


東海大學  
數位創新所

技術報告



建構客戶回購預測模型  
以國內大型購物中心為例

研究生：姜詠韜

指導教授：周忠信 教授

中華民國一零八年一月

# Customer Repurchase Intention Prediction

## - A Case Study of a Shopping Mall

By

Yung-Tao Chiang

Advisor: Prof, Jung-Sing Jwo

January 2019

Taichung, Taiwan

# 建構客戶回購預測模型-以國內大型購物中心為例

學生：姜詠韜

指導教授：周忠信 教授

東海大學數位創新碩士學位學程

## 摘要

零售業競爭激烈，為提高消費意願，企業積極整合線上及線下發展營銷服務。電商平台滿足了消費者線上購物的便利性，近期消費者更關注的是線下的活動及購物體驗。如何有效將對的促銷訊息傳送給有需要的正確顧客，是推動線下商務活動成功的重要關鍵。本論文透過大數據及數據挖掘技術，以某一台灣北區之購物中心作為實證案例。依其年度行銷活動規畫作為背景，首先挖掘出顧客行為與活動檔期回購機率關係。透過會員行為、購買紀錄及賣場環境等數據，建立回購預測模型並產生回購目標客群名單作為行銷人員促銷對象。在此模型下該購物中心可以依其目標，如營收或獲利為主，產出對應目標客群名單，進而達成該購物中心在此次活動檔期之預設目標。

中文關鍵字：客戶回購預測、推薦系統、大數據、決策樹、零售

# Customer Repurchase Intention Prediction

## - A Case Study of a Shopping Mall

Student: Yung-Tao Chiang

Advisor: Prof, Jung-Sing Jwo

Program of Digital Innovation at Tunghai University

### **Abstract.**

The retail business is highly competitive. To promote sale revenues, the online and offline service integration has become a newly strategic trend. In recent years, consumers have been enjoyed with online e-commerce which provides the convenience of online shopping. Yet, consumers are still paying their attention to and enjoy those offline activities and shopping experiences. Therefore, to deliver the right promotion channel and to target the right customers are the key to success in offline business. In this paper, we use the big data and modeling techniques to analyze the operation of a shopping mall as a case study in Taiwan. On the basis of the annual sales and business activities, we can explore the repurchase intentions through customers' behavior and their transactional activities. We also establish a predictive model by using data, including custom members' behavior, purchase records, store environment information, etc. From the related data, a return custom member list can be generated for the marketing promotion target. The store managers of the Shopping mall who is targeting to maximize revenue or profit can make their decision by using this model to generate a list of its target customers. This could lead the shopping mall to achieve its pre-set goals for their business goal.

*Keywords : customer repurchase intention prediction, recommender system, big data, decision tree, retailing*

# 目錄

摘要.....	3
Abstract.....	4
表目標.....	6
圖目錄.....	6
Chapter 1 緒論.....	7
Chapter 2 文獻探討.....	9
2-1 推薦系統.....	9
2-2 客戶關係管理、客戶輪廓與客戶分群.....	10
2-3 大數據與數據挖掘.....	12
2-4 迴歸分析與結構風險最小化.....	14
2-5 分類和迴歸樹.....	15
2-6 混淆矩陣.....	16
2-7 ROC 曲線.....	17
2-8 成本效益分析.....	18
Chapter 3 案例背景.....	20
3-1 背景.....	20
3-2 數據結構.....	21
Chapter 4 大數據分析.....	24
4-1 研究方法及工具.....	24
4-2 決策樹模型的建立.....	26
4-3 數據驗證.....	30
4-4 實務應用.....	35
Chapter 5 結論.....	38
5-1 結論.....	38
5-2 限制與展望.....	38
參考文獻.....	40

# 表目標

表 4-1 2018 預測與實際資料混淆矩陣 (Score >0) .....	31
表 4-2 2018 預測與實際資料混淆矩陣 (Score >0.1) .....	33
表 4-3 2018 預測與實際資料混淆矩陣 (Score >-0.1).....	34
表 4-4 分數等級與精準性比較.....	35
表 4-5 分數等的成本效益試算比較 .....	36

# 圖目錄

Figure 2-1 multiple regression .....	14
Figure 2-1 structural risk minimization .....	15
Figure 2-2 混淆矩陣.....	16
Figure 2-2 ROC & AUC .....	17
Figure 2-3 AUC (Area Under The Curve) .....	18
Figure 4-1 數據分析的作業流程圖.....	24
Figure 4-2 數據預處理:Data Model (UML) .....	25
Figure 4-3 系統操作 SAP PA – Variable 選擇.....	27
Figure 4-4 系統操作 SAP PA –決策樹模型-變數貢獻度分析.....	28
Figure 4-5 系統操作 SAP PA –決策樹呈現 Level 1 & 2.....	28
Figure 4-6 系統操作 SAP PA – 決策樹呈現 Level 1 & 2 & 3.....	29
Figure 4-7 系統操作 SAP PA –預測模型的績效與品質 .....	29
Figure 4-8 系統操作 SAP PA – 2017 年隨機抽樣 25%的混淆矩陣.....	30
Figure 4-9 2017-2018 母親節檔期會員回購統計 .....	31
Figure 4-10 回購人數與比例分析.....	32
Figure 4-11 獲客費用與回購率 .....	33
Figure 4-12 分數等級與精準性比較.....	35
Figure 4-13 分數等的成本效益試算比較.....	36

# Chapter 1 緒論

邁爾.舍恩伯格(Viktor Mayer-Schönberger)在「大數據資本主義」(Reinventing Capitalism in the Age of Big Data)一書提到，數據正在取代金融成為市場經濟的核心驅動力 [1]。而近年來大數據與深度學習等相關技術的快速發展，更於不同領域中綻放出許多創新應用。特別是在包含零售、百貨、與餐飲等行業上 [2]。

傳統上，零售、百貨、與餐飲等行業競爭激烈，特別是在線上(online)電子商務極為發達的今天。為吸引顧客回到線下(offline)實體店面消費，各店家無不卯盡全力，尋找各種可能方法以提高機會。透過建立會員關係，辦活動、提折扣等，是當前最常見的線下作法。然而如何有效將對的促銷訊息傳送給正確有需要的顧客，顯然是推動線下商務活動成功的重要關鍵。

傳統線下商務促銷的做法，以使用數據庫分析的客戶關係管理(Customer Relationship Management, CRM) 作為選擇顧客的依據 [3]。在大數據與人工智慧頗為成熟的今天，”推薦系統”(Recommender System, RS)早已成為無論是線上或線下商務活動的重要推手 [4]。然而針對線下商店，推薦系統儘管可以有效協助定義出目標客群，但如何可以在促銷的投放成本、營收與獲利之間找到平衡點，有時候對於店家而言實更加重要。

面對資訊過載(information overload)的現在，推薦系統的應用是主要解決方式之一。現行推薦系統的主流應用是在資訊的搜尋及商品推薦，透過搜尋引擎及電商平台的商品推薦功能，讓使用者更簡易的操作服務。隨著企業的資訊化提升，透過 ERP(Enterprise Resource Planning, ERP)、POS(Point of Sale, POS)、CRM 及智能零售(smart retail)的持續建置，企業未來將能收集到更多的商品及消費訊息，營銷人員也面臨了資訊過載的挑戰。這麼多的會員、商品及交易資訊，那一些對營銷活動有幫助、是否可以提升行銷活動決策效率、甚至更進一步改善活動績效等，是企業亟需具備的新能力。

為探討上述問題，本論文透過大數據及數據挖掘技術，以某一台灣北區之購物中心作為實證案例。依其年度行銷活動規畫作為背景，首先挖掘出顧客行為與活動檔期回購機率關係。透過會員行為、購買紀錄及賣場環境等數據，建立回購預測模型並產生回購目標客群名單，作為行銷人員促銷對象。在此模型下該購物中心可以依其目標，如營收或獲利等，產出對應目標客群名單，作為該購物中心在此次活動檔期之主要行銷對象。本論文所使用之數據皆為該實證案例之真實數據，其活躍會員約 10 萬人，使用數據達 90 萬筆，活動檔期則以 2018 年之母親節做為探討對象。

本論文架構共分五章，其中第一章介紹研究動機與目的；第二章為相關文獻探討；第三章為本論文案例背景介紹；第四章則為實證數據分析與結果探討；第五章為本論文之結論與未來展望。



# Chapter 2 文獻探討

本論文旨在探討實體門店顧客的重複購買模型，藉由會員消費歷史預測未來參與活動的重複購買行為。本章將針對推薦系統、客戶關係管理、客戶輪廓與客戶分群、大數據與數據挖掘、迴歸分析與結構風險最小化、分類與迴歸樹、混淆矩陣、ROC 曲線以及成本效益分析等背景先行介紹。

## 2-1 推薦系統

互聯網(Internet)與電子商務的崛起帶來使用者大量的資訊，以往要查詢商品資訊，會需要透過走訪的方式取得，現在透過互聯網即可快速連結；對於服務的提供者，以往難以記錄與使用者的互動行為，但現在透過線上系統的服務，所有歷程都可以留下歷程，協助服務提供者分析使用軌跡。然而隨著資訊量的快速增長，使用者無法從大量訊息中獲取真正所需的訊息時，就會發生資訊過載問題 [5]。

推薦系統是解決資訊過載的一種方法，推薦系統屬於資訊過濾的一種應用，將使用者可能喜好或需要的資訊或內容推薦給使用者。常見推薦系統的運作原理，首先是收集使用者歷史數據，再透過大數據相關領域中的推薦技術計算，最後再對使用者作個性化的推薦。現在的推薦系統會主動收集使用者的即時訊息及採用計畫，調整推薦的策略，以產生更符合使用者需求的推薦結果 [4]。推薦系統已經廣泛應用於很多領域，其中最典型的應用就是電子商務領域，但隨著人工智慧及大數據技術的快速發展，也開始帶動不同領域的應用 [5]。

推薦系統的技術有很多種，這裏介紹兩種主要方法 [6]：

- 協同過濾推薦(collaborative filtering recommendation)

系統假設為使用者主動推薦感興趣的內容時，可以先找到與這個使用者有相同興趣的其他使用者，將他們感興趣的內容推薦給此使用者。日常生活中，就像是有共同興趣的朋友，相互分享自己感興趣的主題 [7]。協同過濾推薦方法是從使用者角度出發，具有以下特點：

- 自動獲取：使用者不需填入任何資料，只要持續使用系統，系統即可透過瀏覽或購買行為以隱性的方式建立推薦模型。
- 共享使用經驗：由使用者行為所產生，避免了需要討論過濾法在計算上的完整性及精確性。
- 推薦新訊息的能力：多數推薦法是基於使用者自身的偏好，但透過協同過濾的方式，可以提供使用者完全不相似的資訊，進而發現未發現的潛在興趣偏好。
- 初始數據量不足會有無法有效推薦的稀疏（sparsity）問題。
- 基於內容推薦(content-based recommendation) [8] [9]

系統假設每個內容與使用者會有特徵，藉由特徵對不同的使用者作不同的內容推薦。例如，使用者購買了科幻小說，就會為他推薦該作家其他的作品。此方法的發展上，多是使用機器學習或數據挖掘手法，計算並找出內容與使用者的特徵並找出相關匹配程度，依匹配程度高低提供使用者個性化的推薦內容。常用推薦模型建立方法有決策樹、迴歸分析與類神經網路等。基於內容推薦法的特點有：

- 沒有數據稀疏問題，不需要大量用戶群體、交易記錄或評分記錄，即可產生推薦列表。
- 可對特定內容(如冷門的產品)或使用者(特殊興趣的愛好者)產生推薦列表。
- 建立方式主要是依使用者或內容特徵，並可作推薦解釋增加信服度。
- 為了能取得有效的推薦模型，內容的特徵需要有良好的結構性及數據品質。

## 2-2 客戶關係管理、客戶輪廓與客戶分群

企業希望能與能藉由強化客戶的經驗，進而提升營收。客戶關係管理主要目

的是協助企業管理客戶價值、分類識別及盤點商機推進，包含有下列四個維度 [3]：

- 找出有機會經營的客戶群
- 吸引客戶
- 保留客戶
- 做好客戶發展

企業必須系統化地建立客戶的資訊及知識，以便更深入地了解每個客戶的行為特點和需求。透過深入分析這四個維度的資訊，將可協助企業深入的了解客戶，從而最大限度的提高客戶對企業的價值。

客戶輪廓(customer profile) 是指獲取客戶數據，收集喜歡(或不喜歡)及特徵的訊息，貼上標籤，以建立分析數據以掌握有關客戶行為整體資訊。之後在企業面向客戶的各個流程中，向決策者提供信息以促進盈利行為的流程 [10]。輪廓特徵(profiling characteristics)包含有：

- 人口統計屬性：人口統計數據概念的一些示例包括年齡，性別，種族，原籍國，種族，婚姻狀況，教育水平，收入水平，居住類型，家庭中的成年人數量，更進一步的如家庭兒童數量或汽車數量。
- 行為屬性：是與反映交易歷史數據的相關屬性。如：產品消費模式，購買頻率，購買量或同時購買某些產品。
- 心理屬性：即描述實體獲得的定性方面的屬性。對於人們來說，例子可能包括與人格相關的屬性(如:樂觀)，興趣(如:露營)，生活方式(如:小康)。

客戶輪廓的使用，可幫助企業驗證其產品或服務的目標受眾(target audience, TA) 與消費者群組的關係。針對已驗證的會員，透過商業分析的手法建立以客戶為中心的商務模型。例如，當新客戶填入資料並驗證後，該客戶的特徵就可以

依定義細分被識別，企業可以檢視客戶與客戶之間潛在的共性關係。(如:顏色及類型的偏好)

客戶輪廓在實務上，主要是透過分析交易數據來找出分類模型。具體的應用有分析挖掘未來有潛力的產品組合，而這個產品組合會由有相同特徵的客戶群內的個人一起購買。例如，表達了對特定主題感興趣的客戶可以購買與該主題相關的各種書籍(如:露營)。識別對該特定主題表示興趣的客戶後，企業可以更加明確的向具有類似興趣的其他客戶，推薦關於該主題的產品。

客戶分群(customer segment)是指依各種客戶特徵、行為或潛在特性，對既有客戶群重新組合分群[8]。舉例而言，依據年齡層將所有會員分為青年、中年、長青等不同群組，分析師再根據不同群組找出不同的購買模式。如果這些模式存在顯著差異，則未來實際商務上即可根據客戶所屬群組來發展不同互動方式。客戶分群是一種建模過程，主要方法如下 [10]：

- 確認分析中包含的客戶屬性集合
- 分析數據並依相依性分群組
- 使用數據相依性質，通過區分來重新組織客戶分群
- 依據預期的業務行為對這些細分進行分類
- 使用細分模型對所有客戶進行分類

## 2-3 大數據與數據挖掘

大數據現在正在所有科學和工程領域迅速擴展，包括物理、生物和生物醫學科學。隨著網絡數據存儲、連結、收集能力的快速發展，各式業務活動以指數級增長生成數據。如 IBM 在”企業與大數據”一文提到 [11]，至 2012 年的統計，全球已有創建 2.5 quintillion bytes 的數據量，而 90%的數據都是在過去兩年內生成的；另一個例子是公共圖片共享網站-Flickr，平均每天從 2012 年 2 月到 3 月收

到 180 萬張照片。假設每張照片的大小為 2 兆位元組 (MB)，則每天產生 3.6 TB 的存儲空間。

大數據的處理技術涉及到大量、複雜及不斷增長的多樣數據集源。由於這種巨大的數據增長而無法輕易獲得有組織的知識和信息，也無法輕易理解或自動提取。在 Wu, Xindong 於 IEEE 的發表 Data mining with big data 中提到大數據具有大量、異構(heterogeneous)、分散式及去中心化的特性，數據挖掘是要在這之間探索其複雜性，並找出數據之間的關係 [12]。這樣的背景下，數據科學或數據挖掘的學科在資訊時代的發展愈是重要 [13]。

數據挖掘主要是從數據中發現有共性的知識，可用於指導決策制定和預測決策 [14]。例如，數據挖掘可以通過將客戶分成具有不同特徵和需求的組來提高營銷活動的響應率；也可以預測現有客戶即將流失的可能性 [15]。

常見的數據分析技術包含有：

- 關聯分析(association)
- 分類分析(classification)
- 聚類分析(clustering)
- 預測分析(forecasting)
- 迴歸分析(regression)
- 序列挖掘(sequence discovery)

數據挖掘是一種數據建模的方法，主要是從數據中發現共性，並找出有效益的知識 [16]。它可用於指導管理決策的制定和預測決策的效果。例如，數據挖掘可以通過將客戶分成不同特徵和需求的群組，進而提高營銷活動的回購率；也可以用於預測現有客戶轉換不同商品銷售或離開的可能性 [15]。

每種數據挖掘技術都可以下列一種或多種數據構建模型。前面談到四大 CRM 的應用維度，都可以由不同的數據分析技術從大量數據中提取或檢測隱藏的客戶特徵和行為，來實現數據建模型的目標 [17]。

## 2-4 迴歸分析與結構風險最小化

迴歸是一種數據挖掘技術，用於在給定特定數據集的情況下預測一系列數值(可稱為連續值)。進階的應用上，在多變量變項(multivariate data)的分析中，常使用的預測分析方法是多元迴歸分析(multiple regression analysis)法。以數學公式來表達，聯合  $p$  個自變項  $X$  來預測或解釋依變項  $Y$ ，表示被預測值(predicted value)， $b_0$  表示截距(intercept)， $b_1 \dots b_p$  表示每個自變項的加權係數值，迴歸分析的過程，便是運用最小平方方法(least square)來估計出這些迴歸係數和截距，並以  $X_1 X_2 \dots X_p$  符號來表示這項指標稱作「決定係數」(coefficient of determination)。決定係數的是指在依變項的總變異量中，可以被  $p$  個自變項聯合解釋的變異量百分比 [18]。

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

Figure 2-1 multiple regression

應用迴歸分析做預測，當所有假設都有效時，結果會是最佳解；但當一些假設無效，就會造成迴歸預測的無效。穩建迴歸(Robust Regression)是迴歸分析的其中一種算法，相較於二元迴歸法(squares regression)，適合解決數據樣本出現較多異常點(outliers)的狀況。對預測模型的影響當大 [19]。

傳統機器學習方法中普遍採用的經驗風險最小化原則，在樣本數目有限時是會出現過學習現象。統計學習理論提出了一種策略，在每個子集中尋找最小經驗風險，在子集間折衷考慮經驗風險和置信範圍，取得實際風險的最小。由 Vapnik 和 Chervonekis 於 1974 年所提出的結構風險最小化(Structural risk minimization, SRM)是用於從有限訓練數據集學習模型選擇的歸納原理。它描述了容量控制的一般模型(general model)，並提供了假設空間複雜度(space complexity)與擬合訓練數據質量(如:經驗誤差-empirical error)之間的衡量方式 [20]。

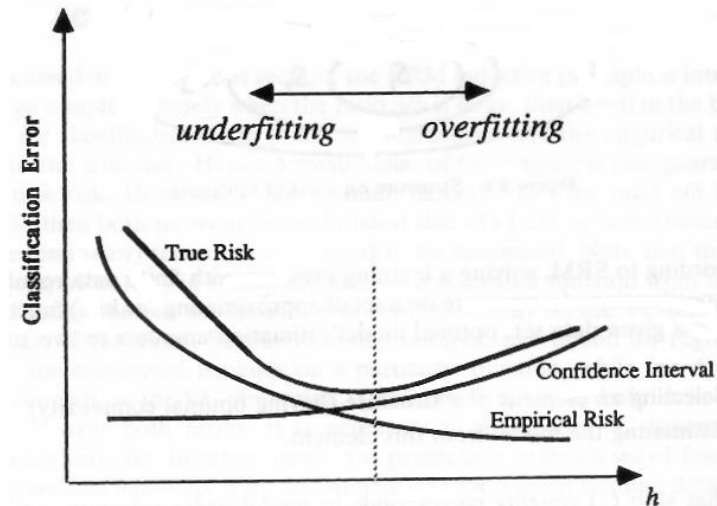


Figure 2-2 structural risk minimization

## 2-5 分類和迴歸樹

CART (Classification And Regression Trees) 是由 Breiman、Friedman、Olshen 與 Stone 所提出，全名為「分類和迴歸樹」 [21]。其做法為建立一棵決策樹，根據交互驗證或測試集檢驗之結果，將樹修剪成最佳樹型。在建立決策樹時，預測因子能對資料做不同預測，決定所選取之預測因子，它能降低資料的失序現象而被選取。CART 屬於非參數統計技術，在資料分析上不需特定的資料分配，其特色還有 [22]：

- (1) 可產生相互獨立的群體，直接和自動提供重要解釋變數的交互作用情形；
- (2) 結果易於解釋和瞭解，可以全面描述每個群體裡的個體特性，每個群體也有其結果變數的平均估計值；
- (3) CART 在進行資料分析時，可以處理複雜及多變數的資料結構，以及同時處理連續和間斷變數；
- (4) CART 不需對樣本空間進行特定結構的假設，除了可以找出變項間的關係，還可以清楚地呈現出分類模型，藉以了解變項間的交互作用情形。

已有很多研究運用決策樹進行分析、解決問題，如 Cho 應用在基於 Web 使用的個性化推薦系統 [23]；Taewijit 以速食業為案例，用決策樹作為客戶回購意向和關聯產品分析的概率方法之一 [24]；而 Chen 基於決策樹方法作應用內購買

重要元素提取 [25]；Khan 通過比較 Park Inn 和 Grand Hotel 確定了在客戶保留中起關鍵作用的因素，並發現 Grand Hotel 的客戶是根據服務而保留，而 Park Inn 的顧客則依據食品質量保留 [26]；Lin 通過決策樹方法預測集裝箱航運業的客戶保留可能性，應用決策樹方法來預測客戶不切換到新服務提供商的可能性 [27]。

## 2-6 混淆矩陣

為了鑑別模型的好壞，我們會以混淆矩陣(Confusion Matrix)及重要指標驗證 [28]。Figure 2-3 混淆矩陣為混淆矩陣的設計方式。

		Prediction outcome		
		positive	negative	
Actual value	positive	<i>TP</i>	<i>FN</i>	<i>TP + FN</i>
	negative	<i>FP</i>	<i>TN</i>	<i>FP + TN</i>
		<i>TP + FP</i>	<i>FN + TN</i>	

Figure 2-3 混淆矩陣

- TP：True Positive，預測為正向(Positive)且預測準確(True)。
- FP：False Positive，預測為正向(Positive)且預測錯誤(False)。
- FN：False Negative，預測為負向(Negative)且預測錯誤(False)。
- TN：True Negative，預測為負向(Negative)且預測準確(True)。

分析回購會員的預測模式與派送應用，優先關注下列指標：

- 準確性 (Accuracy)：(TP+TN) / Total N
  - 預測準確的整體佔比
  - 預測會來，實際有來；預測不來，實際沒來的
- 精確性 (Precision)：TP / (TP+FP)



- 在預測正向下，實際的「精準度」
- 預測會來，且實際有來的比例

## 2-7 ROC 曲線

ROC (Receiver Operating Characteristic) 曲線是一種座標圖式的分析方式，可用於比較找出最佳的信號(預測)模型 [29]。在同一模型中設定最佳閾值 (Threshold)。多數分析軟體已有內建 ROC 曲線功能，可用於判斷預測模型優劣參考。Figure 2-4 說明 ROC 與 AUC 的關係。

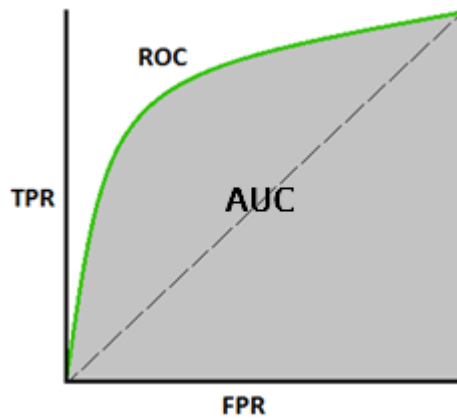


Figure 2-4 ROC & AUC

- $TPR \text{ (Sensitivity)} = TP / (TP+FN)$
- $FPR \text{ (1-Specificity)} = FP / (TN+FP)$
- AUC (Area Under The Curve)

曲線下方的面積，在比較不同分類模型時，可以將每個模型的 ROC 曲線都畫出來，比較曲線下面積做為模型優劣的指標。Figure 2-5 為 AUC=1、AUC=0.8 與 AUC=0.5 的比較。

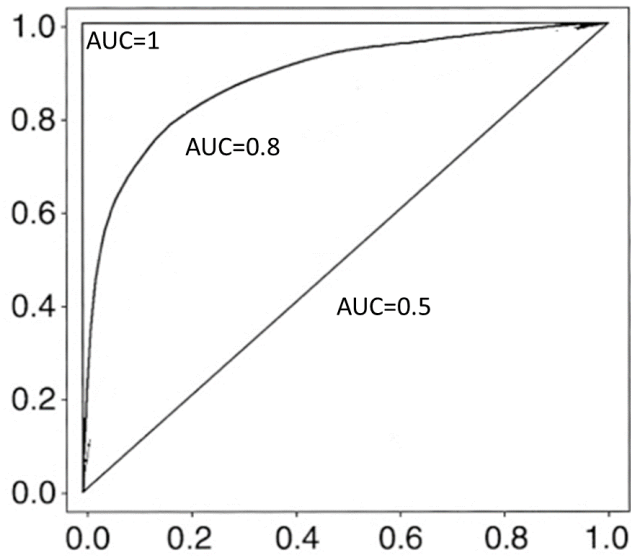


Figure 2-5 AUC (Area Under The Curve)

- AUC = 1，是完美分類器，採用這個預測模型時，存在至少一個閾值能得出完美預測。
- $0.5 < \text{AUC} < 1$ ，優於隨機猜測。這個模型能有預測價值。
- AUC = 0.5，跟隨機猜測一樣（例：丟銅板），模型沒有預測價值。
- $\text{AUC} < 0.5$ ，比隨機猜測還差；但只要總是反預測而行，就優於隨機猜測。

## 2-8 成本效益分析

成本效益分析(Cost Benefit Analysis, CBA)是通過比較專案的全部成本和效益來評估專案價值的一種方法 [30]。應用原理是針對某項支出目標，提出多個實現該目標的方案，運用相同的技術方式，計算出每種方案的成本和收益。並依據一定的原則，以比較的方法，選擇出最優的決策方案。

本研究探討回購與成本效益的關係。客單價及獲客單位成本可以用去年同期計算，進一步計算下列指標：

- 命中：預測正向且實際有回來的會員
- 命中會員營收：命中會員數量 \* 客單價
- 獲客成本：預測正向名單數量 \* 獲客單位成本

- 命中會員淨利： $(\text{命中會員數量} * \text{平均毛利}) - \text{獲客成本}$
- 營收費用比： $\text{命中會員營收} / \text{獲客成本}$
- 淨利費用比： $\text{命中會員淨利} / \text{獲客成本}$

# Chapter 3 案例背景

## 3-1 背景

百貨業的行銷活動規畫除了內容規畫要有新意之外，還有另一重點就是曝光。活動曝光的方式有很多種，從早期的電視廣告到後來的直郵廣告(Direct Mail, DM) [31]。在 2000 年後，隨著電商與線上交易的興起，迎來了 eMail 行銷的高峰期。但隨著 eMail 的泛濫，線上行銷的效果逐年下降，購物中心對於客戶互動的方式又從線上走到線下實體。透過精美的活動設計及邀請，讓會員有限定、限量及專屬感。但實體 DM 寄送到會員的成本遠較 eMail 的方式高，一份 DM 到會員手上的成本如果是 20 元，10 萬會員的購物中心就會有 200 萬元的前期費用，而一年有 12 個以上的檔期活動，光是郵寄費用就是可觀的投入。採用全面寄送方式，除增加費用成本、整體回購效益不高外，過於頻繁廣告也會降低活動參與動機。所以對行銷人員來說，名單與內容同樣需要慎密思考及規畫。

現有百貨零售業面對線上競爭及線下促銷挑戰，競爭實乃十分劇烈。各種活動層出不窮的行銷，逐漸使百貨零售業走向微利時代 [32]。年節、母親節及週年慶等大檔期，是百貨業的年度活動重點，也佔了大部份的營收 [33]。本論文以國內某大型購物中心為研究背景。此零售業客戶主要業務為專櫃經營，全台已有多家分店，每家分店有 50 個以上不同的專櫃品牌進駐，範圍包含食、衣、住、行及育樂類型。透過多年的積累，至今的總會員數量已超過 30 萬人，年活躍會員數量超過 10 萬人。在經營上，團隊努力每季持續引入新專櫃新品牌，以增加顧客對購物中心的新鮮感；為保持曝光，除了支持大型活動，一年 12 個月不停的辦各式主題活動吸引消費者來店；同時也透過會員專屬折扣、直郵行銷、電子行銷等方式，持續吸引會員回購消費。

本論文在探討上，為降低區域的影響及聚焦研究，我們鎖定了北區的單一門店，並以兩年的母親節檔期作為建模及驗證標的。

## 3-2 數據結構

本案例之購物中心提供之數據分兩大類，分別是會員資料與交易資料。數據欄位說明如下：

- 會員資料
  - 人口統計屬性：會員編號、性別、卡種、年齡、居住地、職業、學歷、開卡地點、收入等級、婚姻、成員數量
  - 心理屬性：興趣
  - 行為屬性：累計金額、消費頻率
- 交易資料
  - 商品：區域、分店、樓層、專櫃名稱、專櫃編號
  - 消費記錄：日期/時間、金額
- 其他：天氣、溫度(平均/最高/最低)、人流量、車流量

每個會員的人口統計數據主要來自於辦會員卡的過程中取得，除了期初建立資料及性別為固定，年齡可透過辦卡日期來推算，其他的數據都會有變動的可能，如職業、婚姻、家庭成員、居住地等，這部份可透過活動及廣告的寄送以維持正確性，但這會有賴購物中心對資料的經營狀況。不過行為屬性反而就比較簡單了，只要會員在消費時有輸入會員卡號，不用另外專人填寫，即可透過系統計算的方式掌握每個會員的累計消費金額、消費頻率，甚至可以細分到不同商品、不同類別作統計。本研究在行為屬性中有加入計算累計金額及消費頻率。

流通業相當重視商品與會員之間的關係，每個商品做市場規畫時會鎖定訴求購買的客群，之後也會研究主力購買的客群，並進一步研究與市場規畫的差異。購物中心的分析方式也相同，特點在於場域的規模放大，從單一的門店區域的管理放大到多層樓多門店的管理；所銷售的產品規模也放大，一般來說，購物中心會把消費品項的管理交由專櫃品牌處理，消費分析的重點的不是每個商品的銷售狀況，而是個別專櫃品牌的銷售量及業績與毛利貢獻，意即專櫃品牌就是購物中

心的商品；不同於單一品牌的商品分類，購物中心會以食、衣、住、行、育及樂等構面分析吸引的客群及營收的貢獻。

有別於一般門店的 POS 或品牌商的前後台。為能支持更快速、即時、大量的交易，購物中心的資訊系統會因功能性作區分。專櫃及分店數據一般是維護於 ERP 系統，交易數據則會在每次交易後存於 POS 系統。除了商品、會員及銷售資訊外，購物中心的特點就是在於場域，主打便利性讓消費者一站式滿足多樣性的購物需求，其衡量吸引力的重點就會放在來客數、提袋數的計算。影響場域型來客數最大的就是環境，如天氣好的時候，人們會想去踏青，而不是購物；但當天太熱時大家會想到百貨公司、電影院走走；當天氣冷時，就會想買衣服，想吃美食。由於對來客的影響甚大，相較於一般門店，購物中心從早期重視維護天氣及溫度的數據。本次研究以購物中心為背景，有天氣、溫度及人流量、車流量等環境的資訊合併收集，透過預測工具計算對消費的影響程度。

本研究先收集前一年交易數據所建立的預測模型，並透過實際的資料驗證預測模型的效果。母親節檔期設定為母親節的前兩週，以單一門店當年度母親節檔期前的所有資料作為建模資料，母親節檔期有回購作為預測目標。建模數據(Training Data)為 01/01/2017 到 05/13/2017(母親前最後一日)的交易資料，總計為 41 萬筆；驗證數據(Test Data)為 01/01/2018 到 05/14/2018 的交易資料，總計約為 50 萬筆。

我們以回歸(regression)分析與建立決策樹的方法，透過數據建立會員回購的預測模型。先以第一年的會員消費行為，找出影響回購的因子及建立模型。首先我們找出當年度所有有購買活動的會員，接著再篩選出母親節檔期回來購買的會員，透過當年度一月到四月的消費行為資料，如購買的專櫃品牌、時間、金額等所收集到的所有可能的影響因子導入分析工具中，進一步計算出不同欄位對會員”參與”及”不參與”母親節檔期回購的影響關係，建立此購物中心專屬的母親節回購的分類回歸樹(CART)預測模型。

再將第二年母親節檔期前(01/01/2018 至 30/04/2018)的所有交易數據導入前一年的預測模型，即可預測新的年度所有消費會員於母親節檔期會參與回購的分數，之後我們可以依分數來推算那些會員是高機率會在母親節檔期回購的，分數愈高即代表回購的機會愈大。結合實務應用，我們可以選定一分數標準，高於此分數的即作為預測母親節檔期會回購的預測名單。最後再以實際回購的狀況比對，即可檢驗預測模型在新年度的準確度，並檢視採用技術所發展的模型是否確有可預測性。

# Chapter 4 大數據分析

## 4-1 研究方法及工具

參考數據庫知識發現(Knowledge Discovery in Databases, KDD)流程 [34]，發展預測模型的發展步驟如 Figure 4-1 所示。分為數據收集(Selection)、預處理(Preprocessing)、轉換(Transformation)、數據挖掘(Data Mining)及衡量效益(Evaluation)等步驟。

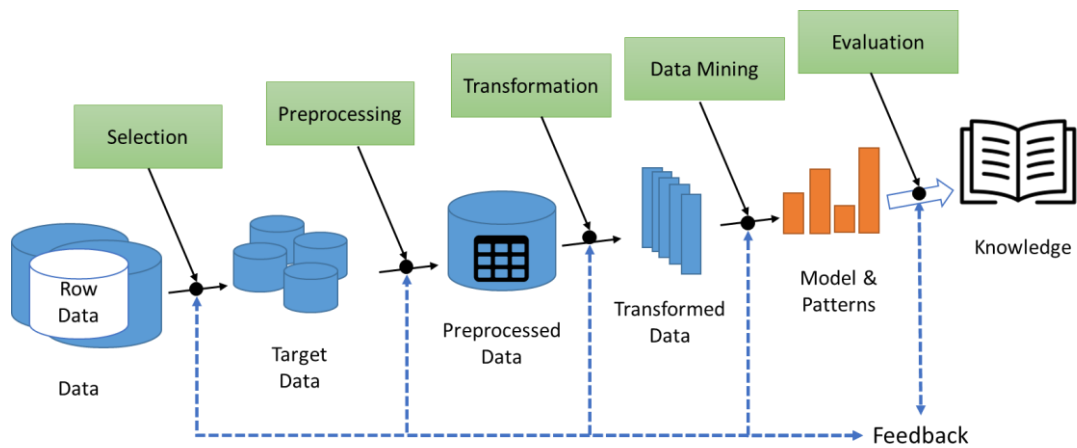


Figure 4-1 數據分析的作業流程圖

各步驟說明如下：

- 數據收集：本次回購模型的規畫，需要掌握門店、會員、專櫃等資訊。客戶使用中的交易系統包含有 ERP、POS 及 CRM，由於涉及多系統的整合，透過建置企業資料倉儲(enterprise data warehouse, EDW)分散在各系統的數據以擷取、轉換和載 (Extract-Load-Transform, ETL)的技術進行彙整。
- 預處理：為確保數據的正確性、一致性及以獲得正確的結果，在建模前會先做前處理，包含格式一致、清洗、補值、排除例外。本次處理過程包含有數據型態的一致，如日期調整為 YYYY/MM/DD；數字型態拿除千分位。性別的部份，調整男、女、F、M 為一致，保留未填寫的類型；縣市、地址等字串型資料，會檢視整體資料筆數及分佈狀況，如有過於分散的問題，為避免



影響分析的效能，就會排除在本次分析之外。Figure 4-2 為預處理後預定進行轉換及建模的數據模型。

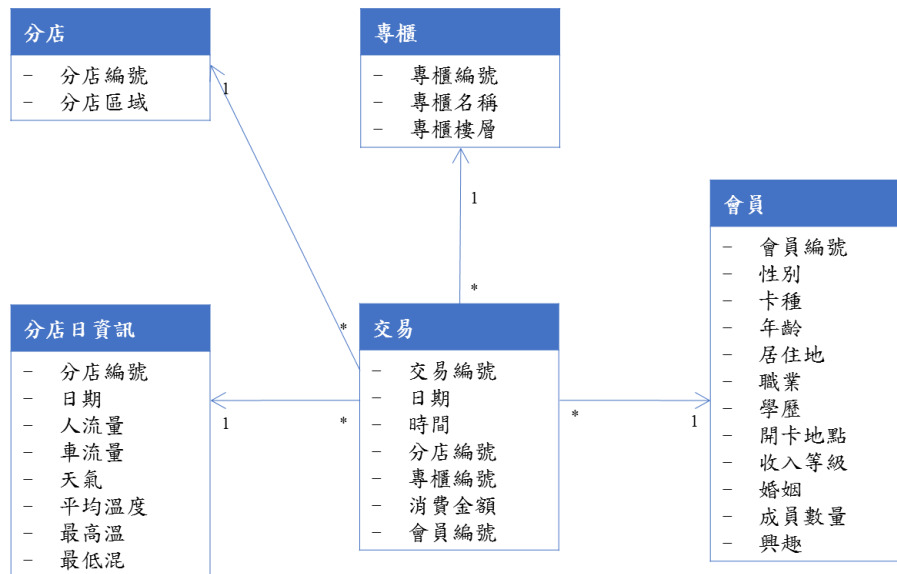


Figure 4-2 數據預處理:Data Model (UML)

- 數據轉換：同時，也可以將一些有用的變量在整理的過程中取得，如隱藏在交易資料當中的客戶行為屬性，包含像有累計金額及消費頻率。本研究在分析前數據處理，主要有透過年度交易歷史數據整理計算”累計消費金額”及”消費次數”；另外，為設定預測模型的目標，在每筆交易數據新增欄為”母親節消費”，標示 母親節活動檔期期間(05/01/2017~05/13/2017) 有來消費的識別。會員行為資料包含有會員、產品及交易數據。本研究使用歷史交易數據建模，收集了2017,2018至母親節檔期結束前的所有會員交易記錄。每筆交易記錄有會員卡號、購買專櫃編號、分店編號及交易時間等資訊，再透過程式合併會員主檔、專櫃主檔及分店主檔的詳細資訊，作為建模分析的基礎。
- 建立模型：本論文運用決策樹方法來建立客戶的回購預測模型，並透過混淆矩陣衡量模型的準確度狀況。本研究使 SAPPA (SAP Predictive Analytics) 工具建立模型，採用多元回歸分析的穩健回歸分析法(robust regression

analysis)，並透過結構風險最小化(structural risk minimization)的演算法 [35]，自動生成模型，幫助減化人工建模、驗證及可視化的分析工作。 [36]

- 衡量商務應用：建立預測模型與行銷決策的關係，加入平均客單價及獲客成本等因素，依不同預測準確度對行銷活動投放效益進行模擬分析，進而協助行銷人員快速決策。

## 4-2 決策樹模型的建立

導入 2017 1 月至 5 月整理後的交易數據，導入所有數據模型變數作為建模的基礎。Figure 4-3 裏，如 A 區所示，可以看到先前收集及預處理的所有變數，都會出現左邊 Explanatory Variables Selected 的選單中。我們預測的目標為參與員親節活動，如 B 區選擇”母親節消費”的變數作為建模的目標變數。接下來我們要排除建模過程程式不用考慮的變數，協助模型的建立更快速及更符合我們的期望。如 C 區所示，將已知”會員編號”、”日期”及”消費年度”等變數不具可預測性，可以透過設定至 Excluded Variables，即可在建模的過程被排除；我們希望未來的預測模型不受大小月份的狀況影響，也透過設定的方式將”消費年月”排除。最後，我們先不做預做任何因子影響的判斷，如 D 區所示，不調整變數權重後即啟動執行，PA 即自動運算分析及建立預測模型。

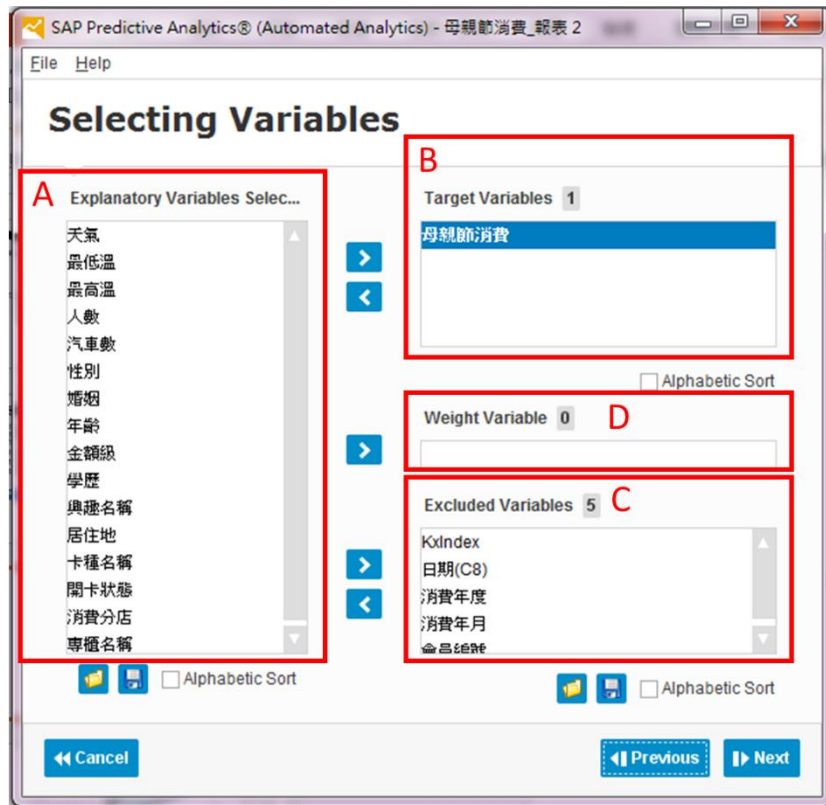


Figure 4-3 系統操作 SAP PA – Variable 選擇

決策樹演算法會以樹狀的方式結構化的計算出各因子所位於的影響層級，並計算出各因子的數值(或值距)之個別權重，作為未來回購的預測模型。運算完成後，SAPPA 會產出之預測模型。決策樹有較佳的可讀性，我們可透過 Figure 4-4 系統操作 SAP PA – 決策樹模型-變數貢獻度分析觀察到透過數據的運算，不同變數影響消費者於活動期間回購的影響度不同。首先是溫度，第二是天氣(晴/雨)，而第三是專櫃的類型。

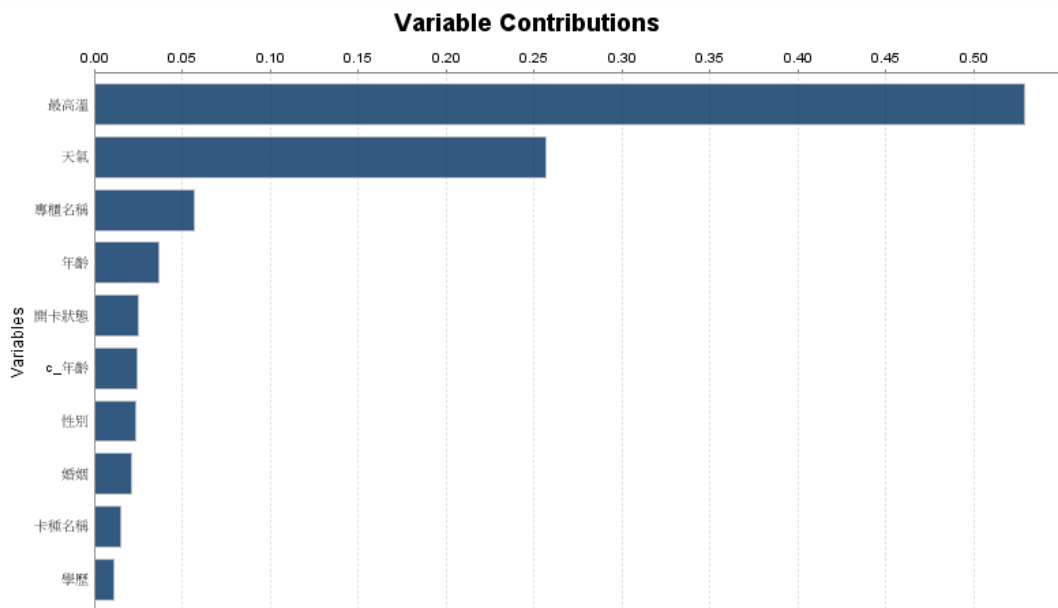


Figure 4-4 系統操作 SAP PA –決策樹模型-變數貢獻度分析

Figure 4-5 所展開樹的第一階與第二階，我們用視覺化樹的型式觀察分析的全數資料(Whole Population)的 41 萬筆資料，可依溫度狀況細分在不同的節點。往下展開，我們可以選擇其他的貢獻因子，包含天氣、專櫃及卡種等。參考 Figure 4-6 系統操作 SAP PA – 決策樹呈現 Level 1 & 2 & 3 Figure 4-6，可以看到增加專櫃之後的節點分配狀況。

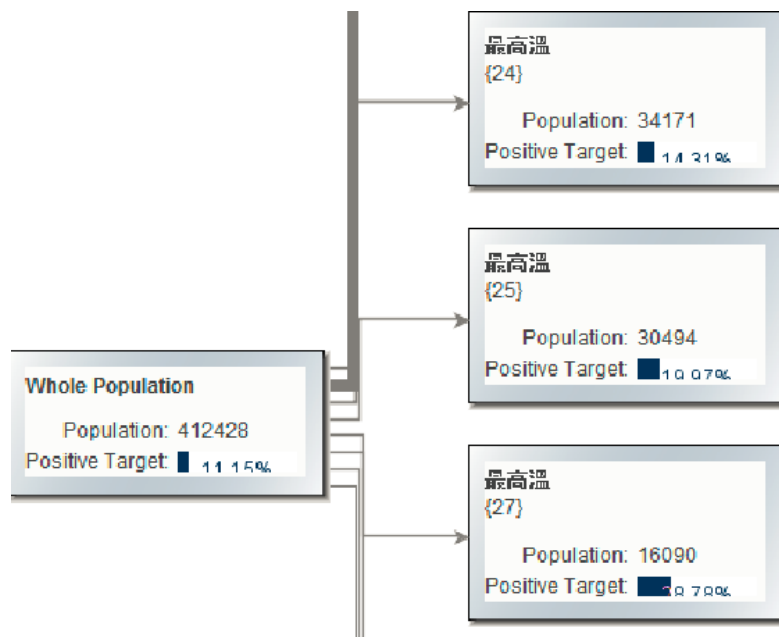


Figure 4-5 系統操作 SAP PA –決策樹呈現 Level 1 & 2

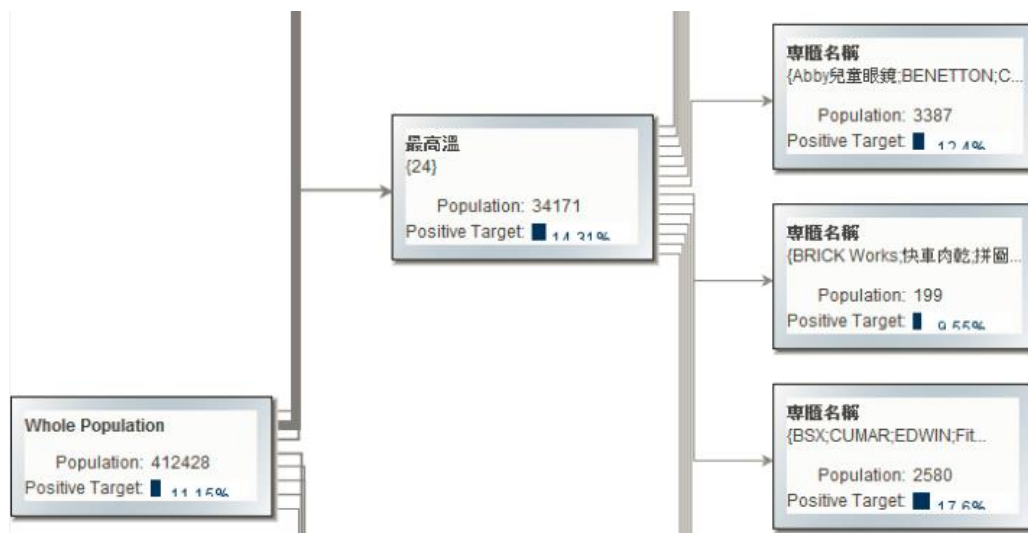


Figure 4-6 系統操作 SAP PA – 決策樹呈現 Level 1 & 2 & 3

工具提供數據品質檢視及可視化的功能，可將建模的狀況以 ROC Curve 作視覺的展示。Figure 4-7 為預測分析的績效及品質，取樣到 7.5% (約 3 萬筆資料)，TPR 可到 68.56%；取樣到 15% (約 6 萬筆) 時，TPR 即可到 100%。依藍色面積可辨識， $AUC > 0.5$  表示此模型具有預測價值。



Figure 4-7 系統操作 SAP PA – 預測模型的績效與品質

如 Figure 4-8 所示，首先以 2017 年數據隨機抽樣 25% (103,500)作為混淆矩陣驗證。其具體指標如下：

- Accuracy(準確性) = 99% - (TP+TN) / Total N
- Precision(精確性) = 99% - (TP+TN) / Total N

Confusion Matrix

	Predicted Y (11695)	Predicted KxMissing (9...
True Y (11589)	11589 11.2%	0 0%
True KxMissing (91911)	106 0.1%	91805 88.7%

Total Population: 103,500

Figure 4-8 系統操作 SAP PA – 2017 年隨機抽樣 25%的混淆矩陣

## 4-3 數據驗證

本案例之購物中心在 2017 年 1 到 4 月的活躍會員數量共有 12 萬人，在 2017 年的母親節檔期參與人數為 2.35 萬人次，回購率則為 18%。到了 2018 年，本案例之購物中心活躍會員數量來到 13 萬人，母親節檔期回購消費的會員數提升到 2.4 萬人次，回購率則與 2017 維持相同的 18%。Figure 4-9 為 2017 與 2018 年母親節檔期本案例之購物中心會員實際回購統計比較。

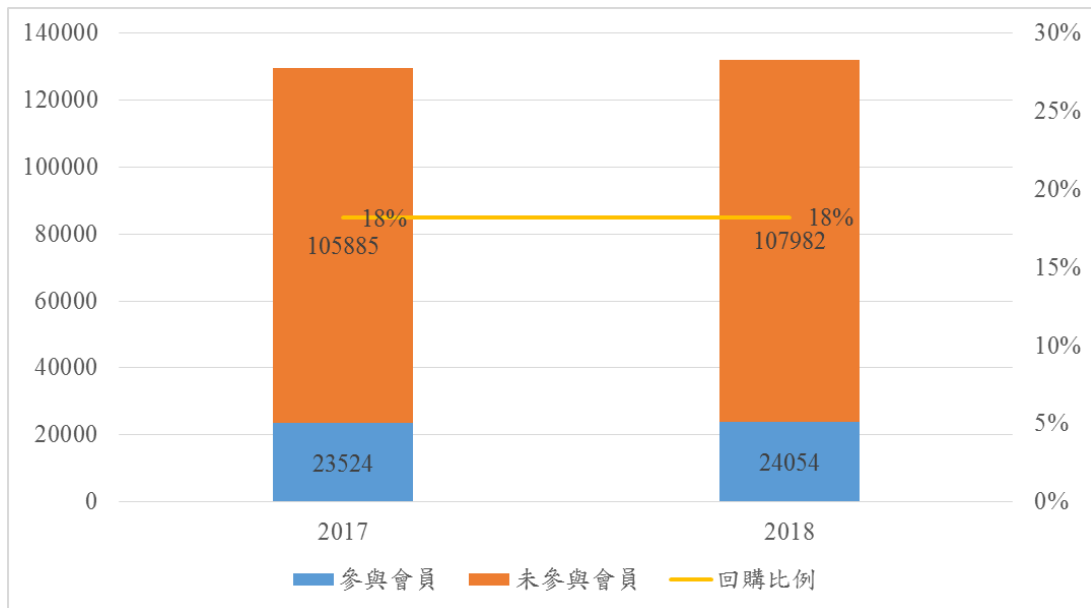


Figure 4-9 2017-2018 母親節檔期會員回購統計

為驗證預測模型成效，我們以 2018 年實際銷售數據驗證：

- 期間：01/01/2018 至 05/14/2018
- 整體：交易 50 萬筆，年度會員 132,036 人。
- 母親節：交易 55,245 筆，參與會員 24,054 人。

將 2018 年的交易數據用前述建立之決策樹模型計算，將依會員及交易計算出分數(Score)，分數愈高代表該會員回購機率愈高。首先以 Score > 0 作為預測會來的會員，整體的預測成果與實際發生之數據作合併比對，如**錯誤！找不到參照來源**。所示並以混淆矩陣檢視預測成果：

表 4-1 2018 預測與實際資料混淆矩陣 (Score > 0)

Confusion Matrix		預測		
		Y	N	小計
實際	Y	7129 (7.96%)	16925 (18.90%)	24054
	N	2776 (3.10%)	62739 (70.05%)	65515
小計		<b>9905</b>	<b>79664</b>	<b>89569</b>

從表 4-1 中可以觀察到：

- 預測回購人數：9905，建議寄送的名單數量
- 預測不回購人數：79664，建議不寄送的名單數量
- Accuracy(準確性) = 78.01%，整體模型的準確性
- Precision(精確性) = 71.97%，建議預測名單準確性

與全面寄送的方式相比，預測名單的數量為 9905 位，建議預測名單的的精確性為 71.97%，較實際採用全面寄送命中率的 18% 提升了 4 倍。透過 Figure 4-10 可以作全面寄送與透過預測模型計算的比較。再以每個會員製作及寄送 DM 費用(獲客費用)20 元計算，採用預測模型的名單的方式可降低 244 萬。Figure 4-11 為全面觸及使用預測模型(Score>0)的獲客費用與回購率比較。

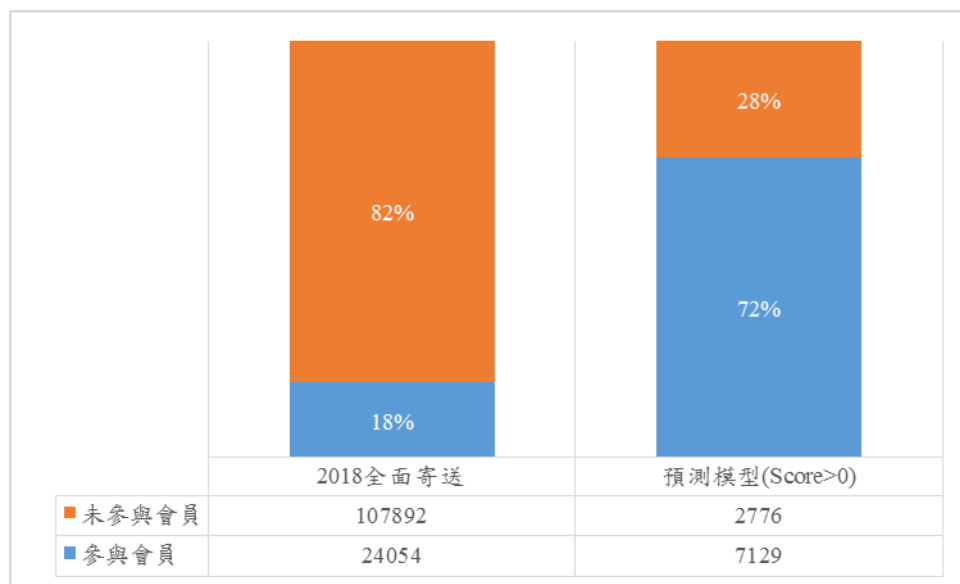


Figure 4-10 回購人數與比例分析



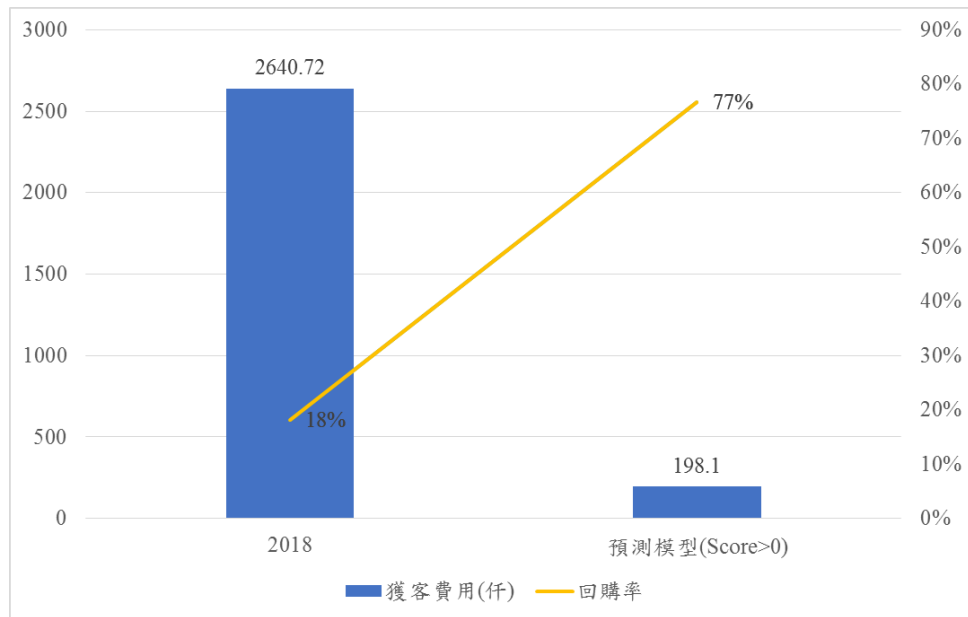


Figure 4-11 獲客費用與回購率

進一步假設隨會員回購的分數不同，實際回購的準確度應會隨著分數的提升而提升。採用分數愈高，標準愈高，預測回購的名單數量會減少，準確度應提升；採用分數愈低，標準下降，預測回購的名單數會增加，但整體準確度也下降。本論文將回購的分數分別以  $\text{Score} > 0.1$  及  $\text{Score} > -0.1$  的兩個級別再做比較，整理篩選出不同名單數量，並用混淆矩陣檢視成果。

- $\text{Score} > 0.1$

表 4-2 2018 預測與實際資料混淆矩陣 ( $\text{Score} > 0.1$ )

Confusion Matrix		預測		
		Y	N	小計
實 際	Y	4442	19612	24054
	N	768	64747	65515
小計		<b>5210</b>	<b>84359</b>	<b>89569</b>

- 預測回購人數：5210，建議寄送的名單數量
- 預測不回購人數：84359，建議不寄送的名單數量
- $\text{Accuracy}$ (準確性) = 77.25%，整體模型的準確性
- $\text{Precision}$ (精確性) = 85.26%，建議預測名單準確性

如表 4-2 所整理，採用 Score>0.1 的名單與全面寄送的方式相比，預測名單的數量下降為 5210 位，建議預測名單的準確性為 85.26%，較實際全面觸及命中率的 18% 提升了 4.73 倍；與 Score >0 相比，預測名單數量減少了 4695 位，準確性下降了 1%，而精確性提升了 13%。名單數量減少近 90%，整體準確性差距不大，但精確性大幅提升。

- Score > -0.1

表 4-3 2018 預測與實際資料混淆矩陣 (Score >-0.1)

Confusion Matrix		預測		
		Y	N	小計
實 際	Y	9698	14356	24054
	N	15702	49813	65515
小計		<b>25400</b>	<b>64169</b>	<b>89569</b>

- 預測回購人數：25400，建議寄送的名單數量
- 預測不回購人數：64169，建議不寄送的名單數量
- Accuracy(準確性) = 66.44%，整體模型的準確性
- Precision(精確性) = 38.18%，建議預測名單準確性

如表 4-3 所整理，與全面寄送的方式相比，預測名單的數量加大為 25400 位，建議預測名單的精確性下降為 38.18%，雖較實際全面觸及命中率的 18% 仍提升了 2.12 倍；但與 Score >0 相比，預測名單數量增加了 15495 位，準確性下降了 12%，而精確性也下降了 32%。名單數量大增，準確性與精確性大幅下降。

為便於比較，我們更進一步將三個等級放在一起。如表 4-4 的整理，我們依分數等級合併比較的預測人數、準確性及精確性。

表 4-4 分數等級與精準性比較

Score	>0.1	>0	>-0.1
實際參與會員人數	24054		
預測人數(名單)	5210	9905	25400
Accuracy(準確性)	77%	78%	66%
Precision(精確性)	85%	72%	40%
命中會員數量	4442	7129	9698

比對 2018 年實際回購的會員人數 24054 人，參考 Figure 4-12 用圖形的方式呈現各分數級別的差異。

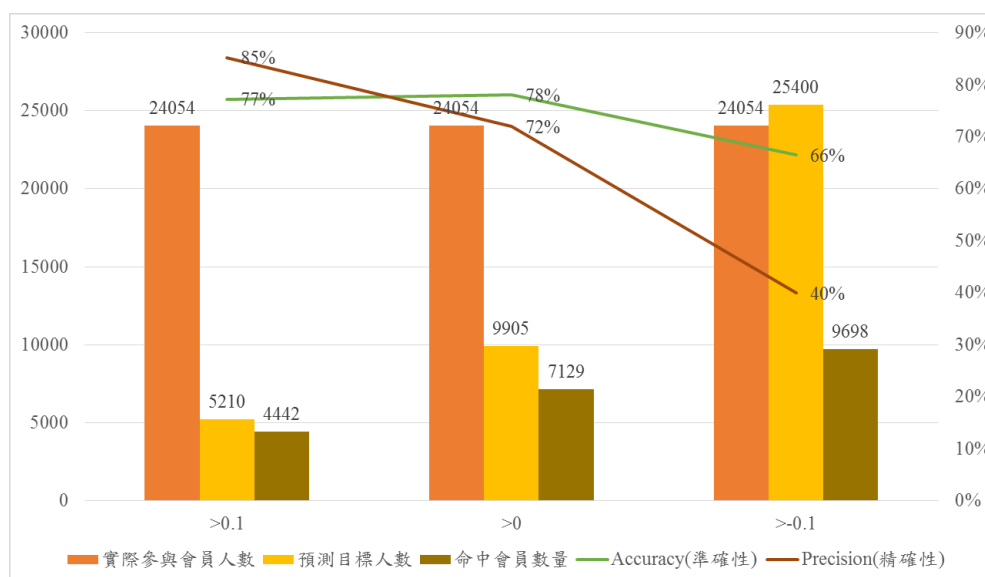


Figure 4-12 分數等級與精準性比較

透過實際數據的驗證，我們掌握了預測模型在不同分數下的準確性及精確性績效，並且進一步思考如何協助市場行銷的決策。

## 4-4 實務應用

行銷人員在應用上，不一定能理解預測模型的計算方式、品質及產出模型成果。在做行銷活動決策的重點會放在預計花多少錢？會得到多少顧客？；如果預

算增加或減少，又會對活動產生之影響等。為了能讓行銷人員更容易應用本研究之預測模型，我們將預測的會員人數及準確率，進一步與營收、毛利及費用合併分析模擬。

結合本次預測的三個分數等級，表 4-5 為比較表，並用 Figure 4-13 作視覺化呈現。

表 4-5 分數等的成本效益試算比較

Score	> 0.1	> 0	> -0.1
命中會員營收(仟)	1,777	2,852	3,879
命中會員毛利(仟)	444	713	970
獲客成本(仟)	104	198	508
命中會員淨利(仟)	340	515	462
淨利費用比	4.26	3.60	1.91

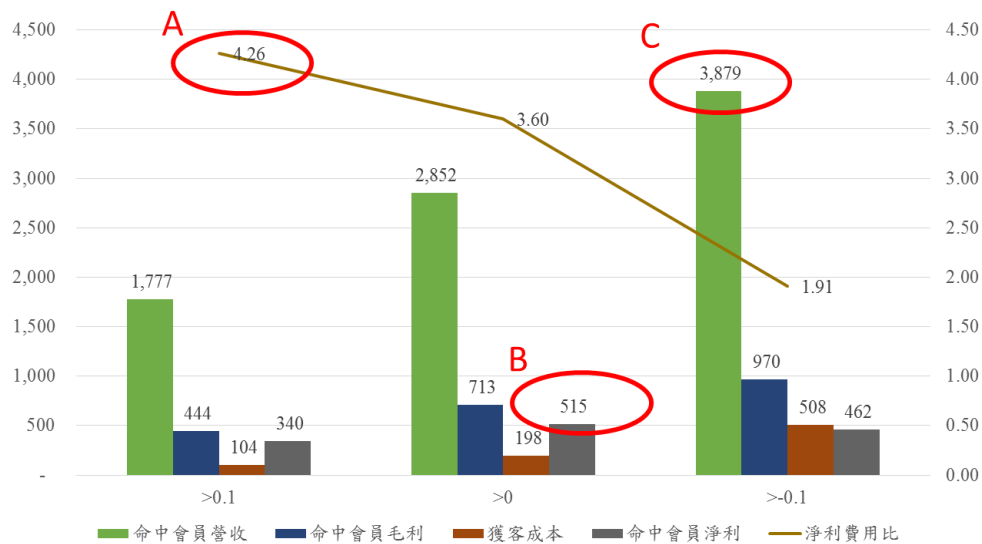


Figure 4-13 分數等的成本效益試算比較

- Score > 0.1：參考 Figure 4-13 分數等的成本效益試算比較的 A 區所示，淨利費用比 4.26 為三組中的最高。相較於全面客戶觸及，此分數模型目標在追求好的獲利效率(淨利費用比)，追求控制費用以獲取高度的回客率。適合用於成熟穩定的活動類型。

- $\text{Score} > 0$  : 參考 Figure 4-13 分數等的成本效益試算比較的 B 區所示，淨利金額 515(仟) 為三組中的最高。此分數模型以總體淨利為目標，同時放寬成長及獲利的增長，適合用於年度重大活動。
- $\text{Score} > -0.1$  : 參考 Figure 4-13 分數等的成本效益試算比較的 C 區所示，命中會員營收金額 3,879(仟)為三組中最高。為提高營收，放大會員預測的名單數量，雖可提命中的名單總數，進而拉高營收，但隨著命中率的下降，獲客成本增加也影響最終淨利。

行銷人員未來在做活動規畫時，可以因應不同的策略選擇採用不同的分數等級。大檔期的活動會計算 DM 的邊際效益，極大化的觸及客戶，追求費用控制下的獲利最大化；而大檔期前後的活動的目標放在獲客效率的最大化(成本低，回客率高)；如果是新產品及新體驗，則會以獲利及費用可平衡的目前下，最大化的規畫觸及的會員名單。

# Chapter 5 結論

## 5-1 結論

隨著大數據與機器學習等技術日趨成熟，如何將之應用在零售行業中，是頗具商業價值的發展方向。本研究以台灣大型購物中心二個年度的銷售資料，先作系統化的數據整理，再應用數據分析技術建立回購預測模型，最後以第二年的實資料驗證模型的成效。實証結果証實以下發現：

- 購物中心的會員活動回購有預測性，使用第一年的資料建模，第二年的數據驗證。Accuracy(準確性)=78.01%，Precision(精確性)=71.97%，優於採用全面派送的回購率。
- 透過三組 Score 預測模型與模擬分析的整合，進一步衡量預測準確率、營收及淨利等關係。營銷人員在活動前，依不同活動、預算及業績目標，選擇推進方案取得推薦客戶名單，實現事前模擬行銷活動的投放成效。
- 預測模型的準確度已不是衡量的惟一考量，實務應用時，準確性及精準性的影響只是決策的一環。以本次研究來說，對營銷人員重視的是成本、來客數、回購比例及毛利費用比，同時需考量不同活動的不同訴求，才有最適合選擇的方案。
- 未來可延用本研究的手法，透過掌握銷售資料，建立預測回購模型，持續應用回購模提供活動決策的參考。

## 5-2 限制與展望

- 本研究假設外在環境及產業無劇烈變化，也沒有巨大競爭對手或事件影響市場。透過歷史資料建立模型並進一步預測未來會員的回購機會。
- 本研究有採用會員及交易資料作為驗證模型的準確性及精確性。未來如有更多活動資料(如活動特徵、名單)及會員行為資料(如在店時間)，即可驗證不同的因子對回購行為的影響程度，進一步提高模型的準確性。

- 然而零售業範圍大且各有特色，對應不同商品、客群會而會有不同的經營模式，不能直接將複用本研究的預測模型。但可透過本研究所採用決策樹模型及驗證手法，找出企業自身專屬的回購模型，為企業提供更有效的行銷策略規畫參考。

## 參考文獻

- [1] V. & R. T. Mayer-Schönberger, “Reinventing Capitalism in the Age of Big Data.,” Basic Books., 2018.
- [2] E. W. X. L. & C. D. C. Ngai, “Application of data mining techniques in customer relationship management: A literature review and classification.,” Expert systems with applications, 2009.
- [3] R. S. Swift, “Accelerating customer relationships: Using CRM and relationship technologies,” Prentice Hall Professional, 2001.
- [4] S. & N. J. Senecal, “The influence of online product recommendations on consumers,” Journal of retailing, 2004.
- [5] M. L. B. & D. L. R. J. L. Montaner, “A taxonomy of recommender agents on the internet.,” Artificial intelligence review, 2003.
- [6] B. K. G. K. J. & R. J. Sarwar, “Item-based collaborative filtering recommendation algorithms,” In Proceedings of the 10th international conference on World Wide Web, 2001.
- [7] B. K. G. K. J. & R. J. Sarwar, “Item-based collaborative filtering recommendation algorithms.,” In Proceedings of the 10th international conference on World Wide Web.ACM., 2001.
- [8] M. & S. Y. Balabanović, “content-based, collaborative recommendation.,” Communications of the ACM, 1997.
- [9] M. J. & B. D. Pazzani, “Content-based recommendation systems.,” Springer,, 1998.
- [10] D. & R. A. Loshin, “Using information to develop a culture of customer centricity: customer centricity, analytics, and information utilization,” Elsevier, 2013.
- [11] “Bring Big Data to the Enterprise,” IBM, What Is Big Data, 2012.
- [12] X. Z. X. W. G. Q. & D. W. Wu, “Data mining with big data,” IEEE transactions on knowledge and data engineering, 2014.
- [13] C. C. Aggarwal, “Data mining: the textbook,” Springer, 2015.
- [14] J. J. P. a. M. K. Han, “Data mining: concepts and techniques,” Elsevier, 2011.
- [15] C. a. O. P. Giraud-Carrier, “Characterising data mining software.,” Intelligent Data Analysis, 2003.
- [16] R. a. D. C. Y. Ling, “ustomer relationship management: An analysis framework



- and implementation strategies.,” *Journal of computer information systems*, 2001.
- [17] E. A. J. E. L. T. P. & S. R. Turban, “*Businesss Intelligence and Analytics: Systems for Decision Support*,” Prentice Hall, London, 2007.
- [18] R. H. & M. R. H. Myers, “*Classical and modern regression with applications*,” Belmont, CA: Duxbury press., 1990.
- [19] P. J. & L. A. M. Rousseeuw, “*Robust regression and outlier detection*,” John wiley & sons., 2005.
- [20] J. B. P. L. W. R. C. & A. M. Shawe-Taylor, “*Structural risk minimization over data-dependent hierarchies.*,” *IEEE transactions on Information Theory*, 1998.
- [21] B. F. J. H. O. R. A. & S. C. J. Leo, “*Classification and regression trees.*,” Wadsworth International Group., 1984.
- [22] A. E. B. D. A. D. D. S. & E. J. M. Clarke, “*Decreasing costs and improving outcomes in systemic lupus erythematosus: Using regression trees to develop health policy.* *Journal of Rheumatology*,” *Journal of Rheumatology*, 1984.
- [23] Y. H. K. J. K. & K. S. H. Cho, “*A personalized recommender system based on web usage mining and decision tree induction*,” *Expert systems with Applications*, 2002.
- [24] S. & T. T. Taewijit, “*A Probabilistic Approach for Customer Repurchase Intention and Association-Based Product Network Analysis: An Empirical Study of Fast Food Market*”.
- [25] L. S. L. M. R. & C. J. R. Chen, “*A Decision Tree Based Method for Extracting Important Elements of In-Applications Purchase.*,” 2016 Third International Conference on. IEEE, 2016.
- [26] S. Khan, “*Determination of customer retention in hotel industry.*,” *Journal of Applied Economics and Business* 1, 2013.
- [27] L. H. C. K. K. & C. R. H. Lin, “*Predicting customer retention likelihood in the container shipping industry through the decision tree approach*,” *Journal of Marine Science and Technology*, 2017.
- [28] S. R. B. R. A. L. & V. D. K. E. Visa, “*Confusion Matrix-based Feature Selection*,” MAICS, 2011.
- [29] E. R. D. D. M. & C.-P. D. L. DeLong, “*Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach.*,” *Biometrics*, 1988.
- [30] E. & H. J. B. S. Quah, “*Cost-benefit analysis.*,” Routledge., 1970.

- [31] M. & C. S. Morimoto, "Consumers' attitudes toward unsolicited commercial e-mail and postal direct mail marketing methods: intrusiveness, perceived loss of control, and irritation.," *Journal of Interactive Advertising*, 2006.
- [32] 蔡杉源, "百貨零售業體驗行銷策略運用之探索性研究：以漢神百貨開店慶暨母親節活動為例," 2003.
- [33] 張婷婷, "百貨零售業特定檔期市場區隔策略分析：以 A 百貨為例," 2015.
- [34] U. P.-S. G. & S. P. Fayyad, "From data mining to knowledge discovery in databases," *AI magazine*, 1996.
- [35] H. Edelstein, "KXEN and the ABCs of SVMs.," *Information Management*, 13, 2003.
- [36] M. & M. E. P. Bera, "US6523015B1. Robust modeling," U.S. Patent No. 6,523,015. Washington, DC: U.S. Patent and Trademark Office, 2003.
- [37] A. H. D. Q. M. a. D. S. Kracklauer, "Customer management as the origin of collaborative customer relationship management.," Springer, Berlin, Heidelberg, 2004.
- [38] M. C. C. A. L. & C. H. H. Chen, "Mining changes in customer behavior in retail marketing," *Expert Systems with Applications*, 2005.
- [39] M. J. & B. D. Pazzani, "Content-based recommendation systems," Springer, 2007.
- [40] J. O. F. H. A. & G. A. Bobadilla, "Recommender systems survey.," *Knowledge-based systems*, 2013.