預防的參考，以提升企業競爭力並創造最大之獲利。目前現有文獻多利用資料探勘、統計分析或其他智慧型機器學習演算等方法來預測客戶流失的可能性，在電信業建立預測客戶流失的模型之研究已非常廣泛。整體而言，電信業者客戶流失預測之相關研究可涵蓋流失客戶的預測，識別流失原因分析與進行客戶保留策略等三個構面，但現今研究著重在提高預測模型的準確度，目前大多學者都採用單一數據集，比較各種不同演算法之準確度以找出最佳化之模型，然而針對客戶流失原因分析與客戶保留策略的研究探討則相對少數 (Verbeke et al., 2011)。其中在電信業流失客戶預測具有代表性之相關研究如下：

Wei & Chiu (2002) 使用決策樹及多專家分類器(Multi-classifier class-combiner approach)進行預測，以通聯記錄時間序列資料進行客戶流失預測，研究結果(一)客戶流失前通話行為異於平常的變化，愈接近退租日愈趨明顯，因此預留前置期間愈短，模型預測能力愈佳，(二)分析資料擷取的觀測期間不宜分段過多，分段後的比較時段天數變少，前後時段通話行為變化愈趨不明顯，不同用戶間變異會擴大，因此觀測期間分段後之比較時段天數不宜短於 10 天。

Au et al. (2003) 使用電信服務提供商的服務線路 (Service line)數據進行流失客戶預測，蒐集 3 個月內服務終止與否的資料，針對其選定年份回推 6 年期間的相關資訊，包含服務類型、服務租用、市場資訊、客戶人口特性及服務線路交易歷史。研究結果顯示影響客戶流失之重要變數包含帳戶類型、客戶區隔、用戶所在區域、過去六個月使用量、自 1993 年以來活躍與否。

Neslin et al. (2006)提出比較各種方法以找出最準確的客戶流失預測模型與建構挽救利潤函數，利用電信資料應用在類神經網路，決策樹與羅吉斯回歸等模型來預測電信客戶流失率，並比較準確度，其研究結果為應用的方法的準確率為選擇模型之重要影響因子，並基於流失機率，流失成本等參數建構挽救利

潤最大函數與數學模型，在流失預警系統指出高危險流失群之後，進行挽留政策，以利企業透過客戶流失預測與挽救策略，獲取最大的利益。

Keramati et al. (2014) 採用伊朗電信公司數據，先是分別使用四種分類技術，分別是類神經網路 (Artificial Neural Network, ANN)、決策樹 (Decision Tree)、最近鄰居法 (K-Nearest Neighbors, KNN) 及支援向量機 (Support Vector Machine, SVM)，類神經網路明顯優於另外三種，接著該研究提出一種混合方法，綜合前述四種技術，在精確率 (Precision Rate) 和召回率 (Recall Rate) 方面均獲得超過 95% 的準確性；該研究結果影響客戶流失之重要變數為使用頻率、抱怨次數、第二次使用。

Vafeiadis et al. (2015) 採用 UCI 數據資料庫中的電信客戶的公開資料集，比較許多現行的熱門演算法例如多層類神經網路 (Artificial Neural Network, ANN)、決策樹 (Decision Tree)、支援向量機 (Support Vector Machine, SVM)、單純貝式分類器 (Naïve Bayes) 與 Adaboost 等應用在客戶流失預測方面的效果，其中以 Adaboost 的整體學習演算法的表現最好，其最佳值可以達到 97% 的準確率且 F 數值平均有超過 84%。

Hassouna et al. (2016) 使用英國某家電信公司資料，預測可能流失之消費客戶，資料分為兩組共 17 個變數，兩組資料均是由 50% 流失客戶及 50% 未流失客戶組成，其客戶合約效期包含 12 個月及 18 個月，變數內容包含人口特性、合約效期、數據使用量、語音使用量、手機維修次數及客訴次數等，研究結果顯示影響客戶流失之重要變數包含合約長度、語音訊務量變化、客戶任期、手機維修次數、涵蓋率及資料傳輸量變化。

Jamil & Khan (2016) 則針對客戶流失原因分析在電信業的領域做更深入的探討，此研究認為預測模型的準確度並非企業所最關注的，更重要的是找出客戶流失的重要因素，此研究運用自動學習的方法 (Active Learning Based Approach) 建構在支援向量機 (Support Vector Machine, SVM) 與決策樹 (Decision Tree) 的客戶流失預測模型，利用規則方法的模型不僅可擁有預測模型的準確度，且具有可解讀性與可判斷性之模型。其中模型準確率更可高達 94%，更重要的是可找出重要客戶流失的原因，讓企業可提早因應，進行保留客戶的策略。

上述研究所針對的客戶均為消費客戶，本研究的主要對象則為企業客戶，因現行文獻中尚未有針對企業客戶之研究，故將取用其提及之重要變數及重要結論，並加入一些企業客戶專有的變數，納入預測模型。

參、研究方法

一、資料來源及定義

本研究原始資料來自國內某家大型電信公司，研究對象為主要據點落在中部、南部及東部之企業客戶共 212,644 家(非流失企業客戶 208,283 家，流失企業客戶 4,361 家)，且其 105 年第三季以前仍租用市內電話門號或行動電話門號者，預測 105 年第四季市內電話業務與行動電話業務流失可能性。

與多數研究不同的是資料範圍同時包含市內電話業務及行動電話業務，參考過往實務經驗，企業客戶租用之市內電話及行動電話會有相互取代情形，該電信公司在實務面之行銷產品包裝亦是提供市話電話業務及行動電話業務之整合優惠包裝，以滿足企業客戶需求。

本研究之資料集時間範圍(Time window of data sets)：以 105 年 Q1~105 年 Q3 為資料集時間範圍，以 105 年 Q4 判斷客戶是否流失。(詳如圖 1)

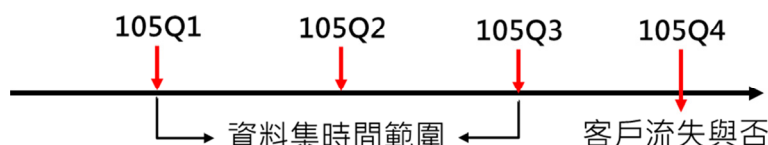


圖 1：本研究資料集時間範圍(Time Window of data sets)

資料來源:本研究整理

二、變數選擇

本研究取用變數總共 40 個，依變數取用來源主要分為三類，分別是「基本變數」、「一般變數」及「企業客戶變數」，說明如下 (詳如表 1)：

1.基本變數 (2 個)：相對於消費客戶流失預測模型取用之人口特性變數(如年齡、性別、居住區域等)，訂定企業客戶特性變數，包含企業客戶資本額及企業客戶產業別，作為本研究之基本變數。

2.一般變數 (15 個)：彙整相關文獻提及有關一般客戶之重要變數，根據可取得情況及本研究方向挑選變數，並當作基準變數(Baseline)，主要包含企業客戶服務等級、企業客戶主要據點、任期、市話及行動訊務量變化。

3.企業客戶變數 (23 個)：依據企業客戶特性新增數個變數，包含是否有專人服務、企業客戶前半年異動情形、是否申請優惠折扣、市話及行動門號數變

化、電信業務營收變化等。企業客戶變數主要是根據多年分析企業客戶特徵而選出，例如企業客戶租用門號數量會隨其性質及規模而有所不同，而門號數變化不一定與訊務量變化相關，某些客戶之門號僅做受話專用，因此除了納入訊務量變化外，同時也將門號數變化列為變數之一。

表 1：企業客戶流失預測取用變數

變數取用來源	變數名稱	資料型態	描述
基本變數	企業客戶產業別	Character	包含工程建築、公部門機構、製造及資訊業、醫療業、服務貿易、交通運輸、其他等七項
基本變數	企業客戶資本額	Double	企業客戶設立公司時於經濟部登記之資本額資料
一般變數	企業客戶服務等級	Character	依月營收貢獻度劃分為五種
一般變數	企業客戶主要據點	Character	分為中部、南部、東部
一般變數	企業客戶任期	Integer	客戶來到本公司的月數
一般變數	市話訊務量_105Q3較前季成長率	Double	市話業務_105年7~9月訊務量較105年4~6月訊務量增減比例
一般變數	市話訊務量_105Q2較前季成長率	Double	市話業務_105年4~6月訊務量較105年1~3月訊務量增減比例
一般變數	市話訊務量_105Q3較104同季成長率	Double	市話業務_105年7~9月訊務量較104年7~9月訊務量增減比例
一般變數	市話訊務量_105Q2較104同季成長率	Double	市話業務_105年4~6月訊務量較104年4~6月訊務量增減比例
一般變數	行動訊務量_105Q3較前季成長率	Double	行動業務_105年7~9月訊務量較105年4~6月訊務量增減比例
一般變數	行動訊務量_105Q2較前季成長率	Double	行動業務_105年4~6月訊務量較105年1~3月訊務量增減比例
一般變數	行動訊務量_105Q3較104同季成長率	Double	行動業務_105年7~9月訊務量較104年7~9月訊務量增減比例
一般變數	行動訊務量_105Q2較104同季成長率	Double	行動業務_105年4~6月訊務量較104年4~6月訊務量增減比例
一般變數	市話行動訊務量_105Q3較前季成長率	Double	市話及行動業務_105年7~9月訊務量較105年4~6月訊務量增減比例
一般變數	市話行動訊務量_105Q2較前季成長率	Double	市話及行動業務_105年4~6月訊務量較105年1~3月訊務量增減比例
一般變數	市話行動訊務量_105Q3較104同季成長	Double	市話及行動業務_105年7~9月訊務量較104年7~9月訊務量增減比例
一般變數	市話行動訊務量_105Q2較104同季成長	Double	市話及行動業務_105年4~6月訊務量較104年4~6月訊務量增減比例
企業客戶變數	企業客戶類型	Character	依中小企業白皮書及營收貢獻度分級成四種，資本額8千萬元以上、資本額8千萬元以下&月貢獻9000元以上、資本額8千萬元以下&月貢獻9000元以下及政府/公營/學術
企業客戶變數	專人服務	Character	Y/N, 是否指派專人服務
企業客戶變數	前一季異動數量	Integer	105年7~9月市話及行動異動數量
企業客戶變數	前半年異動數量	Integer	105年4~9月市話及行動異動數量
企業客戶變數	前一季是否有異動	Character	Y/N, 105年7~9月是否有申裝/退租市話及行動
企業客戶變數	前半年是否有異動	Character	Y/N, 105年4~9月是否有申裝/退租市話及行動
企業客戶變數	優惠折扣申請	Character	Y/N, 105年是否申請優惠折扣
企業客戶變數	CT全區營收_105Q3較前季成長率	Double	電信業務_105年7~9月全區營收較105年4~6月全區營收增減比例
企業客戶變數	CT全區營收_105Q2較前季成長率	Double	電信業務_105年4~6月全區營收較105年1~3月全區營收增減比例
企業客戶變數	CT全區營收_105Q3較104同季成長率	Double	電信業務_105年7~9月全區營收較104年7~9月全區營收增減比例
企業客戶變數	CT全區營收_105Q2較104同季成長率	Double	電信業務_105年4~6月全區營收較104年4~6月全區營收增減比例
企業客戶變數	市話門號數_105Q3較前季成長率	Double	市話業務_105年7~9月門號數較105年4~6月門號數增減比例
企業客戶變數	市話門號數_105Q2較前季成長率	Double	市話業務_105年4~6月門號數較105年1~3月門號數增減比例
企業客戶變數	市話門號數_105Q3較104同季成長率	Double	市話業務_105年7~9月門號數較104年7~9月門號數增減比例
企業客戶變數	市話門號數_105Q2較104同季成長率	Double	市話業務_105年4~6月門號數較104年4~6月門號數增減比例
企業客戶變數	行動門號數_105Q3較前季成長率	Double	行動業務_105年7~9月門號數較105年4~6月門號數增減比例
企業客戶變數	行動門號數_105Q2較前季成長率	Double	行動業務_105年4~6月門號數較105年1~3月門號數增減比例
企業客戶變數	行動門號數_105Q3較104同季成長率	Double	行動業務_105年7~9月門號數較104年7~9月門號數增減比例
企業客戶變數	行動門號數_105Q2較104同季成長率	Double	行動業務_105年4~6月門號數較104年4~6月門號數增減比例
企業客戶變數	市話行動門號數_105Q3較前季成長率	Double	市話及行動業務_105年7~9月門號數較105年4~6月門號數增減比例

企業客戶語音業務流失預測之研究

變數取用來源	變數名稱	資料型態	描述
基本變數	企業客戶產業別	Character	包含工程建築、公部門機構、製造及資訊業、醫療業、服務貿易、交通運輸、其他等七項
基本變數	企業客戶資本額	Double	企業客戶設立公司時於經濟部登記之資本額資料
一般變數	企業客戶服務等級	Character	依月營收貢獻度劃分為五種
一般變數	企業客戶主要據點	Character	分為中部、南部、東部
一般變數	企業客戶任期	Integer	客戶來到本公司的月數
一般變數	市話訊務量_105Q3較前季成長率	Double	市話業務_105年7~9月訊務量較105年4~6月訊務量增減比例
一般變數	市話訊務量_105Q2較前季成長率	Double	市話業務_105年4~6月訊務量較105年1~3月訊務量增減比例
一般變數	市話訊務量_105Q3較104同季成長率	Double	市話業務_105年7~9月訊務量較104年7~9月訊務量增減比例
一般變數	市話訊務量_105Q2較104同季成長率	Double	市話業務_105年4~6月訊務量較104年4~6月訊務量增減比例
一般變數	行動訊務量_105Q3較前季成長率	Double	行動業務_105年7~9月訊務量較105年4~6月訊務量增減比例
一般變數	行動訊務量_105Q2較前季成長率	Double	行動業務_105年4~6月訊務量較105年1~3月訊務量增減比例
一般變數	行動訊務量_105Q3較104同季成長率	Double	行動業務_105年7~9月訊務量較104年7~9月訊務量增減比例
一般變數	行動訊務量_105Q2較104同季成長率	Double	行動業務_105年4~6月訊務量較104年4~6月訊務量增減比例
一般變數	市話行動訊務量_105Q3較前季成長率	Double	市話及行動業務_105年7~9月訊務量較105年4~6月訊務量增減比例
一般變數	市話行動訊務量_105Q2較前季成長率	Double	市話及行動業務_105年4~6月訊務量較105年1~3月訊務量增減比例
一般變數	市話行動訊務量_105Q3較104同季成長率	Double	市話及行動業務_105年7~9月訊務量較104年7~9月訊務量增減比例
一般變數	市話行動訊務量_105Q2較104同季成長率	Double	市話及行動業務_105年4~6月訊務量較104年4~6月訊務量增減比例
企業客戶變數	企業客戶類型	Character	依中小企業白皮書及營收貢獻度分級成四種，資本額8千萬元以上、資本額8千萬元以下&月貢獻9000元以上、資本額8千萬元以下&月貢獻9000元以下及政府/公營/學術
企業客戶變數	專人服務	Character	Y/N, 是否指派專人服務
企業客戶變數	前一季異動數量	Integer	105年7~9月市話及行動異動數量
企業客戶變數	前半年異動數量	Integer	105年4~9月市話及行動異動數量
企業客戶變數	前一季是否有異動	Character	Y/N, 105年7~9月是否有申裝/退租市話及行動
企業客戶變數	前半年是否有異動	Character	Y/N, 105年4~9月是否有申裝/退租市話及行動
企業客戶變數	優惠折扣申請	Character	Y/N, 105年是否申請優惠折扣
企業客戶變數	CT全區營收_105Q3較前季成長率	Double	電信業務_105年7~9月全區營收較105年4~6月全區營收增減比例
企業客戶變數	CT全區營收_105Q2較前季成長率	Double	電信業務_105年4~6月全區營收較105年1~3月全區營收增減比例
企業客戶變數	CT全區營收_105Q3較104同季成長率	Double	電信業務_105年7~9月全區營收較104年7~9月全區營收增減比例
企業客戶變數	CT全區營收_105Q2較104同季成長率	Double	電信業務_105年4~6月全區營收較104年4~6月全區營收增減比例
企業客戶變數	市話門號數_105Q3較前季成長率	Double	市話業務_105年7~9月門號數較105年4~6月門號數增減比例
企業客戶變數	市話門號數_105Q2較前季成長率	Double	市話業務_105年4~6月門號數較105年1~3月門號數增減比例
企業客戶變數	市話門號數_105Q3較104同季成長率	Double	市話業務_105年7~9月門號數較104年7~9月門號數增減比例
企業客戶變數	市話門號數_105Q2較104同季成長率	Double	市話業務_105年4~6月門號數較104年4~6月門號數增減比例
企業客戶變數	行動門號數_105Q3較前季成長率	Double	行動業務_105年7~9月門號數較105年4~6月門號數增減比例
企業客戶變數	行動門號數_105Q2較前季成長率	Double	行動業務_105年4~6月門號數較105年1~3月門號數增減比例
企業客戶變數	行動門號數_105Q3較104同季成長率	Double	行動業務_105年7~9月門號數較104年7~9月門號數增減比例
企業客戶變數	行動門號數_105Q2較104同季成長率	Double	行動業務_105年4~6月門號數較104年4~6月門號數增減比例
企業客戶變數	市話行動門號數_105Q3較前季成長率	Double	市話及行動業務_105年7~9月門號數較105年4~6月門號數增減比例
企業客戶變數	市話行動門號數_105Q2較前季成長率	Double	市話及行動業務_105年4~6月門號數較105年1~3月門號數增減比例
企業客戶變數	市話行動門號數_105Q3較104同季成長率	Double	市話及行動業務_105年7~9月門號數較104年7~9月門號數增減比例
企業客戶變數	市話行動門號數_105Q2較104同季成長率	Double	市話及行動業務_105年4~6月門號數較104年4~6月門號數增減比例

資料來源:本研究整理

三、變數的重要性(Importance of Features)

我們首先採用線性迴歸模式來衡量變數的重要性，線性迴歸的依變數通常為連續型變數，然而本研究結果產出之依變數為二元類別型變數(流失/未流失)，因此使用羅吉斯迴歸分析。

羅吉斯迴歸分析(Logistic Regression)

(一) p-Value

首先運用羅吉斯迴歸計算所有自變數 p-value，找出對依變數具有影響性的自變數，P-value 小於 0.05 的自變數如表 2 所示，與之前研究的結果雷同的是，業務量變化相關的自變數(包括門號數和營收變化)對於客戶流失有顯著的影響，此外，企業客戶服務等級、是否有專人服務、企業客戶產業別、企業客戶主要據點、是否申請優惠折扣、前半年是否有異動、企業客戶任期，以及企業客戶服務等級等自變數也有影響，相關說明如下：

表 2：羅吉斯迴歸-自變數 P-value

	Estimate	Std. Error	z value	Pr(> z)	
企業客戶服務等級D級	-1.49E+00	1.39E-01	-10.66	< 2e-16	***
企業客戶產業別_其他	-1.29E+00	8.49E-02	-15.24	< 2e-16	***
專人服務	5.24E-01	4.38E-02	11.97	< 2e-16	***
優惠折扣申請	7.41E-01	6.66E-02	11.12	< 2e-16	***
企業客戶任期	-1.46E-02	1.57E-03	-9.31	< 2e-16	***
企業客戶產業別_製造及資訊業	-3.25E-01	5.56E-02	-5.85	5.00E-09	***
企業客戶產業別_公部門機構	-6.59E-01	1.13E-01	-5.82	5.90E-09	***
前一季是否有異動	4.05E-01	7.24E-02	5.59	2.30E-08	***
企業客戶服務等級C級	-6.35E-01	1.14E-01	-5.55	2.80E-08	***
企業客戶產業別_交通運輸	-4.21E-01	7.97E-02	-5.28	1.30E-07	***
企業客戶服務等級E級	-1.70E+00	3.31E-01	-5.13	2.90E-07	***
企業客戶產業別_醫療業	-4.56E-01	9.07E-02	-5.03	4.90E-07	***
前半年是否有異動	2.86E-01	6.41E-02	4.46	8.30E-06	***
行動門號數_105Q2較前季成長率	-5.09E-01	1.19E-01	-4.27	1.90E-05	***
前半年異動數量	-1.38E-02	3.51E-03	-3.94	8.20E-05	***
市話門號數_105Q2較104同季成長率	1.11E-01	2.92E-02	3.79	0.00015	***
市話門號數_105Q2較前季成長率	-6.41E-01	1.77E-01	-3.63	0.00028	***
企業客戶產業別_服務貿易	-1.82E-01	5.59E-02	-3.25	0.00114	**
市話行動門號數_105Q3較前季成長率	-6.66E-01	2.18E-01	-3.06	0.00224	**
企業客戶主要據點_南部	-9.29E-02	3.15E-02	-2.95	0.00322	**
企業客戶主要據點_東部	-4.94E-01	1.71E-01	-2.89	0.00380	**
市話行動門號數_105Q3較104同季成長率	4.72E-02	1.63E-02	2.89	0.00388	**
市話門號數_105Q3較104同季成長率	-1.05E-01	4.04E-02	-2.59	0.00969	**
CT全區營收_105Q2較前季成長率	-5.95E-02	2.81E-02	-2.12	0.03436	*

資料來源:本研究整理

1. 企業客戶服務等級

依企業客戶月營收貢獻由大到小分為四種等級(A~D)及當年度新增客戶設為 E 級，提供不同等級之服務(例如針對 A 級客戶提供專人服務)。流失比例最高的是 A 級客戶(16.7%)，其月營收貢獻亦是五種等級中最高；流失家數最多的是 D 級客戶(2,921 家)，其月營收貢獻為五種等級中最低；根據實務上的觀察，月營收貢獻高的客戶通常大量地且多樣化地租用電信業務，其電信網路架構複雜，有時會同時租用多家電信業者之網路服務，相互備援，避免發生通訊中斷的狀況，因此影響該類型客戶採購意願的因子並非單純價格因素，而必須額外考慮網路架構是否相容、備援機制、通訊品質等。而月營收貢獻低的客戶通常只租用單一產品，大多為行動電話、市內電話或是寬頻網路，沒有複雜的網路架構，其採購意願大多受到價格影響。針對截然不同的兩種類型客戶，制定行銷策略時應有所差異。(詳如圖 2)

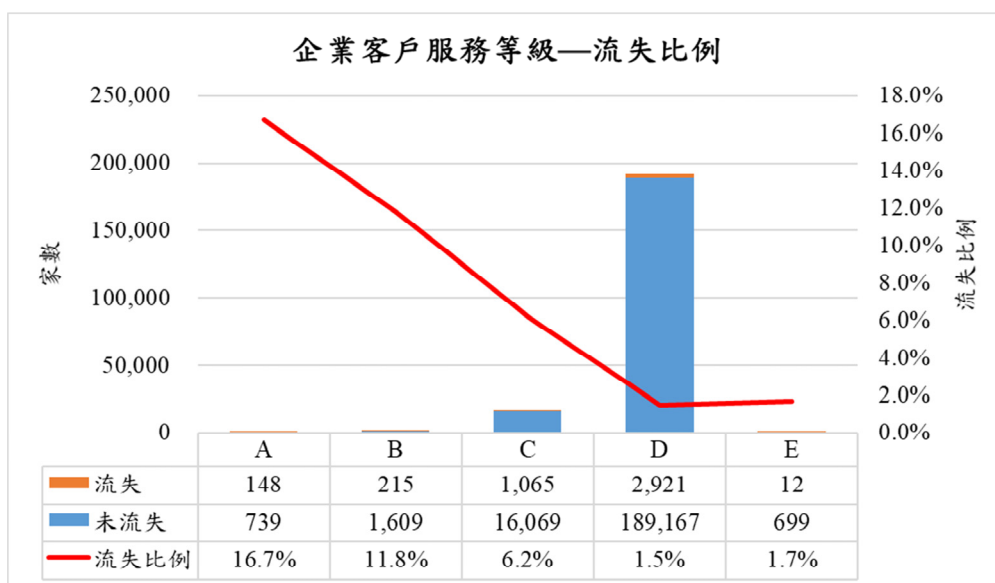


圖 2：企業客戶服務等級-流失比例

資料來源:本研究整理

2. 專人服務

未指派專人服務共 162,645 家，其中流失 2,154 家，流失比例 1.3%；指派專人服務共 49,999 家，其中流失 2,207 家，流失比例 4.4%。顯然指派專人服務之客戶仍有較大幅度流失比例，且其流失家數多於未指派專人服務之流失家數；由於客戶營收貢獻需達到一定程度，才會指派專人服務，推測可能因高貢獻度客戶為電信業者積極爭取對象，其流失可能性相較於未指派專人服務之客戶高；此外，專人服務流程及品質如果不佳，亦有可能導致客戶流失，因此在後續經營策略上，應著重在服務流程及品質的提升。(詳如圖 3)

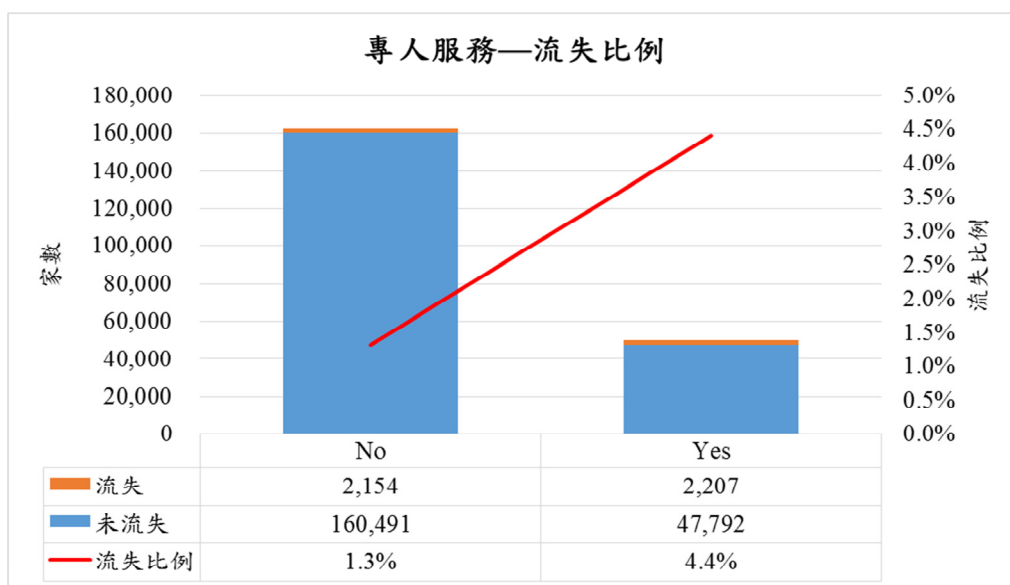


圖 3：專人服務-流失比例

資料來源:本研究整理

3. 企業客戶產業別

統計各產業別客戶流失情形，流失家數最多的是製造及資訊業客戶(1,484 家)，其流失比例為 2.4%，而工程建築業客戶之流失比例最高(3.3%)，但其流失家數(456 家)遠小於製造及資訊業客戶。根據實務上觀察到的現象，電信業者較為關注工程和製造及資訊業客戶，因其具有群聚特性，他們多半聚集在科學園區/工業區內，甚至在客戶營業規模大到一定程度時，其上下游合作夥伴會聚集在同一區域，以利製造流程順利進行。此類型客戶變成電信業者爭相搶

奪對象，電信業者通常會優先建設該區域之網路基礎設施，建設成本固定但網路利用率相較其他地方高，進一步可提升獲利。後續在建置流失預測模型時將企業客戶產業別列入變數之一，期望能有效提高預測精準度。(詳如圖 4)

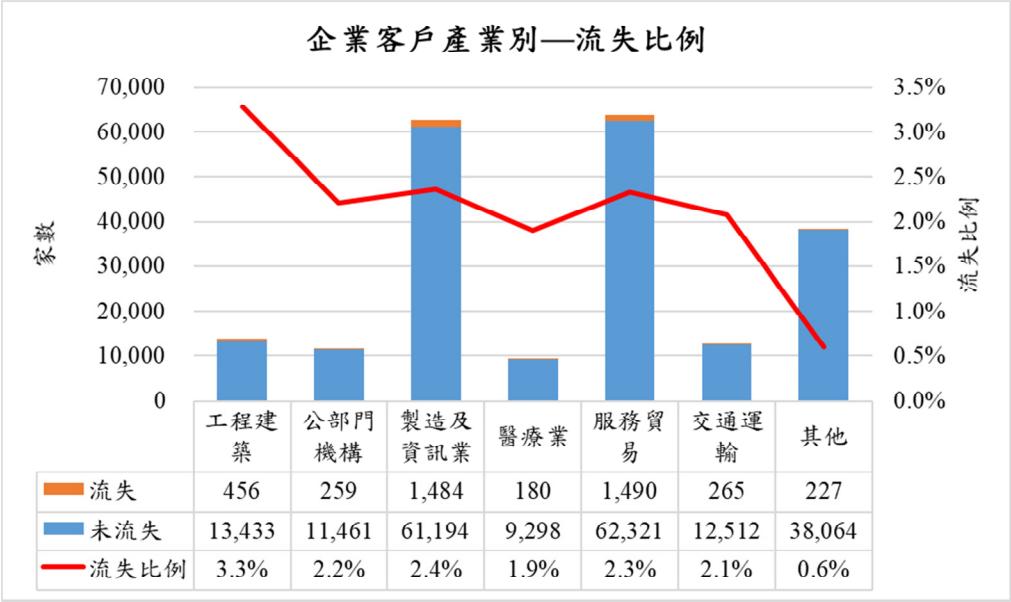


圖 4：企業客戶產業別-流失比例

資料來源:本研究整理

4. 企業客戶主要據點

統計中南部及東部企業客戶流失情形，中南部客戶家數多且流失比例高(中部 2.2%、南部 2.0%)，而東部客戶家數少且流失比例最低(1.2%)，推測可能因企業客戶大多集中在中部及南部，且中南部較東部更加都市化，網路基礎建設早已完備，不像東部地區仍需投入成本建設開發，故中南部成為電信業者搶奪客戶之主要區域。該電信公司在規劃佈置服務人力時，通常大量投注在中南部地區，且提供較多資源以鞏固客戶，以應對中南部的激烈競爭，而建置流失預測模型有助於該公司提高客戶固守率。(詳如圖 5)

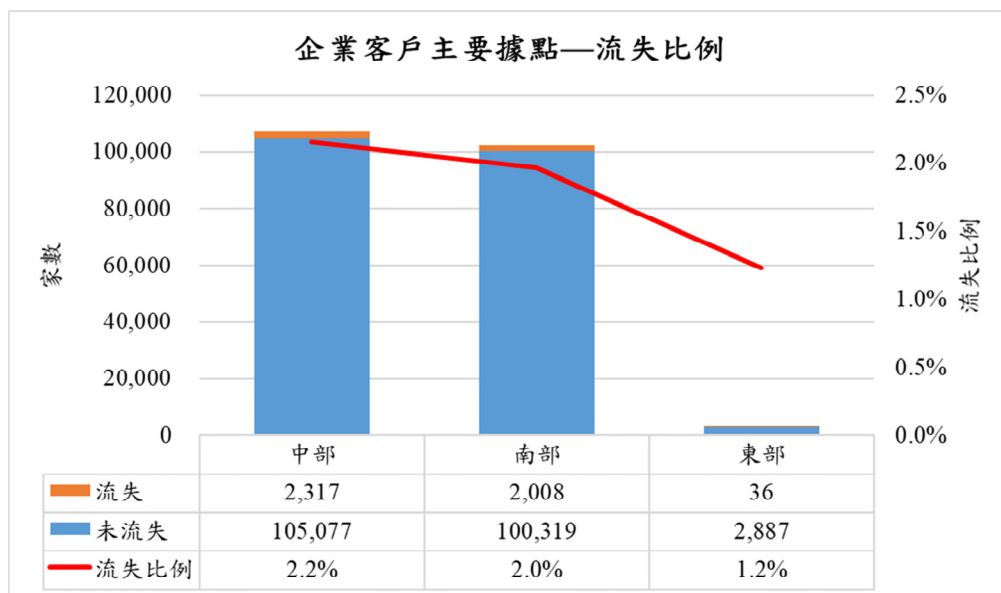


圖 5：企業客戶主要據點-流失比例

資料來源:本研究整理

5. 優惠折扣申請

統計企業客戶 105 年是否申請過優惠，未申請優惠而流失的客戶最多 (4,000 家)但比例不高，推測可能並非單純對價格不滿而離開，而另有其他潛在原因，後續應在電信門市受理退租流程裡，加入一道流程關懷客戶退租原因，以作為後續行銷策略改進參考；另外有 361 家客戶曾經申請過優惠，但仍然流失，且其流失比例高，推測可能因提供優惠內容不符合客戶需求，故客戶依舊流失，然而因客戶願意申請優惠，電信公司可取得客戶期望之產品價格資訊，進一步分析客戶需求，以用來調整產品價格包裝。(詳如圖 6)

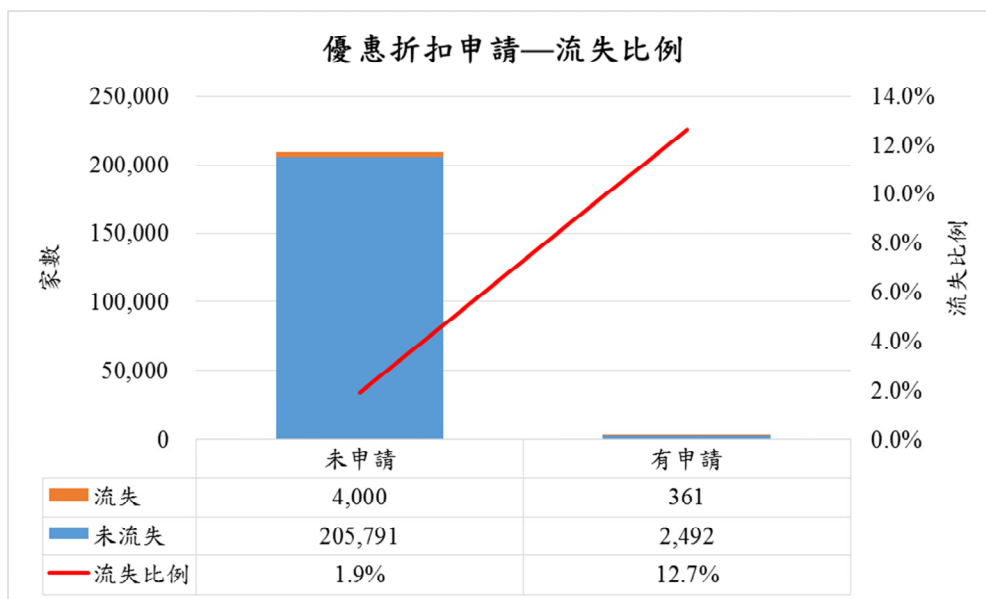


圖 6：優惠折扣申請-流失比例

資料來源:本研究整理

6. 前半年是否有異動

前半年無異動之客戶流失 3,375 家，流失比例 1.7%，前半年有異動之客戶流失 986 家，流失比例 6.1%，推測不少企業客戶是陸續分批退租，而非一次全部退租。(詳如圖 7)

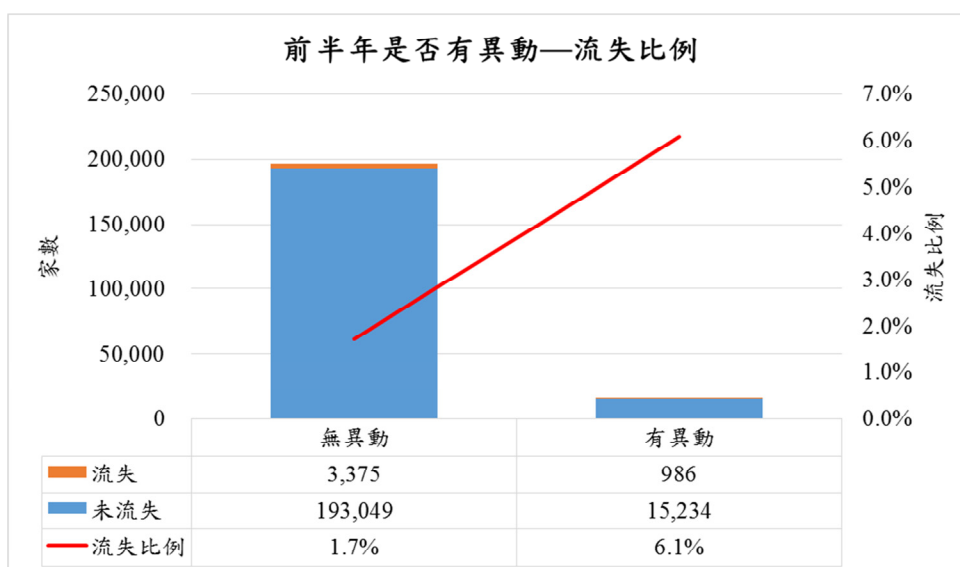


圖 7：前半年是否有異動-流失比例

資料來源:本研究整理

7. 企業客戶任期

依不同企業客戶任期之流失比例排序，分別是任期 3~4 年客戶流失比例 3.6%、任期 1~2 年客戶流失比例 2.8%、任期 2~3 年客戶流失比例 2.7%、任期 4 年以上客戶流失比例 1.9%、任期 1 年以內客戶流失比例 1.1%，推論前三類客戶(任期在 1 年~3 年間)流失比例偏高可能與合約年限相關，目前電信公司最常訂定的合約年限分別為 1 年、2 年及 3 年(少數特定業務會簽約 5 年合約)，可能因合約到期流失至競業，而任期 1 年以內客戶因為綁約而相對較無流失意願，後續可針對合約即將到期客戶制定鞏固方案，以防止客戶流失；此外，任期 4 年以上客戶流失比例只有 1.9%，符合之前研究所顯示的老客戶流失率較低(Kapoor, 2017)，然而流失家數並不少(3,374 家)，此類客戶至少已和電信公司續約過一次，卻流失至競業，推測可能有其他影響因素，後續可合併前述 6 項因子進行分析，以推論出客戶流失原因，進一步亦能用以建置流失預測模型。(詳如圖 8)

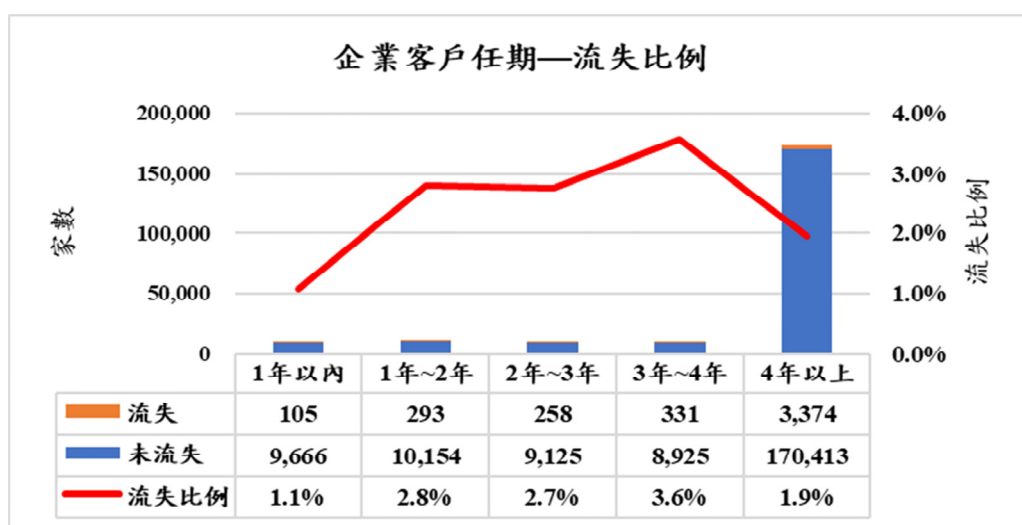


圖 8：企業客戶任期-流失比例

資料來源:本研究整理

依廣義線性模型(Generalized Linear Model, GLM)來計算自變數重要性，並依重要性排序為「企業客戶產業別_其他」、「專人服務」、「優惠折扣申請」、「企業客戶服務等級 D 級」、「企業客戶任期」、「企業客戶產業別_製造及資訊業」、「企業客戶產業別_公部門機構」、「前一季是否有異動」、「企業客戶服務等級 C 級」、「企業客戶產業別_交通運輸」...等。其中，排名前五名之變

數重要性遠大於其他變數，後續建置預測模型將列入參考。(詳如圖 9)

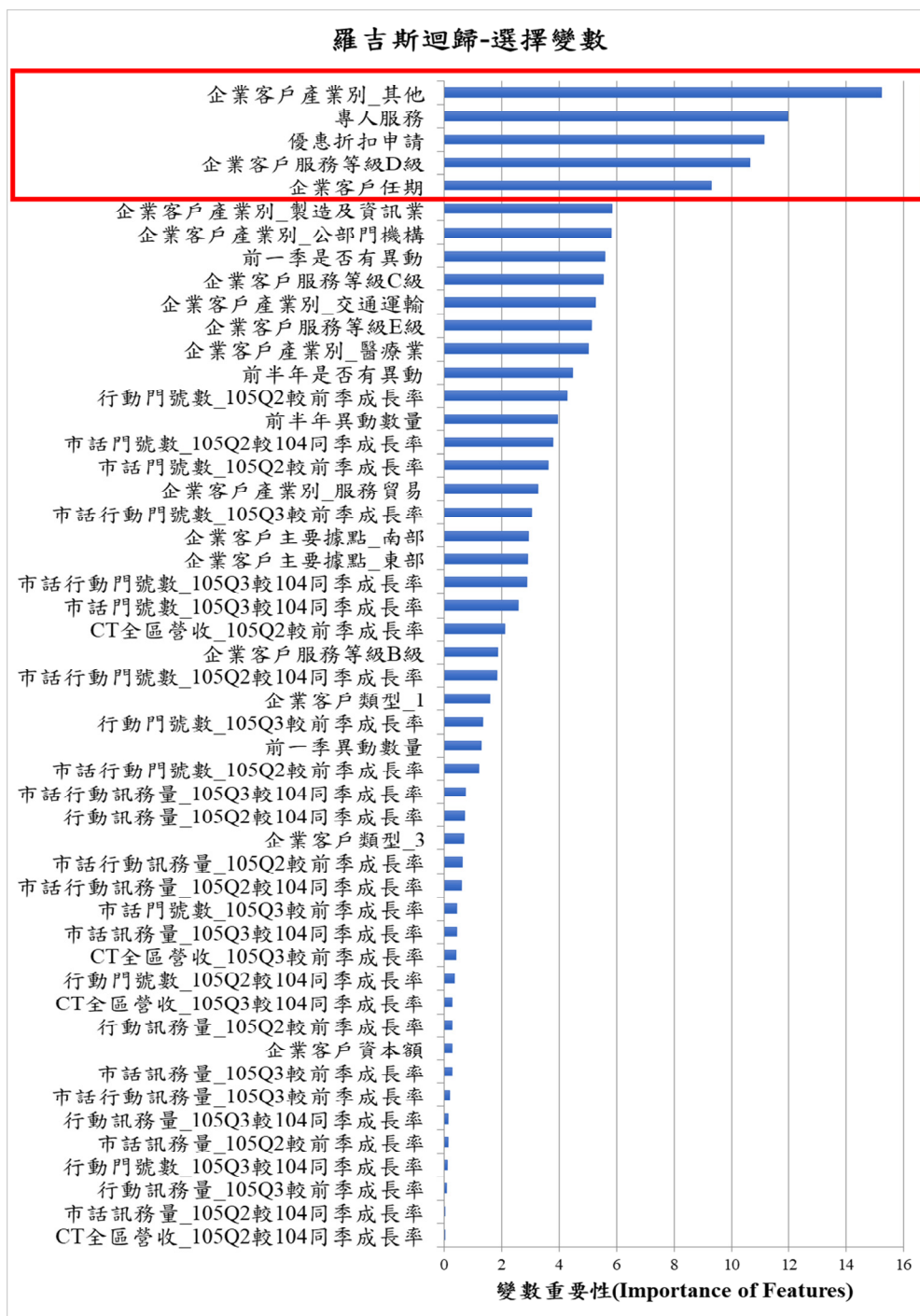
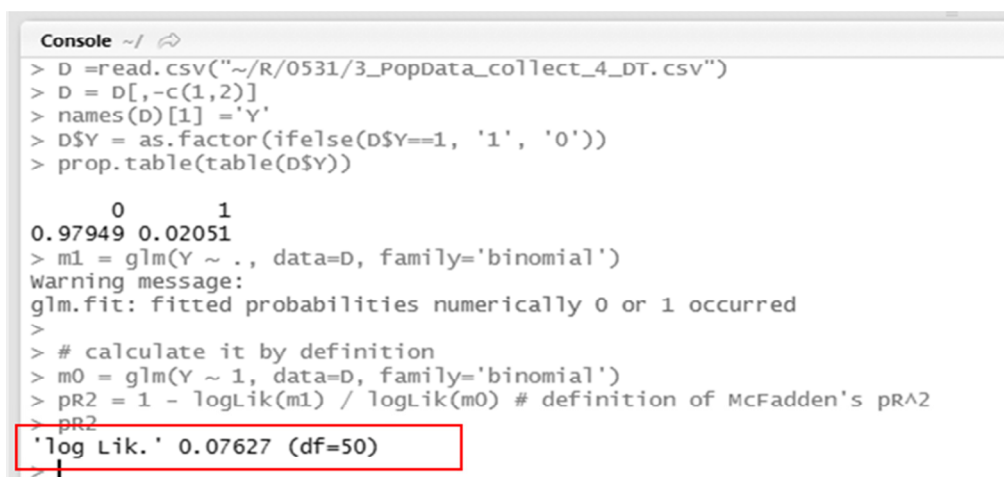


圖 9：變數重要性(Importance of Features)

資料來源:本研究整理

(二) 偽 R 平方 (Pseudo R²)

為檢驗資料解釋力，且考量本研究之依變數為二元變數，因此不採用線性迴歸常用之決定係數(R²)，而採用偽 R 平方(Pseudo R Squared)。圖 10 顯示 McFadden's 偽 R 平方 McFadden (1973) 為 0.07627，解釋力不足，顯示線性模式不適合用來預測，故後續不使用羅吉斯迴歸的預測模型。



```

Console ~/
> D = read.csv("~/R/0531/3_PopData_collect_4_DT.csv")
> D = D[, -c(1, 2)]
> names(D)[1] = 'Y'
> D$Y = as.factor(ifelse(D$Y==1, '1', '0'))
> prop.table(table(D$Y))

      0      1
0.97949 0.02051
> m1 = glm(Y ~ ., data=D, family='binomial')
warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
>
> # calculate it by definition
> m0 = glm(Y ~ 1, data=D, family='binomial')
> pr2 = 1 - logLik(m1) / logLik(m0) # definition of McFadden's pR2
> pr2
'log Lik.' 0.07627 (df=50)

```

圖 10：計算 McFadden's 偽 R 平方 (R 程式執行畫面)

資料來源:本研究整理

Xgboost 分析 (eXtreme Gradient Boosting)

我們接著採用非線性的 Xgboost 方式來進行變數重要性分析，Xgboost 是近年來非常熱門的一種預測演算法，並為許多機器學習競賽優勝的隊伍所採用。Xgboost 可以計算自變數重要性，並依重要性排序 (Chen & Guestrin, 2016)。本研究採用 Xgboost 演算法計算變數之重要性排行如圖 11 所示，其中與業務變化相關的變數，包括「行動訊務量_105Q2 較 104 同季成長率」、「行動訊務量_105Q3 較前季成長率」、「市話行動門號數_105Q3 較 104 同季成長率」、「市話訊務量_105Q3 較 104 同季成長率」、「行動訊務量_105Q3 較 104 同季成長率」等都列入前十個變數，而「企業客戶服務等級」、「企業客戶任期」、「前半年異動數量」、「專人服務」等也是重要變數，而這些變數在廣義線性模型中亦是顯著重要之變數。在前十大變數裡，企業客戶專屬變數占了三個(「市話行動門號數_105Q3 較 104 同季成長率」、「前半年異動數量」和「專人服務」)，這也說明了納入企業專屬變數的必要性。

值得注意的是「企業客戶資本額」在 Xgboost 的非線性模型中的重要性排名第一，「企業客戶資本額」之定義為企業客戶設立公司時於經濟部登記之資本額資料，可說是幾近於客戶營運規模，營運規模越大者通常流失率會越低。過去研究對於一般客戶流失預測裡，也顯示富裕客戶的流失率較低，Xgboost 分析結果與過去研究相互呼應。

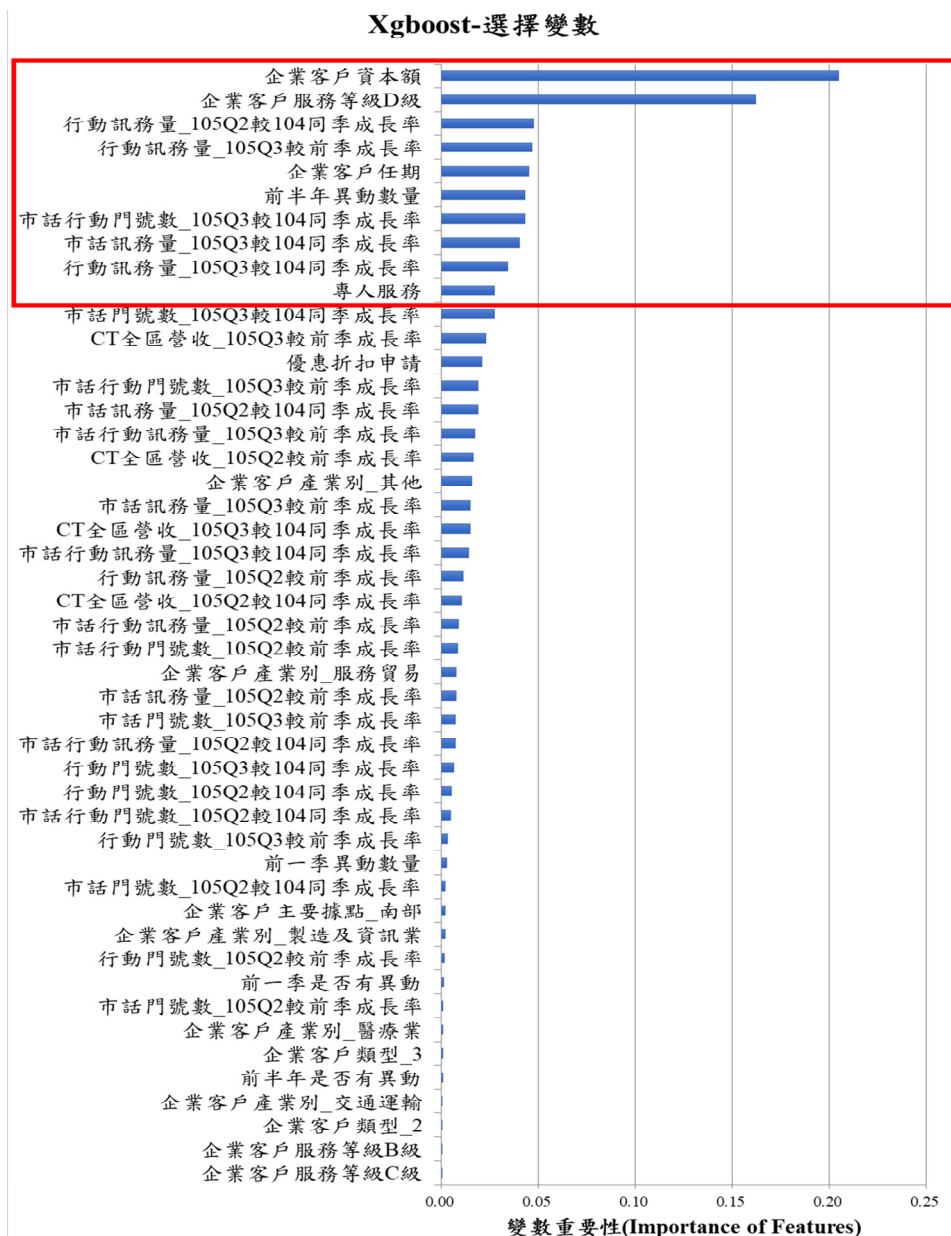


圖 11：Xgboost 變數重要性(Importance of Features)

資料來源:本研究整理

肆、流失預測

本章節承襲前面章節分析結果進行模型建置與評估，比較兩種不同變數組合之預測結果；另外模擬試算執行固守客戶策略後之營收損失，推算預測流失客戶且採取保留客戶措施之最適合閾值(Threshold)，以達到營收損失最小化。

一、模型建置過程與預測結果

本研究採用整體預測模型(Ensemble Predictive Model)演算法，整體預測模型學習演算法的運作是透過多次執行"基礎學習演算法"，並且針對每次產生的結果進行投票，最後整合投票的結果構成一致同意的最佳化成果。典型的整體預測模型學習演算法包括 Random Forest (隨機森林)、Adaboost、Gradient Boosting 與 Xgboost 等(Dietterich, 2002; Chen et al., 2015)。本研究採用上述的整體學習演算法模型，並執行「10 次交叉驗證法(10-fold Cross Validation)」，將整段流程重複執行 10 次，取其平均執行結果做為預測結果，閾值設定為 0.5；另考量本研究資料筆數較多，共有 212,644 筆，流失客戶數與非流失客戶數比例差異甚大，流失客戶比例僅佔 2.05%，本研究使用 R package 中縮減取樣(Down sampling)方式處理樣本數不平均的問題，因此將每一次的訓練資料集做縮減取樣(Down sampling)，測試資料集按照原比例不變，以避免預測結果出現偏差 (Freeman et al., 2012)。關於建置模型所使用之變數，本研究採用兩組不同變數 (詳如表 3)為基準變數(Baseline)與所有變數，再比對兩組預測結果。

表 3：企業客戶流失預測模型建置_變數組合

基準變數	所有變數
企業客戶服務等級	企業客戶產業別
企業客戶主要據點	企業客戶資本額
企業客戶任期	企業客戶服務等級
市話訊務量_105Q3較前季成長率	企業客戶主要據點
市話訊務量_105Q2較前季成長率	企業客戶任期
市話訊務量_105Q3較104同季成長率	市話訊務量_105Q3較前季成長率
市話訊務量_105Q2較104同季成長率	市話訊務量_105Q2較前季成長率
行動訊務量_105Q3較前季成長率	市話訊務量_105Q3較104同季成長率
行動訊務量_105Q2較前季成長率	市話訊務量_105Q2較104同季成長率
行動訊務量_105Q3較104同季成長率	行動訊務量_105Q3較前季成長率
行動訊務量_105Q2較104同季成長率	行動訊務量_105Q2較前季成長率
市話行動訊務量_105Q3較前季成長率	行動訊務量_105Q3較104同季成長率
市話行動訊務量_105Q2較前季成長率	行動訊務量_105Q2較104同季成長率
市話行動訊務量_105Q3較104同季成長率	市話行動訊務量_105Q3較前季成長率
市話行動訊務量_105Q2較104同季成長率	市話行動訊務量_105Q2較前季成長率
	市話行動訊務量_105Q3較104同季成長率
	市話行動訊務量_105Q2較104同季成長率
	企業客戶類型
	專人服務
	前一季異動數量
	前半年異動數量
	前一季是否有異動
	前半年是否有異動
	優惠折扣申請
	CT全區營收_105Q3較前季成長率
	CT全區營收_105Q2較前季成長率
	CT全區營收_105Q3較104同季成長率
	CT全區營收_105Q2較104同季成長率
	市話門號數_105Q3較前季成長率
	市話門號數_105Q2較前季成長率
	市話門號數_105Q3較104同季成長率
	市話門號數_105Q2較104同季成長率
	行動門號數_105Q3較前季成長率
	行動門號數_105Q2較前季成長率
	行動門號數_105Q3較104同季成長率
	行動門號數_105Q2較104同季成長率
	市話行動門號數_105Q3較前季成長率
	市話行動門號數_105Q2較前季成長率
	市話行動門號數_105Q3較104同季成長率
	市話行動門號數_105Q2較104同季成長率

1. 模型一：採用「基準變數(Baseline)」

本模型採用基準變數共 15 個，其來自電信業流失客戶預測之相關文獻提及之重要變數，包含企業客戶服務等級、企客主要據點、企業客戶任期、語音訊務量變化等，其中「企業客戶服務等級」及「企業客戶任期」均落在 GLM 變數

重要性分析及 Xgboost 變數重要性分析之排名前五名；語音訊務量變化之部份變數則在 Xgboost 變數重要性分析排名前十名；「企客主要據點」無論在 GLM 或 Xgboost 變數重要性分析均被視為較不重要變數，考量相關文獻研究結果，仍將此變數加入預測模型。預測結果詳如表 4.2，各項數據平均值分別為 AUC 值 68%，準確率(Accuracy) 68.6%，精確率(Precision) 4.1%，召回率(Recall) 67.6%。

表 4：採用基準變數之預測結果

演算法	AUC	準確率(Acc)	精確率(Pre)	召回率(Rec)
AdaBoost	67.3%	70.7%	3.9%	64.3%
Random Forest	67.8%	65.9%	4.1%	69.8%
Gradient Boosting	68.1%	69.8%	4.2%	66.2%
Xgboost	68.8%	68.1%	4.2%	70.1%
Average	68.0%	68.6%	4.1%	67.6%

資料來源:本研究整理

2. 模型二：採用「基準變數加上其他所有變數」

本模型採用基準變數(15 個)加上其他所有變數(23 個)，總共 38 個變數。除了基準變數外，加入所有表一之變數，預測結果詳如表 5，各項數據平均值分別為 AUC 值 71.6%，準確率 70.1%，精確率 5.1%，召回率 73.5%。

表 5：採用所有變數之預測結果

演算法	AUC	準確率(Acc)	精確率(Pre)	召回率(Rec)
AdaBoost	70.7%	72.1%	4.8%	69.2%
Random Forest	71.5%	67.3%	5.1%	76.4%
Gradient Boosting	71.9%	70.7%	5.0%	73.1%
Xgboost	72.4%	70.3%	5.3%	75.2%
Average	71.6%	70.1%	5.1%	73.5%

資料來源:本研究整理

3. 比較與結論

比較兩種模型結果，以 AUC、準確率、精確率及召回率來看，模型二「採用所有變數」之預測表現較好，尤其在 AUC 及召回率數據有明顯提升，顯示採用本研究所提出的之變數組合能有效改善預測結果。(詳如表 6)

本模型採用資料之流失客戶佔所有客戶比例僅 2.05%，非流失客戶佔 97.95%，兩者比例懸殊，影響精確率(Precision Rate)及準確率(Accuracy Rate)之計算；精確率係指“被正確預測流失客戶數”佔“所有被預測流失客戶數”之比例，受限於本模型資料結構，其精確率數字偏低，可提升程度有限；同樣地，由於非流失客戶數量很多，被猜中機率相對較大，影響準確率數字，故準確率不能完全解釋模型表現是否良好。

由於本研究目的著重在找出流失客戶進行固守，預測模型改善重點放在 AUC 及召回率的提升，期望盡可能找出真正流失客戶群，在模型二預測結果，召回率達到 73.5%，較模型一提升 5.4%，AUC 達到 71.6%，較模型一提升 3%，顯示模型二相對較符合本研究目的。

表 6：比較兩種變數組合之預測結果

	模型一（基準變數）	模型二（所有變數）
AUC	68.0%	71.6%
準確率(Acc)	68.6%	70.1%
精確率(Pre)	4.1%	5.1%
召回率(Rec)	67.6%	73.5%

資料來源：本研究整理

二、營收損失最小化計算

本節主要是以模型二為基礎，模擬試算執行固守客戶策略後之營收損失，推算最適合之閾值，以達到營收損失最小化。

首先根據流失預測模型判斷客戶流失與否之結果，決定是否執行固守方案，「被預測為流失之客戶」執行固守方案，「被預測為非流失之客戶」不執行固守方案，損失營收及執行結果說明如下(詳如圖 12)：

(一)損失營收計算：

1. 本研究設定客戶流失之損失營收為 182,664 單位，計算公式為假設該客戶全數電信業務流失，並且以平均綁約年限來算，流失到競業 24 個月。
2. 執行固守方案的成本為 7,200 單位，這是依據提供固守方案之優惠折扣，且重新綁約 24 個月，每戶電信費用貢獻減少 7,200 單位。

(二)執行結果：

1. 「被預測為流失之客戶」執行固守方案結果

- ☐執行固守方案後，客戶仍然流失，也就是固守失敗，每戶損失營收 182,664 單位。
- ☐執行固守方案後，客戶未流失，每戶損失營收 7,200 單位；其中客戶未流失因素包含兩種，一是固守客戶成功，另一則是客戶未有流失意圖，但被預測為流失客戶，仍然執行固守方案，造成營收減少。

2. 「被預測為非流失之客戶」不執行固守方案結果

- ☐不執行固守方案，客戶流失，也就是流失客戶未被正確預測出，每戶損失營收 182,664 單位。
- ☐不執行固守方案，客戶未流失，也就是客戶未有流失意圖，同時亦未執行固守方案，因此每戶損失營收 0 單位。

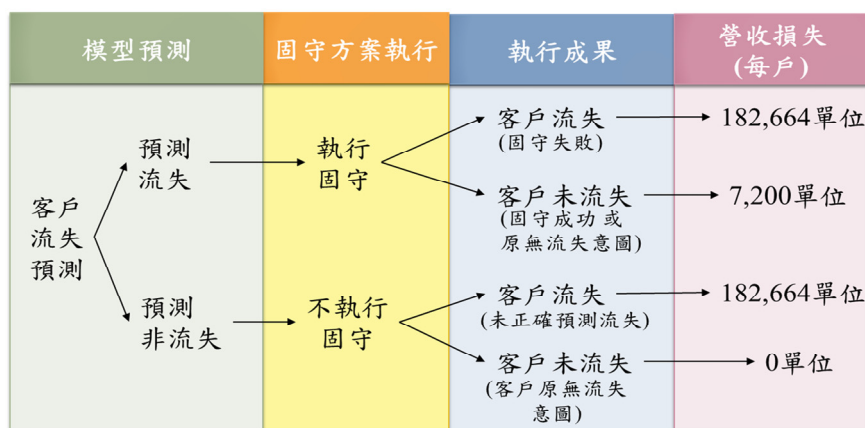


圖 12：固守方案執行過程

資料來源：本研究整理

依據上述各種情況之損失營收，重新以模型二為基礎計算最適合閾值，以得出最小化損失營收。以模型二得出之混淆矩陣之預測結果繪製損失營收矩陣 (詳如表 7)，預測非流失客戶之損失營收分別為 0 單位和 182,664 單位，預測流失客戶之損失營收分別為 7,200 單位和 47,557 單位，其中「預測流失且真實情況為流失客戶」之損失營收計算考量客戶固守成功率，我們採用個案公司 105 年度執行固守案之成功比例為 77%，其損失營收計算公式為

$$0.77 \times 7,200 + (1 - 0.77) \times 182,664 = 47,557。$$

接著將損失營收矩陣套用預測模型二，圖13顯示在各個不同的閾值時每位客戶平均營收損失。可以看出最適合閾值為0.72（在原模型設為0.5），最小化損失營收為3,221單位。當閾值設為0時，即全數視為流失並進行固守，損失營收為8,028單位；閾值設為1時，即不做預測且不執行固守，損失營收為3,746單位；運用預測模型找出流失客戶，執行固守方案，其損失營收相對較低，較完全不作為減少損失 $3,746-3,221=525$ 單位；較過度作為減少損失 $8,028-3,221=4,807$ 。後續可將預測模型廣泛應用，以減少營收損失。

表 7：損失營收矩陣

		預測結果	
		非流失	流失
真實情況	非流失	0	7,200
	流失	182,664	47,557

資料來源：本研究整理

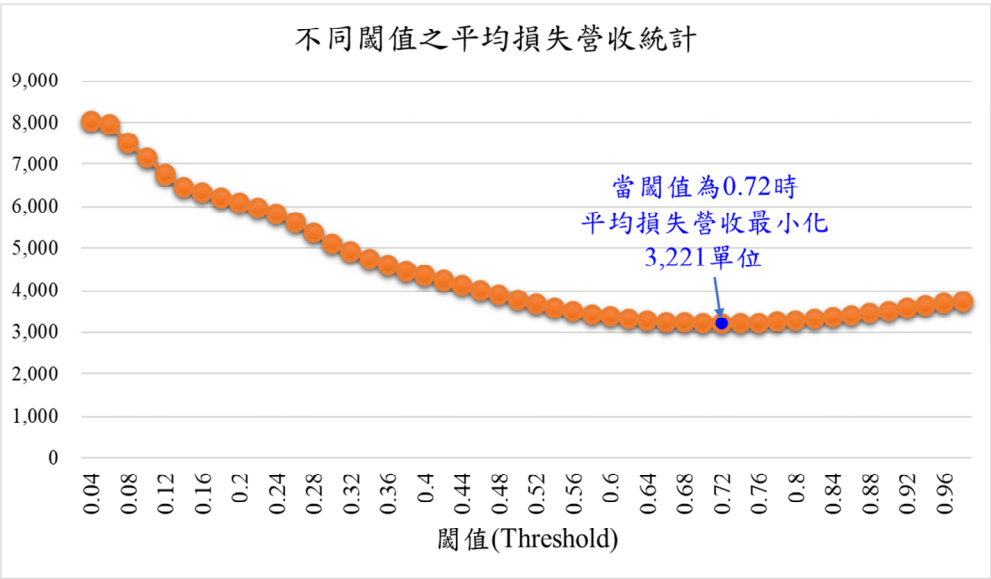


圖 13：不同閾值之平均損失營收統計

資料來源：本研究整理

伍、結論

一、研究結論與貢獻

過去許多電信業務流失預測之相關研究對象為消費客戶，本研究則是著重在企業客戶流失預測，因此在蒐集變數的過程中，本研究在考量企業客戶特性後，彙集消費客戶預測之重要變數，並依其合適程度進行篩選，最後留下 15 個變數作為基準變數；接著依據企業客戶特性，產生企業客戶之專有變數加以用現行最常用之整體學習演算法建立預測模型；結論是「採用所有變數」之變數組合所產生預測模型表現較基本變數模型更為準確，其 AUC 達到 71.6%，召回率最高可達到 76.4%，顯示加入企業客戶專有變數後，更能有效找出流失客戶。

在本研究建置整體學習演算法預測模型並採用了 Xgboost 分析法來決定重要變數，從 Xgboost 分析法導出之前十大重要變數(詳如圖 11)來看，其中六個變數與訊務量及門號數異動相關，包含訊務量本季及上季成長率、門號數異動數量及成長率，因企業客戶通常租用多個門號，且退租時採分批退租，顯示在客戶流失前即會有些許跡象透露出客戶流失意圖，可透過資訊系統自動觀測客戶訊務量及門號數成長率變化，當其降低至某個警戒值之下，系統主動通知專人以進行後續行動，如主動關懷客戶及提供固守方案；另有兩個類別變數與客戶服務相關，分別是企業客戶服務等級及指派專人服務，該電信公司針對企業客戶月營收貢獻區分不同服務等級，且在特定等級以上指派專人服務，根據前述章節分析服務等級最高之企業客戶流失比例最高，該類客戶為競業積極爭取對象，其流失可能性相對較高，而影響該類客戶採購意願之因子除價格因素外，尚有網路架構相容性、備援機制、通訊品質、電信公司提供專人服務之品質...等，電信公司除了提供合適優惠方案外，對外可主動協助規劃客戶網路架構，設定多重路由提供備援機制，對內應定期檢討改善通訊品質、透過檢視客戶滿意度來了解專人服務品質優劣，以鞏固客戶；再者，透過分析企業客戶任期這項變數發現企業客戶流失比例偏高與合約年限相關，因此可針對合約即將到期客戶制定固守方案，以防止客戶流失；企業客戶資本額是 Xgboost 分析法認定之最重要變數，由於實務上亦使用企業客戶資本額推論其應有之電信營收貢獻，判斷客戶使用競業產品的比例，基於 80/20 法則，優先關注高資本額客戶流失情形，在前述提及運用系統觀測客戶訊務量及門號數成長率變化時，此類客戶之警戒等級應設為最高，以盡早採用鞏固行動。

最後，考量實務面作法，在得到預測結果後，依據結果估算執行固守客戶策略所帶來的營收損失，藉由設定最適合閾值，使營收損失最小化，結論是當閾值設為 0.72 時，可產生最小營收損失 3,221 單位。

二、未來研究建議

由於企業客戶是否要租用電信業務的決策過程非常複雜，通常是多人討論決定，並且與企業內部預算規劃相關，有些企業客戶甚至考量其上下游廠商使用何家電信業者，來決定自己要選擇哪家，或是將上下游廠商的行動電話門號組成群組互撥，因此在分析預測企業客戶是否流失，可以從各個不同層面來探討分析。

受限於本次建置預測模型可取得資料內容，僅能依企業客戶訊務量、門號數及營收數進行分析，後續建議可將企業客戶間的關係網路納入，例如透過企業客戶電話互撥、網路互連等資料，先推論出企業客戶間的關係，再探討某家企業客戶流失，是否會連帶影響與它友好的企業客戶流失，接著將探討結果重新納入本次預測模型，應可有效提升模型預測準確度，進而對電信業者固守客戶有所貢獻。

參考文獻

- 國家通訊傳播委員會(NCC)，2016，「我國電信業者營運實績統計數據」，
https://www.ncc.gov.tw/chinese/news.aspx?site_content_sn=1974，accessed on May, 25,
2017。(NCC, 2016, “Taiwan Telecom Industry Operational Performance Statistics,”
https://www.ncc.gov.tw/chinese/news.aspx?site_content_sn=1974，accessed on May, 25,
2017.)
- Au, T., Ma, G., & Li, S., 2003, “Applying and Evaluating Models to Predict Customer
Attrition Using Data Mining Techniques,” **Journal of Comparative International
Management**, Vol. 6, No. 1, 10-23.
- Berson, A. & Smith, S. J., 2002, **Building Data Mining Applications for CRM**, 1st, New
York: McGraw-Hill.
- Chen, T. & Guestrin, C., 2016, “Xgboost: A Scalable Tree Boosting System,” **Proceedings
of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and
Data Mining**, New York, United States.
- Chen, T., He, T., & Benesty, M., 2015, **Xgboost: extreme gradient boosting**. R Package
Version 0.4-2, Vol. 1, No. 4, 1-4.
- Dietterich, T. G., 2002, Ensemble learning. **The Handbook of Brain Theory and Neural
Networks**, 2, 110-125.
- Freeman, E. A., Moisen, G. G., & Frescino, T. S., 2012, “Evaluating Effectiveness of
Down-Sampling for Stratified Designs and Unbalanced Prevalence in Random Forest
Models of Tree Species Distributions in Nevada,” **Ecological modelling**, Vol. 233,
1-10.
- Hassouna, M., Tarhini, A., Elyas, T., & AbouTrab, M. S., 2016, “Customer Churn in Mobile
Markets A Comparison of Techniques,” **International Business Research**, Vol. 8, No.
6, 224-237.
- Jamil, S. & Khan, A., 2016, “Churn Comprehension Analysis for Telecommunication
Industry Using ALBA,” **Proceedings of Emerging Technologies (ICET), 2016
International Conference**, Islamabad, Pakistan.
- Kapoor, A., 2017, “Churn in the Telecom Industry – Identifying customers likely to churn
and how to retain them. Technical report.” Retrived July 10, 2017, from
<https://wp.nyu.edu/adityakapoor/>
- Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., & Abbasi,
U., 2014, “Improved Churn Prediction in Telecommunication Industry Using Data

- Mining Techniques,” **Applied Soft Computing**, Vol. 24, 994-1012.
- Kotler, P., Keller, K. L., Ancarani, F., & Costabile, M., 2014, **Marketing Management** 14th, Boston: Pearson.
- Mattison, R., 2006, **The Telco Churn Management Handbook**, 1st, Oakwood Hills, Illinois: XiT Press.
- McFadden, D., 1973, Conditional Logit Analysis of Qualitative Choice Behavior. P. Zarembka, ed. **Frontiers of Econometrics**, New York: Academic Press 105-140.
- Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H., 2006, “Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models,” **Journal of Marketing Research**, Vol. 43, No. 2, 204-211.
- Peppers, D. & Rogers, M, 2016, **Managing Customer Experience and Relationships: A Strategic Framework**, 3rd, New Jersey: **John Wiley & Sons**.
- Premkumar, G. & Rajan, J, 2017, “Customer Retention in Mobile Telecom Service Market in India: Opportunities and Challenges,” **Ushus-Journal of Business Management**, Vol. 12, No. 2, 17-29.
- Reichheld, F. F. & Sasser, J. W, 1990, “Zero Defections: Quality Comes to Services,” **Harvard Business Review**, Vol. 68, No. 5, 105-111.
- Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C., 2015, “A Comparison of Machine Learning Techniques for Customer Churn Prediction,” **Simulation Modelling Practice and Theory**, Vol. 55, 1-9.
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B, 2012, “New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach,” **European Journal of Operational Research**, Vol. 218, No. 1, 211-229.
- Verbeke, W., Martens, D., Mues, C., & Baesens, B, 2011, “Building Comprehensible Customer Churn Prediction Models with Advanced Rule Induction Techniques,” **Expert Systems with Applications**, Vol. 38, No. 3, 2354-2364.
- Wei, C.-P. & Chiu, I.-T, 2002, “Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach,” **Expert Systems with Applications**, Vol. 23, No. 2, 103-112.

作者簡介

黃三益

黃三益於1994年獲得美國明尼蘇達大學博士學位，1995年開始任教於國立中山大學資訊管理系，現為資管系教授，他的研究興趣包含文字探勘，推薦系統，工作流程系統，和資料庫系統。

Email: syhwang@mis.nsysu.edu.tw

林彥君

國立中山大學資訊管理研究所畢業，現任職於中華電信公司，主要研究領域為機器學習與數據分析。

Email: leoushiuh@gmail.com

賴佳瑜

國立中山大學資訊管理博士，現任教於國立屏東科技大學資訊管理系助理教授。學術背景與工作經歷兼具財務管理與資訊管理領域。主要研究領域為文字探勘、機器學習、專利探勘、專案管理，論文主要發表於PAJAS、Research Policy、International Journal of Project Management等期刊。

Email: chiayulai@mail.npust.edu.tw

