

# 致理技術學院

商務科技管理系

實務專題報告



資料探勘應用於消費者購買行為之研究

—以致理幸福農學市集為例

指導老師：高楊達老師

學生：黃育紘(10033128)

吳貫玉(10033122)

洪秀(10033157)

中華民國 103 年 12 月

# 致理技術學院

商務科技管理系

實務專題報告

資料探勘應用於消費者購買行為之研究  
——以致理幸福農學市集為例

學生：黃育紘(10033128)

吳貫玉(10033122)

洪秀(10033157)

本成果報告書經審查及口試合格特此證明。

指導老師：

中華民國 103 年 12 月

# CTM 實務專題研究授權書

本授權書所授權之實務專題研究為吳貫玉、黃育紘、洪秀共 3 人，在致理技術學院商務科技管理系 103 學年度第 1 學期完成商管實務專題。商管實務專題名稱：資料探勘應用於消費者購買行為之研究—以致理幸福農學市集為例

同意 不同意本組同學共\_\_\_\_\_人，皆同意著作財產權之論文全文資料，授予教育部指定送繳之圖書館及本人畢業學校圖書館，為學術研究之目的以各種方法重製，或為上述目的再授權他人以各種方法重製，不限地域與時間，惟每人以一份為限。

上述授權內容均無須訂立讓與及授權契約書。依本授權之發行權為非專屬性發行權利。依本授權所為之收錄、重製、發行及學術研發利用均為無償。上述同意與不同意之欄位若未鉤選，該組同學皆同意視同授權。

指導教授姓名：

專題生學號簽名：

學號：

中華民國 103 年 12 月 18 日

# 致謝

首先，非常感謝在這一年多帶領及指導我們的高楊達老師，跟著老師一起加入了致理幸福農學市集這個團隊，籌備了 10 多次的市集活動，其中包含了活動策劃、顧客關係管理、產品包裝以及行銷策略，這段期間，老師由領導者轉為輔助者，讓學生由被領導轉為主動思考，提供自己的想法與市集團隊一同討論，提升團隊整體的組織能力，使我們一次又一次的成長，從錯誤中學習，藉由實務的經驗，在未來企業的環境中更有競爭力。

很高興能夠加入商務科技管理系這個大家庭，讓我們學習到多元的知識，將專業知識應用於致理幸福農學市集，專題的研究過程中，讓我們對資料探勘有了更深入的瞭解。最後，感謝致理幸福農學市集所有團隊成員，因為有你們的幫助，讓我們的畢業專題能順利進行，且如期完成，謝謝你們，有你們真好。

# 摘要

隨著消費者在挑選商品上的多樣性，如何抓住消費者的心，找出不同消費族群喜愛的商品，對於企業是一大重要課題。資料探勘的技術可以從大量的資料庫中，獲得銷售產品的排行榜、銷售總金額，找出消費者的特徵，得知潛在顧客及消費者的購買行為，以利於企業之行銷策略之用，藉以滿足顧客需求。

本研究以「致理幸福農學市集」為研究對象，將市集的會員基本資料及交易資料，經過資料庫正規化後，透過 SQL Server 進行資料探勘，以 Apriori 演算法產生相依性網路圖，區分不同類型的消費族群，找出會員所喜愛的產品和可能會喜歡的產品等潛在資訊。因此可以針對不同的會員，進行商品、活動的推薦，或是有新商品推出時，可以優先通知哪一些會有興趣的消費者，讓市集與消費者更有互動，創造更多的商業行為。

關鍵詞：資料探勘、消費者購買行為、關聯法則、Apriori、農學市集

# 目錄

CTM 實務專題研究授權書 .....	i
致謝 .....	ii
摘要 .....	iii
目錄 .....	iv
圖目錄 .....	v
表目錄 .....	vi
第一章 緒論 .....	1
第一節 研究動機 .....	1
第二節 研究目的 .....	1
第二章 文獻探討 .....	2
第一節 市集定義 .....	2
第二節 資料探勘 .....	4
第三節 消費者購買行為 .....	10
第四節 顧客關係管理 .....	13
第三章 研究方法 .....	15
第一節 研究設計與架構 .....	15
第二節 資料庫設計 .....	15
第三節 Apriori 演算法 .....	18
第四節 實作工具(SQL Server).....	20
第四章 研究發現 .....	21
第一節 個案研究-以致理幸福農學市集校內教職員團購為例.....	21
第二節 個案研究-以致理幸福農學市集交易紀錄為例.....	34
第五章 研究結論與建議 .....	36
第一節 結論 .....	36
第二節 後續研究建議 .....	36
參考文獻 .....	37

# 圖目錄

圖 2-1	市場參與者互動模式.....	2
圖 2-2	資料探勘流程圖.....	4
圖 2-3	決策樹範例.....	9
圖 2-4	各類型顧客之行銷策略.....	13
圖 3-1	流程圖.....	15
圖 3-1	教職員團購未經正規化.....	16
圖 3-2	教職員團購正規化結果.....	16
圖 3-3	校內教職員團購 ER-Model 關聯圖.....	17
圖 3-4	實體市集 ER-Model 關聯圖.....	17
圖 4-1	新增資料庫畫面.....	22
圖 4-2	SQL Server 匯入和匯出精靈畫面.....	23
圖 4-3	資料來源與路徑.....	23
圖 4-4	目的地畫面.....	24
圖 4-5	選取來源資料表.....	24
圖 4-6	建立專案.....	25
圖 4-7	新增專案.....	26
圖 4-8	選擇連接伺服器名稱.....	26
圖 4-9	如何定義連接.....	27
圖 4-10	資料來源檢視精靈.....	27
圖 4-11	選取資料來源.....	28
圖 4-12	選取資料表和檢視.....	28
圖 4-13	關聯圖.....	29
圖 4-14	資料採礦精靈.....	29
圖 4-15	選取定義方法.....	30
圖 4-16	關聯規則建立資料採礦結構.....	30
圖 4-17	指定資料表類型.....	31
圖 4-18	指定索引鍵、輸入、可預測.....	32
圖 4-19	建立測試集.....	32
圖 4-20	教職員團購相依性網路圖.....	33
圖 4-21	於相依性網路圖中點選 A18-04.....	33
圖 4-22	市集交易相依性網路圖.....	34
圖 4-23	於相依性網路圖中點選 A27.....	35
圖 4-24	相依性網路圖點選 A50.....	35

# 表目錄

表 2-1 資料庫中的 4 筆交易記錄.....	8
表 2-2 Apriori 演算法計算流程圖.....	8
表 3-1.....	18
表 3-2.....	18
表 3-3.....	19
表 3-4.....	19
表 3-5.....	19
表 3-6.....	19
表 3-7.....	20

# 第一章 緒論

## 第一節 研究動機

未來是大數據(Big Data)的時代，如何將巨大的資料轉換為人們所需要的資訊，會是一件非常重要且具挑戰的事情，我們利用了 SQL Server，將致理幸福農學市集各月的會員交易資料以 Apriori 演算法推導出喜愛 A 類商品的會員會去買 B 類商品的頻率，以後在碰到會員購買 A 類商品時，我們就可以向會員推薦 B 類商品，藉以提高消費者之提袋率。

## 第二節 研究目的

依據上述研究背景與研究動機的闡述，本論文之研究目的為下列幾項：

- 一、研究致理幸福農學市集的消費者購買行為，藉此瞭解會員特徵。
- 二、研究 SQL Server 操作，對致理幸福農學市集之會員交易資料進行資料探勘，進而提出商品組合建議。
- 三、提出致理幸福農學市集後續建議。

## 第二章 文獻探討

### 第一節 市集定義

市場(market)作為一交易場所，最早源自人與人之間的物物交換。《牛津地理學辭典》(mayhem, 2001:269)對市場的定義為：「是買賣貨品和服務的地方，這包括所有便利的安排，使人們可以購買和出售貨品、服務和生產要素，因此沒有一個特定的地點。」下圖為市場參與者的互動模式：

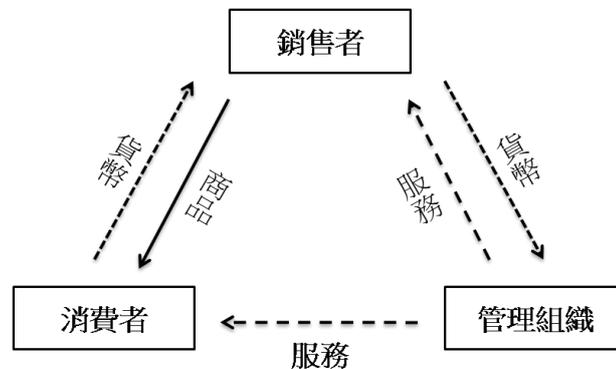


圖2-1 市場參與者互動模式

資料來源：賴鳳雲(2010)

本研究以致理幸福農學市集為例，下面介紹農夫市集與農學市集的相關定義。

#### 一、農夫市集

「農夫市集」亦稱為農民市集或稱綠色市集，簡單來說它是一個由農民所組成，親身販售的交易系統。近年來，綠色生態經濟開始受到關注，有機農業、樂活和環保的意識興起，農夫市集逐漸在台灣發展開來。

農夫市集與傳統市場同為交易場所，卻有所區隔。先以傳統市場來說，其分佈最密集，但是其中生產者的角色並未出現，僅透過銷售者將「匿名」生產者的商品售出賺取利益。傳統市場的管理組織十分消極，通常消費者

不會感受其存在，僅向銷售者收取攤位租金或是負責收攤後的清潔工作。

賴鳳雲(2010)研究中提到農夫市集最大的不同點在於除了商品與服務的提供外，管理組織的角色更強調理念，其向生產端要求生產者即是銷售者，要求生產者遵行無毒農法，要求環境永續；向消費端提供環保、永續的觀念，希望生產與消費更加接近，添加了許多非經濟面的元素。

美國農夫市集協會(The National Association of Farmer's Markets)的定義指出，農夫市集應由農夫(farmers)、種植者(growers)或生產者(producers)所組成，並獨立生產、畜養、採集、釀製、烘培、煙燻、加工等各式產品，在指定的區域裡親自、直接的販售產品給予消費者。

陳嬾伊(2009)研究中提到這種承接了過去了消費系統的新世代市集，具有下列五種特點：

- (一) 農夫親自販售。
- (二) 在地生產在地消費。
- (三) 結合健康與環保。
- (四) 促進社區關係。
- (五) 培植弱勢小農。

## 二、農學市集

林淑玲(2012)研究中提到「農學」意味著生活者要向生產者學習、生產者要跟環境學習；「市集」意味著這個團體將是生產者與生活者的組合，人與人的聚集交會。農學市集簡單地說是從心裡層面上的要求或是身體上的健康來看待農業這件事情，對於土地的一種情感為出發點，想善待土地、自然、生態，一種和諧共處的表現。農學市集的理念是「無毒」、「安全」、「友善」，希望能夠達到消費者和農民之間互相尊重、用心的「友善農業」。本研究的實施場域與農學市集定義上較為相似。

## 第二節 資料探勘

資料探勘(Data Mining)應用了許多統計分析與人工智慧相關的演算法，配合電腦技術自動化，將資料加以分析、歸納與整合，所產生的大量資料中勘查出關鍵資訊，再定義為採礦模型，取得資料所隱含的知識，透過這樣的方式，資料探勘能在龐雜的資料中發掘更多意義與趨勢，有助於決策者知識的取得和輔助決策。

### 一、資料探勘的定義

資料探勘將大量的資料庫進行分析、篩選、統計，挖掘出隱含的意義及關連性，發現出隱藏價值的知識，輔助決策者做重要決策之參考。

資料探勘的過程如圖 2-2 所示：

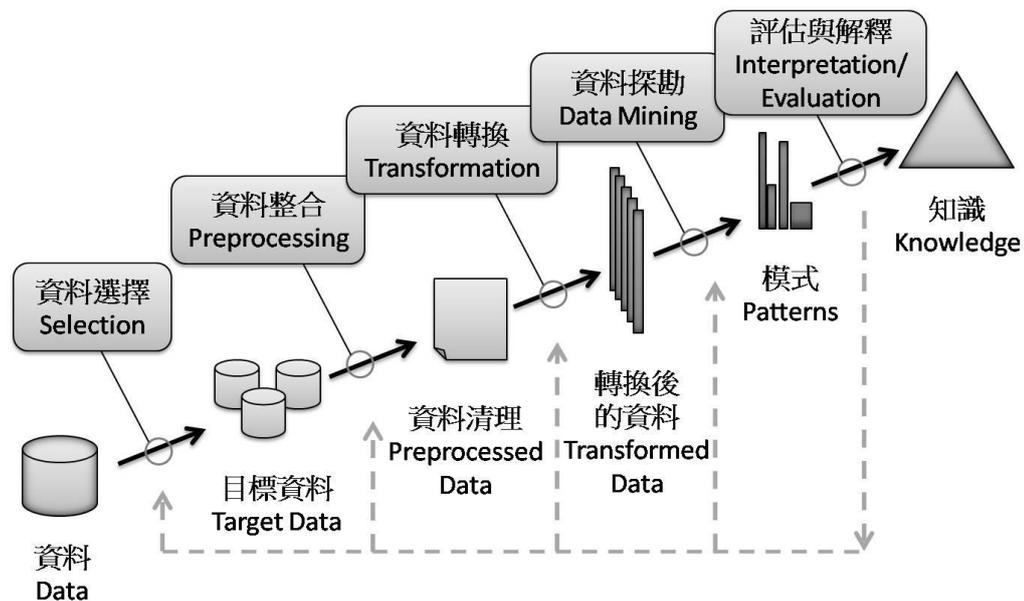


圖2-2 資料探勘流程圖

資料來源：Fayyad(1996)

幾個步驟說明如下：

- (一) 資料選擇：瞭解相關知識以及確定資料探勘的目標資料。
- (二) 資料整合：整合不同來源的資料。

- (三) 資料清理：刪除不需要的資料，選取與研究主題相關的資料。
- (四) 資料轉換：將目標資料轉換成適用目標分析的格式，集合目標資料。
- (五) 資料探勘：應用資料探勘技術，推導資料演算法及分析模式。
- (六) 評估與解釋：評估模型的精確性，將探勘出來的知識於使用者應用。

## 二、資料探勘的功能

一般而言，Data Mining 功能可包含下列五項功能：

- (一) 分類(classification):按照分析對象的屬性分門別類加以定義，建立類組。
- (二) 推估(estimation)：根據既有連續性數值之相關屬性資料，以獲致某一屬性未知之值。
- (三) 預測(prediction):根據對象屬性之過去觀察值來推估該屬性未來之值。
- (四) 關聯分組(affinity grouping)：從所有物件決定那些相關物件應該放在一起。
- (五) 同質分組(clustering)：將異質母體中區隔為較具同質性之群組。

梁衍忠(2012)研究中提到，最佳的決策方法是建構在科學化的研究結果與過程，透過資料探勘的本土化研究做為決策之依據，資料探勘的目的就在於運用廣泛的資訊進行分析與研判，得到的結果加以評估後，根據這些結果進行決策。

許凱嵐(2012)研究中提到，資料探勘可應用的範圍相當廣泛，針對不同的研究分析，必須應用不同的資料探勘方式，才可得到適切的研究結果與知識。隨著資料性質的不同，選擇不同的探勘方式，且在探勘過程中的設

定，對探勘結果的呈現也會有所不同。

宋堯正(2013)研究中提到，資料採礦就是用觀察到的資訊，利用電腦程式模擬後，來得到結果而加以利用與決策，企業內如果能熟練地掌握這個工具，就能快速準確的鎖定問題根因，有效的制定出相對應的行動計畫來解決問題；更深一層的意義就是，從大量資料中進行探索和分析，發現出有意義的模型或規則，並整理成有價值的知識，運用知識管理，傳承給企業內的人，這樣對專業人才的訓練、養成是事半功倍的，且可幫助企業獲取最大的營運績效。

江淑惠(2014)決策者應分析下游的成員需求，並適當組合其資源。資料探勘技術運用於交易資料庫的分析，能協助管理者有效區隔客戶，揭露潛在的管理規則，提升決策品質。

### 三、資料倉儲

在 1993 年 William H.Inmon 將資料倉儲定義為「一個主題導向的、整合的、隨時間變化的、不易失漏性資料的集合，用於支援管理層的決策程序」，其特性為：

- (一) 主題導向 (Subject-oriented)：資料倉儲只儲存與分析主題有關之資料，在將資料轉換為儲存在資料倉儲格式中時，就需將與主題無關之資料清除。
- (二) 整合性 (Integrated)：將來源不同的資料，依統一的格式整合在資料倉儲中，以便於使用者使用。
- (三) 非揮發性 (Nonvolatile)：資料是不需更新的，只需定期或不定期的增加新的資料，但原有的資料是不會改變的。
- (四) 時間變動性 (Time-variant)：資料中包含著明確的時間範圍，而在這時間範圍之資料均可探討其趨向性及變化性，但資料結構組成之成分是不變的。

傳統的資料庫已經無法滿足企業所需，隨著市場競爭的加劇，企業更需要能支援決策的資訊。將日常處理所收集到的各種資料，轉變為具有商業價值的資訊，就是資料倉儲的功能。

郭家禎(2013)資訊科技的進步，各式資料倉儲成本亦隨之降低，企業搜集大量資料不再是相當艱困的任務，但伴隨而來的是大量資料的產生與儲存的速度，遠超過人們所能分析和消化的速度，以至於過去的經驗行銷方式無法快速滿足銀行的需求，應有效運用資料探勘技術，於龐雜資料庫中之隱諱難見、未知或超乎想像等重要資訊。

#### 四、關聯法則

關聯法則(Association Rule)演算法是資料探勘常運用的方法之一，用來分析大量交易資料(Transaction Data)來找出商品項目間的相關性。例如可以在關聯法則運算後，發現有熱門商品 A、B、C，並發現當商品 A 被購買時商品 C 也常常同時被購買，那就可能可以產生一個商品 A 在被購買時會影響到商品 C 被購買機率的關聯法則。關聯法則的演算是根據既有的事實紀錄資料，可以進行對未來發展趨勢的預測推論，梁衍忠(2012)。

從多筆交易資料或多個購物車當中找出經常共同出現的商品項目，並分析這些被找出的項目之間可能存在的交互影響使用關聯法則能分析出大量資料中項目之間，原本沒被發現的潛在相互關聯性。范瑋宸(2011)。

關聯法則中最著名及最常用的演算法為 Apriori 演算法。

Agrawal 等人(1993)提出 Apriori 演算法，之後陸續有人有 Apriori 演算法相關的研究被提出。Apriori 演算法包含兩個部分:產生大型項目集與元素項目集。一個項目集中包含 K 個項目稱為 K-項目集(K-itemset)。項目集出現在資料庫中的交易次數均大於或等於最小支持個數稱為大型項目集，Lk

表示所有大型  $K$ -項目集所形成的集合。 $C_k$  表示所有候選的  $K$ -項目集集合，當  $C_k$  經過最小支持個數過濾則形成  $L_k$ 。

常世杰(2014)研究中提到在不同購物環境下進行 Apriori 演算模型分析，藉此得到正確的關聯規則信賴度閾值設限，管理者將可獲得較高且極為穩定的資料探勘預測結果，進而對未來在市場上的商品組合作出正確決策。

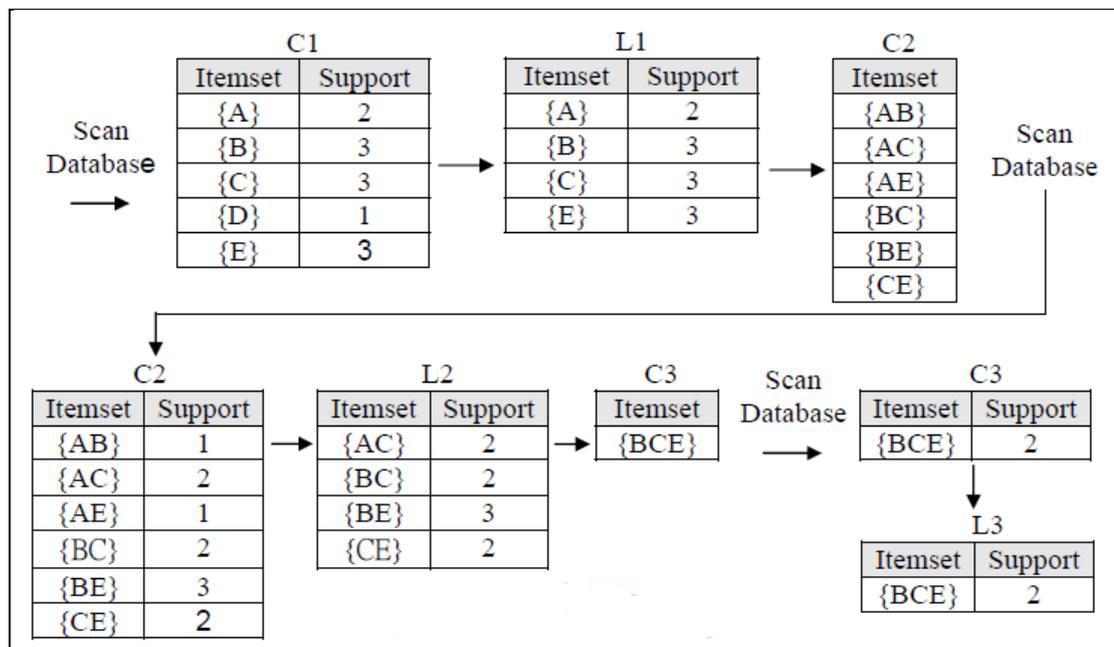
表 2-1、2-2 為 Apriori 演算法之範例，表 2-1 假設資料表中包含四筆交易資料，表示消費者的購買細項。

表 2-1 資料庫中的 4 筆交易記錄

ID	Items
001	ABCE
002	BE
003	ACD
004	BCE

根據表 2-1 的交易紀錄做 Apriori 演算法計算，流程如下：

表 2-2 Apriori 演算法計算流程圖



資料來源：張家蓁，資料採礦應用於消費者網路團購因素探勘之研究(2010)

## 五、決策樹(Decision Tree Model)

決策樹是一個預測模型；他代表的是對象屬性與對象值之間的一種映射關係。樹中每個節點表示某個對象，而每個分叉路徑則代表的某個可能的屬性值，而每個葉結點則對應從根節點到該葉節點所經歷的路徑所表示的對象的值。決策樹僅有單一輸出，若欲有複數輸出，可以建立獨立的決策樹以處理不同輸出。數據挖掘中決策樹是一種經常要用到的技術，可以用於分析數據，同樣也可以用來作預測。

一個決策樹包含三種類型的節點：

- (一) 決策節點：通常用矩形框來表示，用來了解特定族群的特徵，做為決策判斷的依據。
- (二) 機會節點：通常用圓圈來表示，表示在其中指定的結果根據概率確定的點的決策分析中的元素。
- (三) 終結點：通常用三角形來表示，為決策的最後結果。

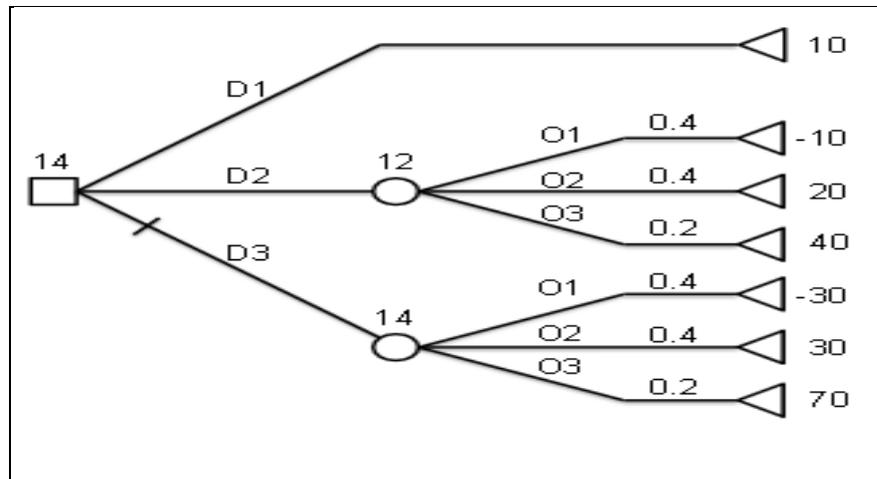


圖2-3 決策樹範例

資料來源：博鴻電子報

宋堯正(2013) 以 Decision Tree Model 分析出來的規則，其簡單易懂的樹狀圖結構讓人一目瞭然。在做 Decision Tree 資料模型可以先使用 Association Rule 來找出一些可能的關連或互相影響的因子，用來提高資料模型預測之準確性。

### 第三節 消費者購買行為

當消費者產生購買行為的動機時，會先滿足基本需求，基本需求滿足之後，會再試圖滿足更高階的需求。如果我們能了解各個消費者的需求、購買的動機，進而推出他們需要的產品，進行行銷活動，就可以大幅提高客單價。

Kotler&Scheff (1996) 將購買決策分為五種過程，並提出購買心態在實際購買前行動已經開始，並且於購買後仍然未結束。

#### (一) 確認需要：

決定購買心態需開始於消費者感到需要或問題的產生，準購買者透過外在的某些因素刺激，並在心理、生理感覺到，當需求狀態出現時，則產生購買衝動的驅使力。

#### (二) 搜集資訊：

消費者受到刺激後會對其事物提高關注力，可能透過各種方式和管道搜集必要的資訊，其驅使力的強度等於搜尋程度、擁有資訊的多寡與額外資訊的難易度、取得評價額外的資訊及過程得到滿足的資訊等因素的影響。

#### (三) 評估方案：

消費行為在理性的前提下，消費者會從搜集來的資訊中，評估各項對自己有利的方案。理性的消費者對於感興趣的物品，會依照自己的慾望與需求，對產品給予不同的重要性，透過對產品信念帶來的影響，建立對其品牌的態度，並會在多重屬性的商品間加以評估抉擇。

#### (四) 購買決策：

在評估方案過程中，會形成個人的品牌偏好，更可能出現購買意圖，

傾向於偏好的品牌，但卻不一定會決定購買原本認定的最好品牌，因為別人的態度與觀感及無法預期的情形都可能會讓消費者臨時改變、取消或者延後購買。消費者在決定購買意圖的執行過程，可能受到任何品牌、供應商、時間、數量及付款方式等因素影響。

#### （五） 購後行為：

消費者購買行為後所遭遇某種程度滿意或不滿意的反應。滿意的程度應取決於對產品的期望與對產品的期望或失望；如符合期望，消費者會感到滿意；如果超出期望，則會產生極大的滿足感。而消費者的滿意程度會反映在其周遭的口碑宣傳、再購行為與品牌的認同度。

江淑惠(2014)行銷策略為因應客戶群的採購行為的差異，應整合適合的變數，規劃適當的銷售人力。

郭家禎(2013)一個完全沒有需求的消費者要他產生購買行為是相當困難的，最好的方式是在他最需要的時候，最他做出適度的促銷活動，所以說只要能預測消費者的購買需求，就能精確的掌握銷售時機，銷售成功的機率將大為提升。

洪薇婷(2012)提到顧客的「性別」、「職業」與消費動機之間息息相關。以台北市非連鎖咖啡館為例，男性顧客以享受空間氣氛的消費動機最多，女性則最常在咖啡館中與親友聚會。學生至咖啡館多以閱讀為目的，停留時間較長，對於上班族則是享受空間氣氛的地點，停留時間多不超過 2 小時。消費時段與顧客的「職業」、「月收入」有關，學生多在平日晚上消費，上班族則在平日或假日白天的比例最高。

消費動機不同的顧客對於設計效果、空間機能和外觀特色的知覺評價有所不同。消費頻率不同的顧客表達了不同程度的購買意願，有重複購買行為者，有較高的意願。

楊和伸(2014)歸納為消費者行為是：

- (一) 消費者行為是動態、情感、認知、行為與購買環境的作用。
- (二) 包含從購買前的思考、購買當下決策與購買後的心理狀態。
- (三) 除了個人主觀之外，也會參考其它人或社會建議，在消費的過程中會追求功能性利益與心理利益。

影響消費者購買決策的因素有：

- (一) 過程中會追求功能性利益與心理利益。
- (二) 受到風潮流行下的衝動購買。
- (三) 非自發性的購買，受到環境因素的影響而影響決策過程。

## 第四節 顧客關係管理

企業透過顧客關係管理的操作，得以鞏固現有顧客、抓住新的顧客，讓企業更親近顧客，來建立顧客的忠誠度。

李婉琦(2013)研究認為顧客關係管理是「一種以顧客為中心的管理理念，把顧客當成企業重要資源的一部分加以管理，並利用資訊科技之協助，發掘顧客長期價值，以達到最終提升企業利益之目標」。

企業在資源有限的情況下，依據不同類型的顧客採取不同的行銷策略，各類型顧客之行銷策略如下圖 2-4，每類顧客說明如下：

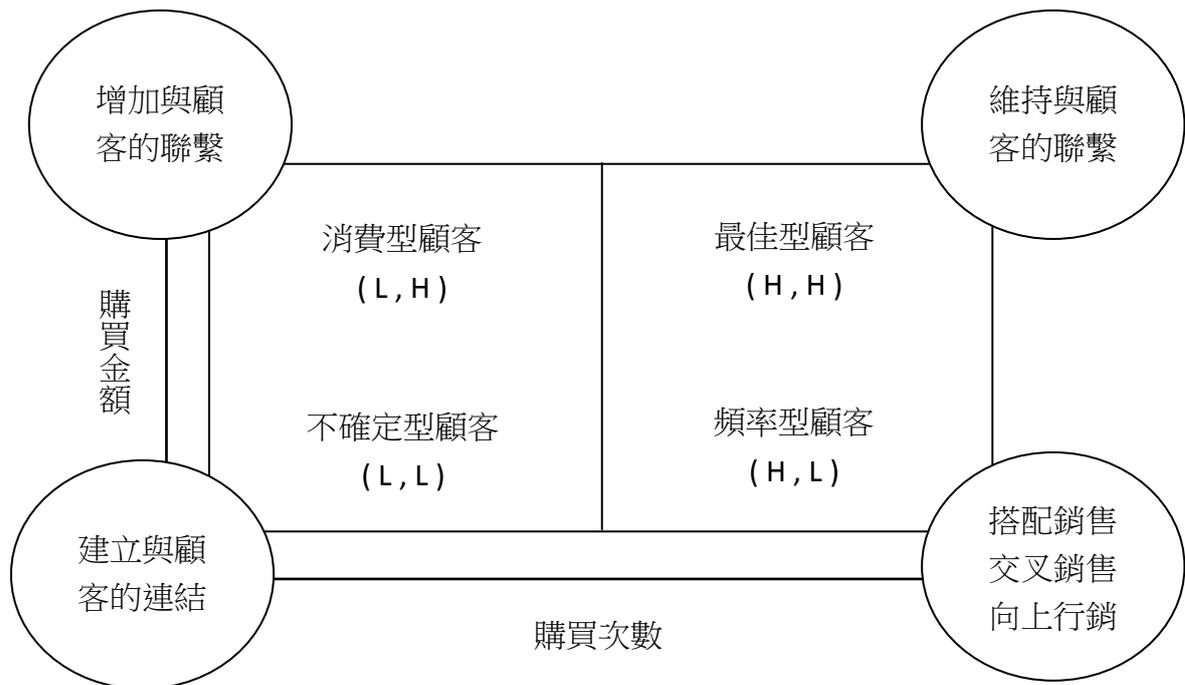


圖2-4 各類型顧客之行銷策略

資料來源：林婉琦, 2013

#### 一、最佳型顧客

此類型顧客對企業貢獻度高，應努力與此類型顧客維持關係。

#### 二、消費型顧客

此類型顧客購買金額較高，但消費次數較低，應設法增加與此類型顧客之聯繫。

#### 三、頻率型顧客

此類型顧客購買金額較少，但消費次數較多，顯示此類型顧客對企業已具有基本忠誠度，因此可透過搭配銷售、交叉銷售及向上行銷，提升此類型顧客之購買金額。

#### 四、不確定型顧客

此類型顧客多屬首次購買且僅買一次之顧客，若能與此類型顧客持續保持聯繫，設法增加回購率，將可成為企業龐大之潛在顧客族群。

我們希望與最佳型顧客維持關係，並將消費型顧客與頻率型顧客透過行銷手法，提升為最佳型顧客。

## 第三章 研究方法

本章節共分四個小節，內容依序為研究設計與架構、資料庫設計、Apriori 演算法，最後則是介紹資料分析軟體 SQL Server Business Intelligence Development Studio(以下皆以 SQL Server 為代稱)。

### 第一節 研究設計與架構

市集從 102 年 6 月開始成立，並於 102 年 12 月開始至 103 年 10 月收集了各月農學市集的會員交易資料，並進行資料整合、篩選等初步工作，接著進行資料庫的建立與正規化，最後使用 Apriori 演算法，從資料庫中探勘會員的購買行為與商品屬性之間的關聯性。

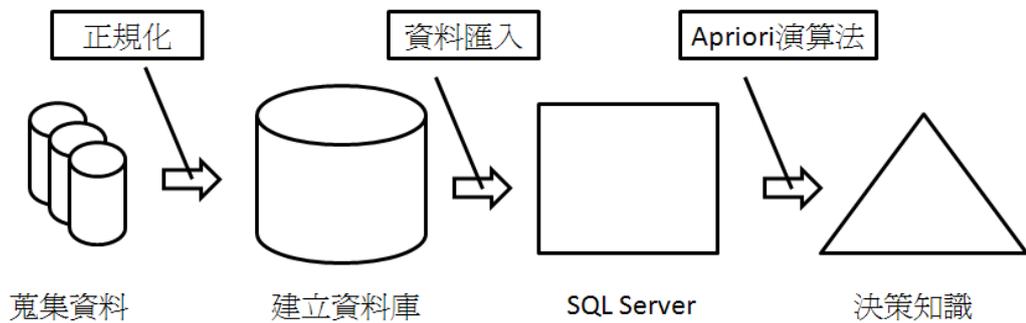


圖3-1 流程圖

### 第二節 資料庫設計

首先將搜集的資料分析、篩選、統計、挖掘並進行正規化，資料庫通常需經 3 階正規，第一階正規化：將重複的資料儲存到不同的紀錄，並加上適當的主鍵；第二階正規化：除去資料表中的部分相依性；第三階正規化：除去資料中的間接相依性，並繪製 ER-Model 的圖形，如下圖 3-1、圖 3-2、圖 3-3、圖 3-4。

<b>產品資料表</b>	<b>教職員資料表</b>	<b>訂單資料表</b>
廠商代碼	教師編號	訂單編號
廠商名稱	教師姓名	教師編號
產品代碼	單位代碼	產品代碼
產品名稱	單位名稱	數量
廠商類別	位置	售價
產品屬性	性別	下單日期
規格	電話	預計送貨日
價格	E-mail	
聯絡人		
廠商電話		

圖3-1 教職員團購未經正規化

<b>廠商資料表</b>	<b>教職員資料表</b>	<b>訂單主檔資料表</b>
廠商代碼	教師編號	訂單編號
類別	教師姓名	教師編號
廠商名稱	單位代碼	下單日期
聯絡人	性別	
廠商電話	電話	
	E-mail	
<b>產品資料表</b>	<b>單位資料表</b>	<b>訂單明細資料表</b>
產品代碼	單位代碼	訂單編號
產品名稱	單位名稱	產品代碼
廠商代碼	位置	數量
規格		售價
價格		預計送貨日
聯絡人		
廠商電話		

圖3-2 教職員團購正規化結果

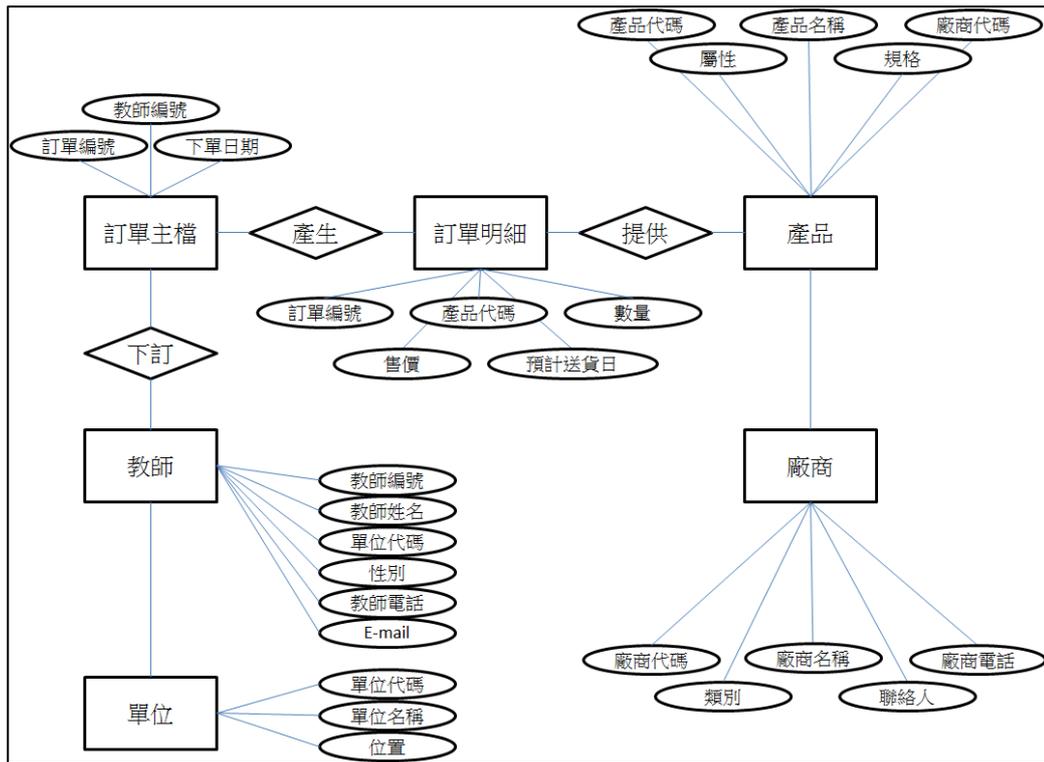


圖3-3 校內教職員團購ER-Model關聯圖

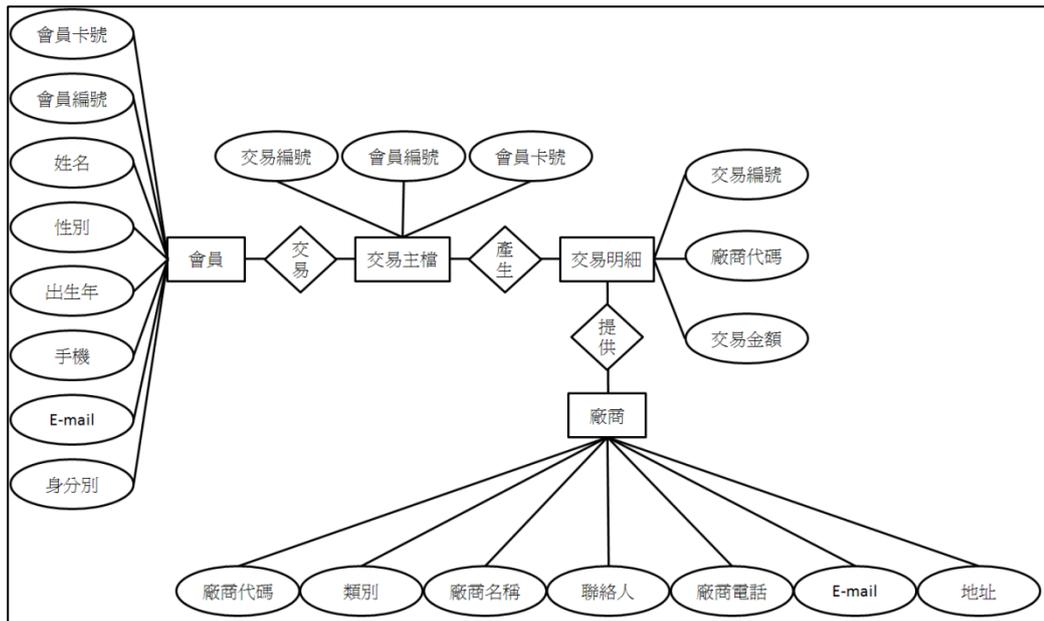


圖3-4 實體市集ER-Model關聯圖

### 第三節 Apriori 演算法

Apriori 演算簡述如下：首先，簡單統計所有含一個元素項目集出現的頻數，並找出那些不小於最小支持度的項目集，即一維最大項目集。從第二步開始循環處理直到再沒有最大項目集生成。循環過程是：第 k 步中，根據第 k-1 步生成的(k-1)維最大項目集產生 k 維元素項目集，然後對資料庫進行搜索，得到候選項目集的項集支持度，與最小支持度進行比較，從而找到 k 維最大項目集。

以下為 2013 年 11 月致理幸福農學市集教職員團購資料中的 4 筆資料加以解釋：

表 3-1

ID	Items
T003	I2,I3,I4,I8
T004	I1,I2,I3,I8
T006	I1,I8
T007	I2,I3

一般來說 minSupport 設定為總樣本數的 20%，故這裡假設最小支持度 minSupport=2，下面用圖例說明演算法運行的過程：

表 3-2

ID	Items
T003	I2,I3,I4,I8
T004	I1,I2,I3,I8
T006	I1,I8
T007	I2,I3

掃描資料庫，對每個元素項目進行支持度計數得到表 3-3：

表 3-3

項集	支持度計數
I1	2
I2	3
I3	3
I4	1
I8	3

比較元素項目支持度計數與最小支持度 minSupport，因 I4 沒有大於 minSupport 故予以刪除，產生 1 維最大項目集表 3-4：

表 3-4

項集	支持度計數
I1	2
I2	3
I3	3
I8	3

由表 3-4 產生元素項目集表 3-5：

表 3-5

項集
{I1,I2}
{I1,I3}
{I1,I8}
{I2,I3}
{I2,I8}
{I3,I8}

掃描資料庫，對每個元素項目集進行支持度計數，得到表 3-6：

表 3-6

項集	支持度計數
{I1,I2}	1
{I1,I3}	1
{I1,I8}	2
{I2,I3}	3
{I2,I8}	2
{I3,I8}	2

比較元素項目支持度計數與最小支持度  $\text{minSupport}$ ，產生 2 維最大項目集

表 3-7：

表 3-7

項集	支持度計數
{I2,I3,I8}	2

直到沒有其他最大項目及生成，演算法中止。

## 第四節 實作工具(SQL Server)

SQL Server 是主從式資料庫伺服器，它本身並沒有像 Access 提供使用者介面元件供我們開發用戶端程式。不過這也正是主從式架構的特點，伺服器本身並不需要著重在使用者介面；反之，透過伺服器所支援的各種存取方式，我們能經由 Access 存取，或是以各種不同的發展工具如 Visual Basic、Delphi、C++、PowerBuilder、ASP 來開發用戶端應用程式，最後在與後端資料庫 SQL Server 組合成一個完整的應用系。

## 第四章 研究發現

本論文以致理幸福農學市集作為研究個案的代表，將其分為校內教職員團購及實體市集交易紀錄兩部分研究，運用微軟的資料庫管理工具 Microsoft SQL Server 2008 的 SQL Server Business Intelligence Development Studio，使用關聯法則 Apriori 演算法來分析團購及市集每月交易紀錄，根據分析結果提供消費者做為購買參考。第一節校內教職員團購資料使用 102 年 9 月至 103 年 10 月，共 509 筆交易；第二節實體市集交易紀錄使用 102 年 12 月至 103 年 10 月，共 2477 筆交易，其中 103 年 2 月因適逢年節，市集暫停 1 次，故共有 10 個月資料。這些數據乍看之下或許不多，但這些是真實的交易資料，如果不是參與市集，絕對無法從任何公司內部抓取資料，因此每筆資料都彌足珍貴。

### 第一節 個案研究-以致理幸福農學市集校內教職員團購為例

本節以校內教職員團購為研究，使用紙本與 E-mail 發送團購 DM 及訂購單，根據訂購單為資料收集的依據，藉由資料分析結果，給予教職員推薦產品，促使交易額成長。本節分為三個部分，第一個部分收集各月份資料，第二部分將資料匯入 SQL Server 建立資料倉儲，第三部分建立資料採礦(Data Mining)模型。

#### 一、各月份團購資料收集

根據前一章節正規化結果，我們得到六個資料表分別為廠商資料

表、產品資料表、教師資料表、單位資料表、訂單主檔資料表及訂單明細資料表。

## 二、匯入 SQL Server Management Studio

根據前面所收集到的資料，用匯入的方式將資料放入資料庫中，一次全部匯入減少花費時間。

### (一) 建立資料庫

資料庫右鍵→新增資料庫。

於資料庫名稱打上我們的資料庫名稱：教職員團購。

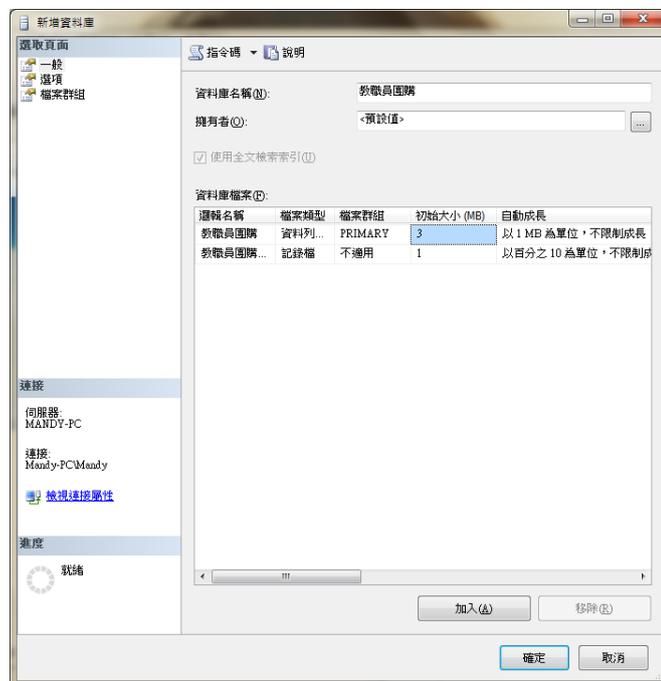


圖4-1 新增資料庫畫面

### (二) 匯入資料表內容

教職員團購右鍵→工作→匯入資料→開啟 SQL Server 匯入和匯出精靈。

開啟匯入和匯出精靈，跟隨步驟按下一步進而完成精靈。



圖4-2 SQL Server匯入和匯出精靈畫面

由於資料搜集時放在 Excel 中，這邊資料來源選擇 Microsoft Excel，下方檔案路徑填入 Excel 所在位置。



圖4-3 資料來源與路徑

將目的地設定本機 SQL Server Native Client 10.0，資料庫教職員團購。

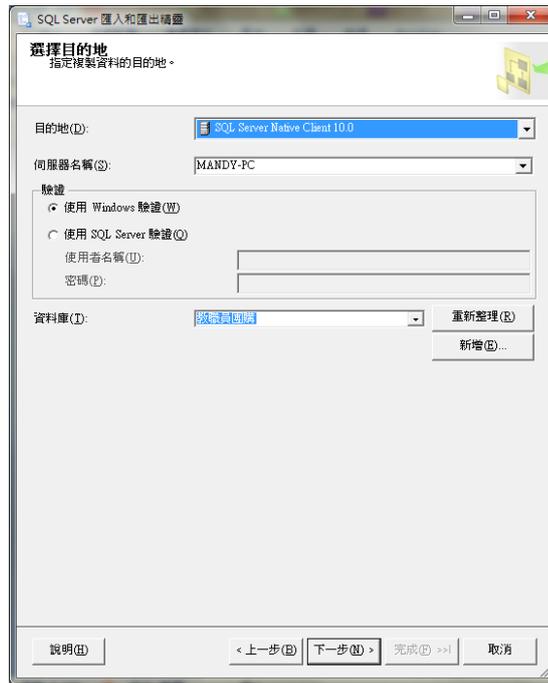


圖4-4 目的地畫面

將要匯入的資料於來源打勾，由於資料庫內尚未新增資料表，SQL 會於目的地使用 Excel 中的名稱做預設。



圖4-5 選取來源資料表

### 三、使用 SQL Server Business Intelligence Development Studio 建立模型

#### (一) 建立新的專案

選擇範本 Analysis Services 專案，輸入專案名稱：教職員團購。

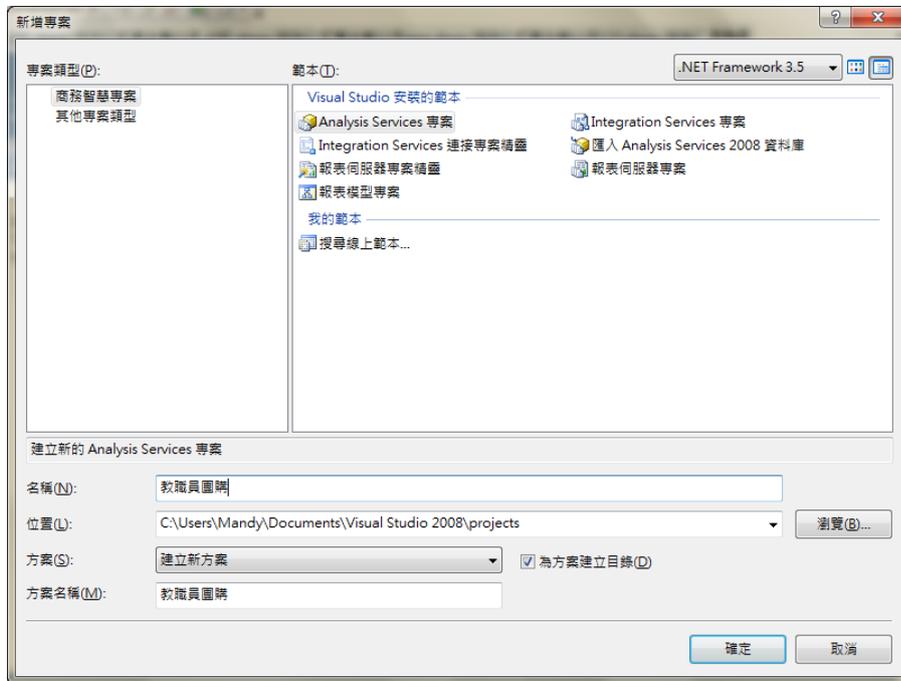


圖4-6 建立專案

於登入畫面輸入帳號密碼獲得資料庫讀取權限，完成專案建立。

## (二) 新增資料來源

開啟資料來源精靈，跟隨步驟按下一步進而完成精靈。

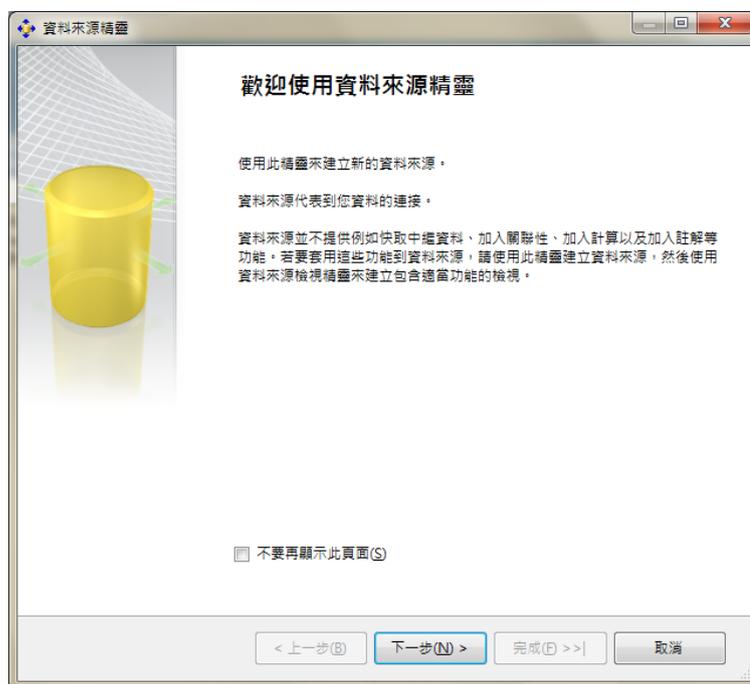


圖4-7 新增專案

資料來源為本機，伺服器名稱 localhost。

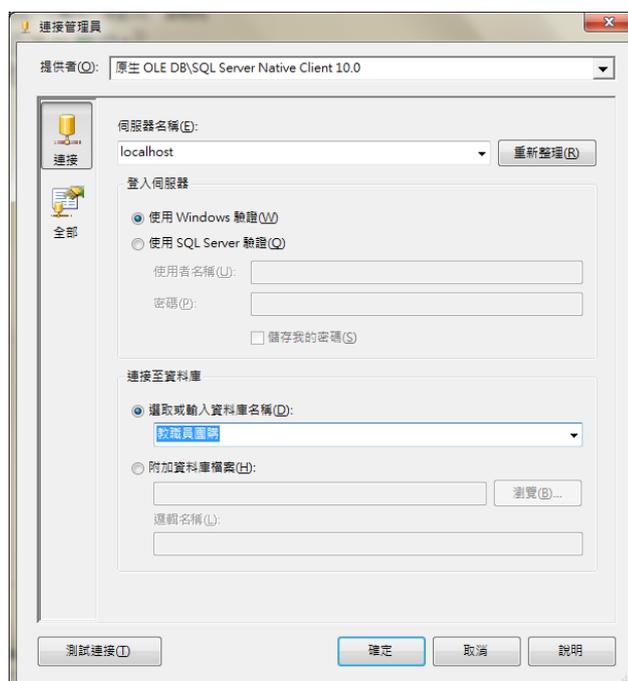


圖4-8 選擇連接伺服器名稱

選擇前面設定的 localhost.教職員團購

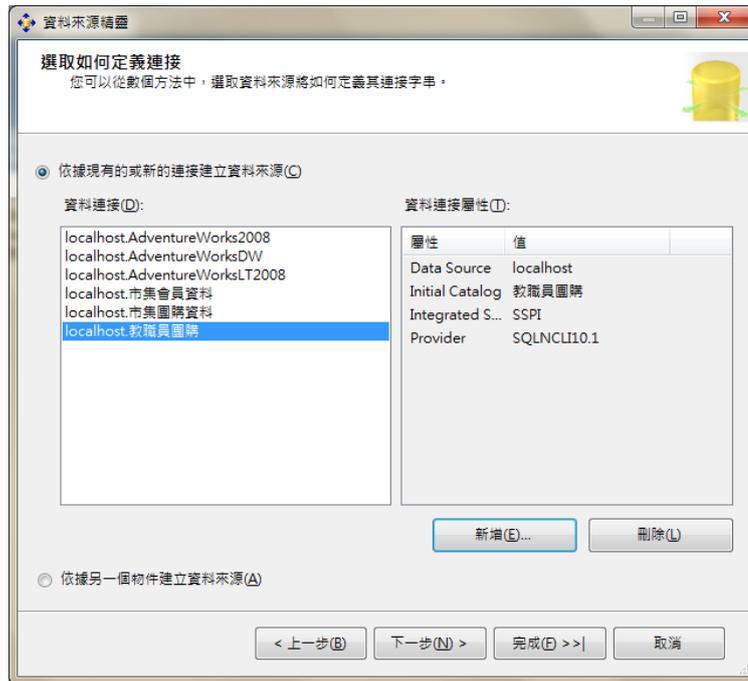


圖4-9 如何定義連接

### (三) 新增資料來源檢視

開啟新增資料來源檢視精靈，跟隨步驟按下一步進而完成精靈。

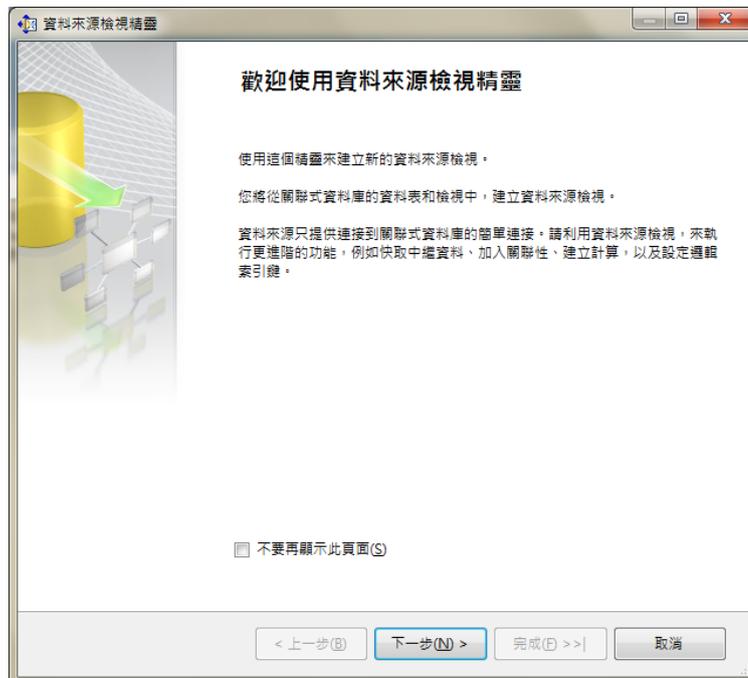


圖4-10 資料來源檢視精靈

選擇上面建立好的資料來源：教職員團購。

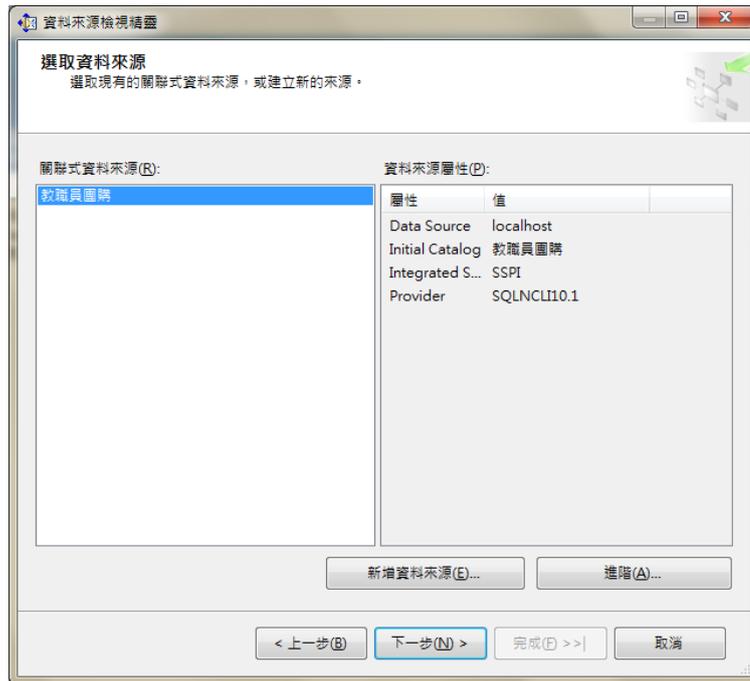


圖4-11 選取資料來源

將所有資料表選取放到右邊，使資料表會於關連圖中出現。

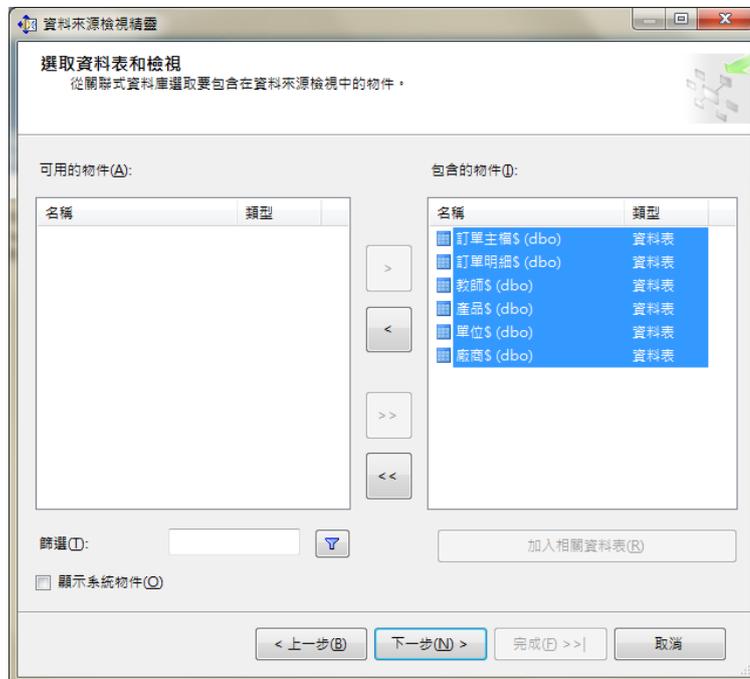


圖4-12 選取資料表和檢視



依照現有資料資料庫來定義採礦結構，這邊選擇：從現有的關聯是資料庫或資料倉儲。

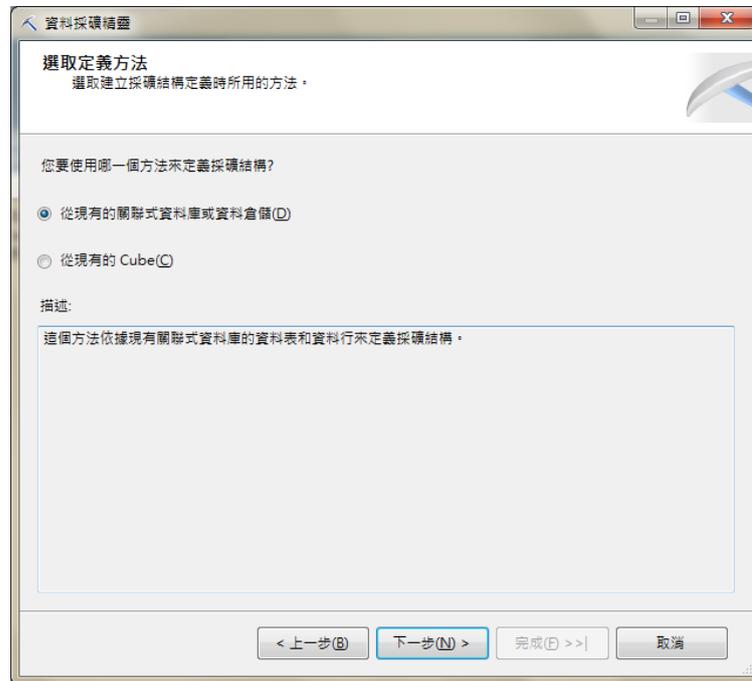


圖4-15 選取定義方法

本論文研究為關連規則 Apriori，這邊選擇 Microsoft 關聯規則。

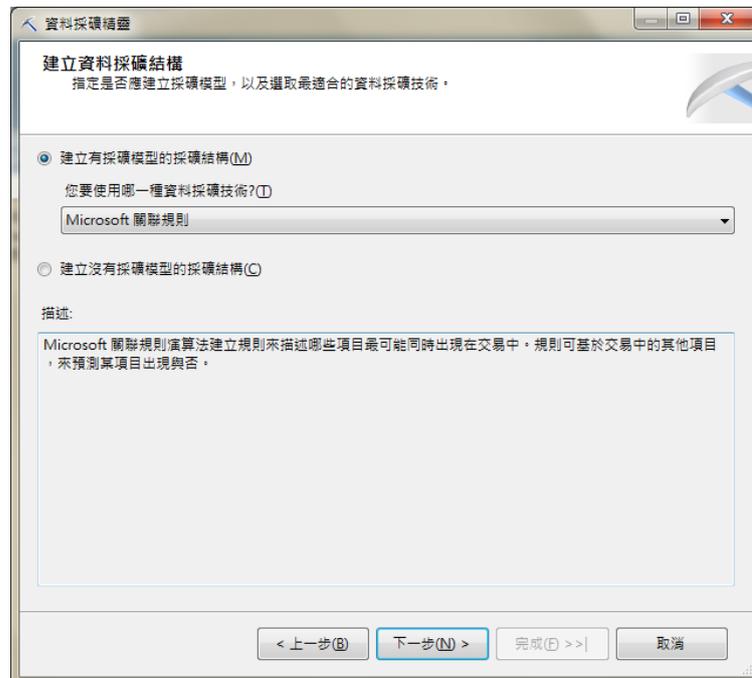


圖4-16 關聯規則建立資料採礦結構

團購資料以訂單做為基本單位，則案例定義為：訂單主檔\$。巢狀資料表透過外鍵與案例資料表相連接，則巢狀定義為：訂單明細\$。

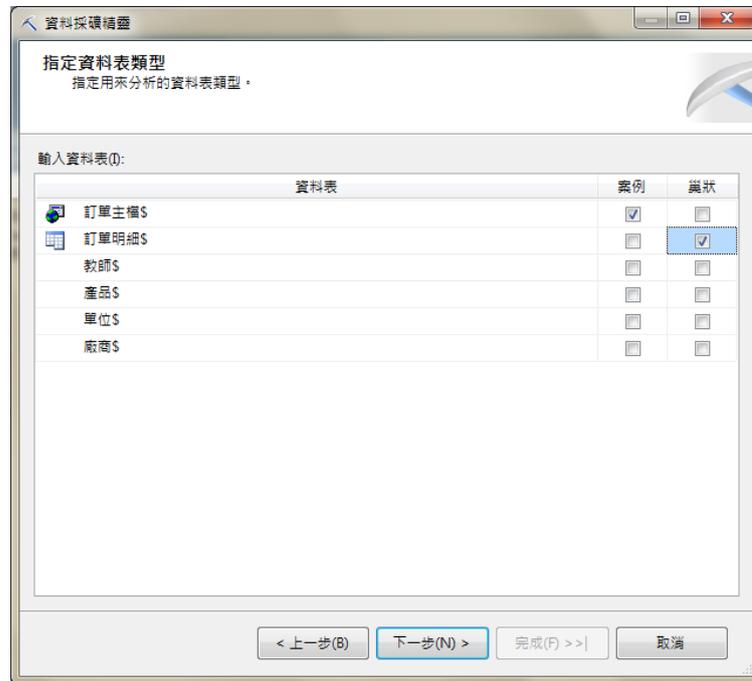


圖4-17 指定資料表類型

分析過程中，以教師編號及產品代碼做分析，則輸入位置選取：教師編號、產品代碼。分析結果，在相依性網路圖中我們希望看到教師編號及產品代碼，則可預測的位置選取：教師編號、產品代碼。

- 索引鍵：是資料表中的一個欄位或一組欄位，它們為資料表提供每一個資料列的唯一識別碼。
- 輸入：將欲分析的資料，勾選後會根據其資料產生分析結果及相依性網路圖。
- 可預測：欲於分析結果中出現的資料，需勾選此項目。

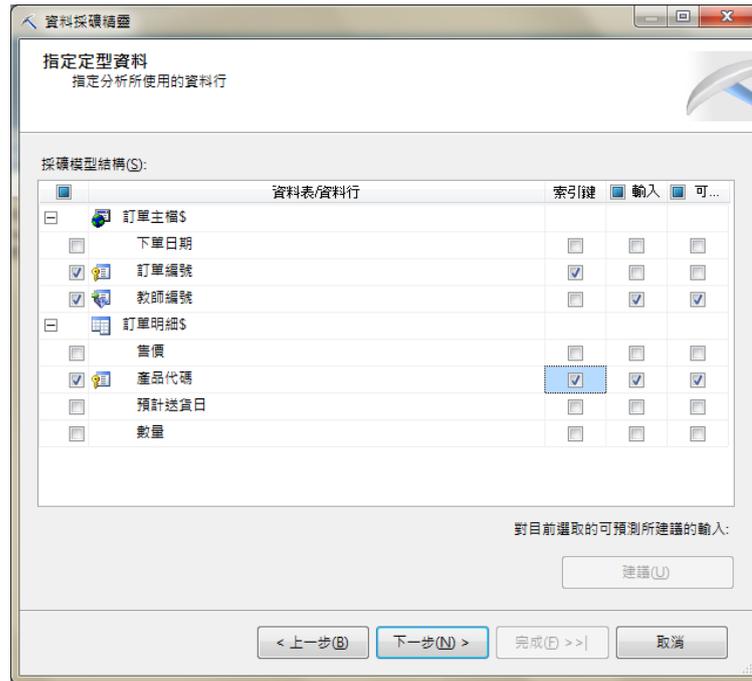


圖4-18 指定索引鍵、輸入、可預測

測試資料百分比使用預設 30%。

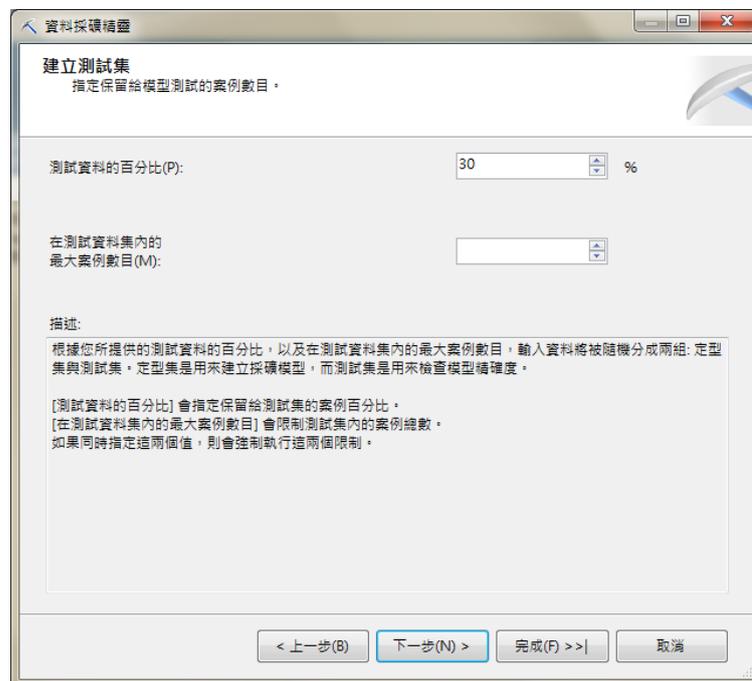


圖4-19 建立測試集

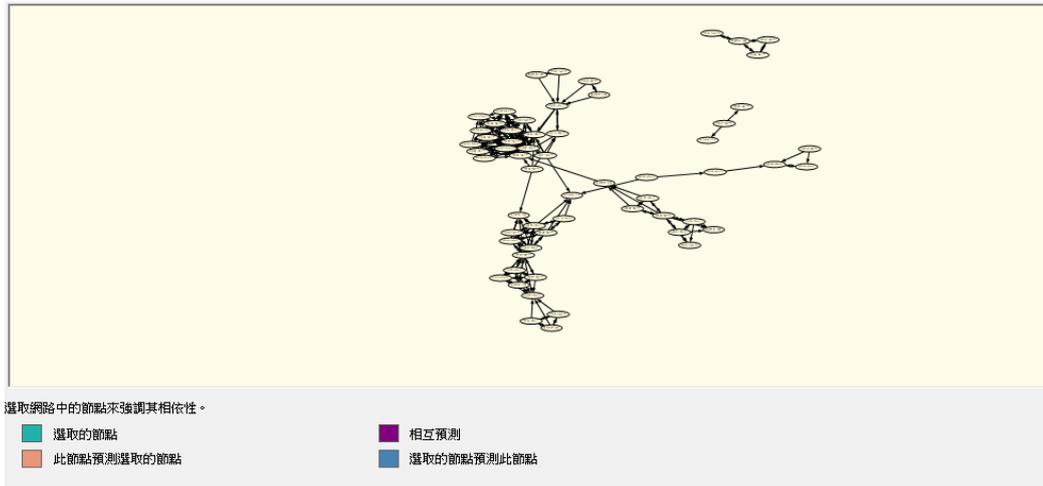


圖4-20 教職員團購相依性網路圖

根據圖 4-20 四種顏色的角色來強調其相依性，■選取的節點，■此節點預測選取的節點，■互相預測，■選取的節點預測此節點。放大其中一塊圖示如圖 4-21：

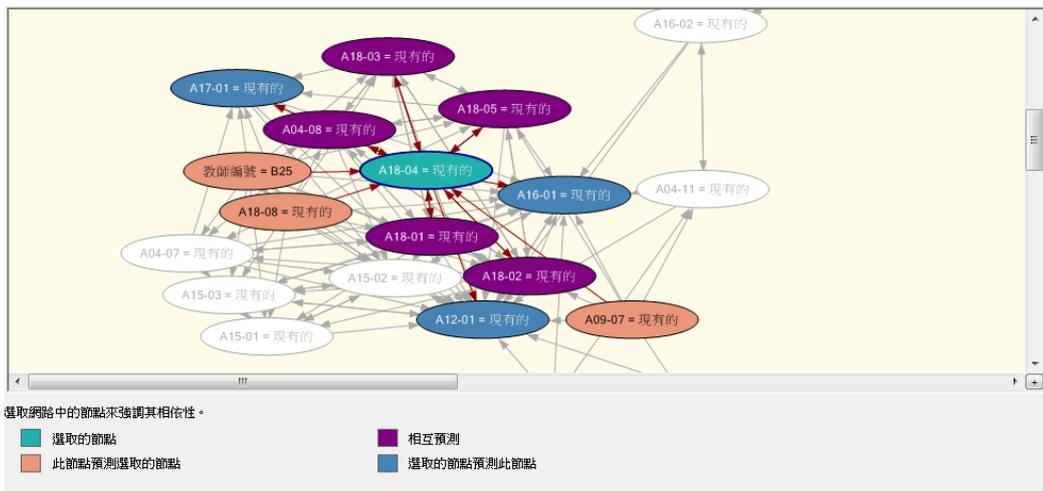


圖4-21 於相依性網路圖中點選A18-04

由圖 4-21 可得知■為點選的 A18-04，■為教師編號 B25 及購買 A18-08 的教職員有可能購買 A18-04 這項商品，■為購買這些商品可能購買 A18-04 且購買 A18-04 也可能購買這些商品，相依性很高，■為購買 A18-04 的消費者對於節點商品可能購買，在實際應用上，我們可以將連結度高的紫色作為團購商品設計，於節慶時推出年節禮盒。

## 第二節 個案研究-以致理幸福農學市集交易紀錄為例

本節以市集交易紀錄為代表，其交易紀錄累積二千多筆，以會員與農友為資料搜集對象，藉由資料分析，我們將能給予會員推薦產品，促使交易額成長。本節分為三個部分，第一個部分收集各月份資料，第二部分將資料匯入 SQL Server 建立資料倉儲，第三部分建立資料採礦(Data Mining)模型。

### 一、各月份市集交易資料收集

根據前一章節正規化結果，我們得到四個資料表分別為廠商資料表、會員名單資料表、交易主檔資料表及訂單總檔資料表。

### 二、匯入 SQL Server Management Studio

根據前面所收集到的資料，用匯入的方式將資料放入資料庫中，一次全部匯入減少花費時間。

### 三、使用 SQL Server Business Intelligence Development Studio 建立模型

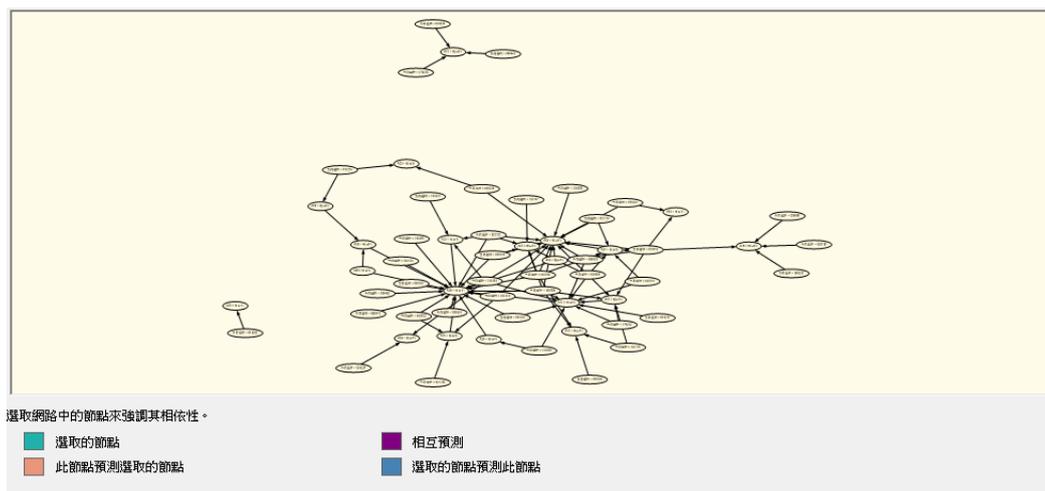


圖4-22 市集交易相依性網路圖

根據圖 4-22 四種顏色的角色來強調其相依性，■ 選取的節點，■ 此節點預測選取的節點，■ 互相預測，■ 選取的節點預測此節點。放大圖示看圖 4-23

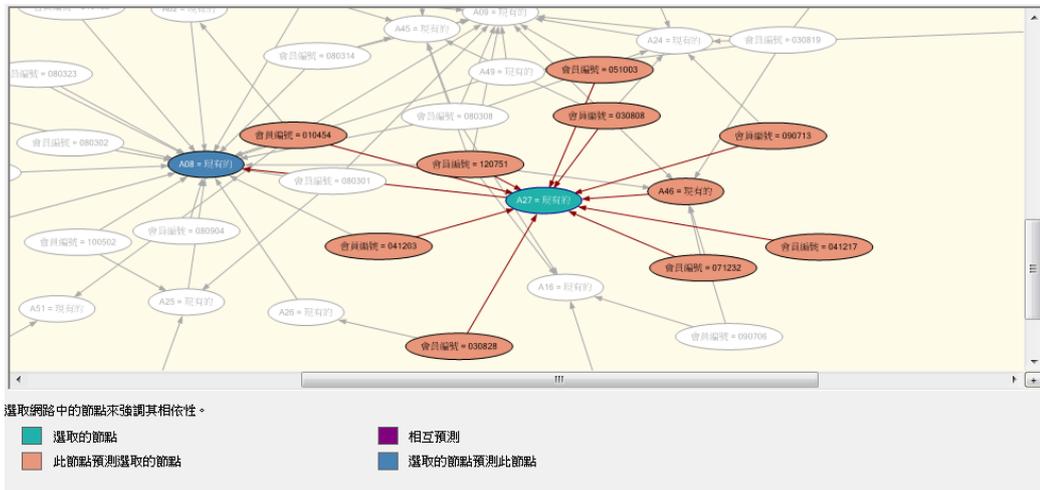


圖4-23 於相依性網路圖中點選A27

由圖 4-23 可得知■為點選的 A27，■為購買的消費者有可能購買 A27 這項商品，■為購買 A27 的消費者對於節點商品可能購買 A08 此項產品。

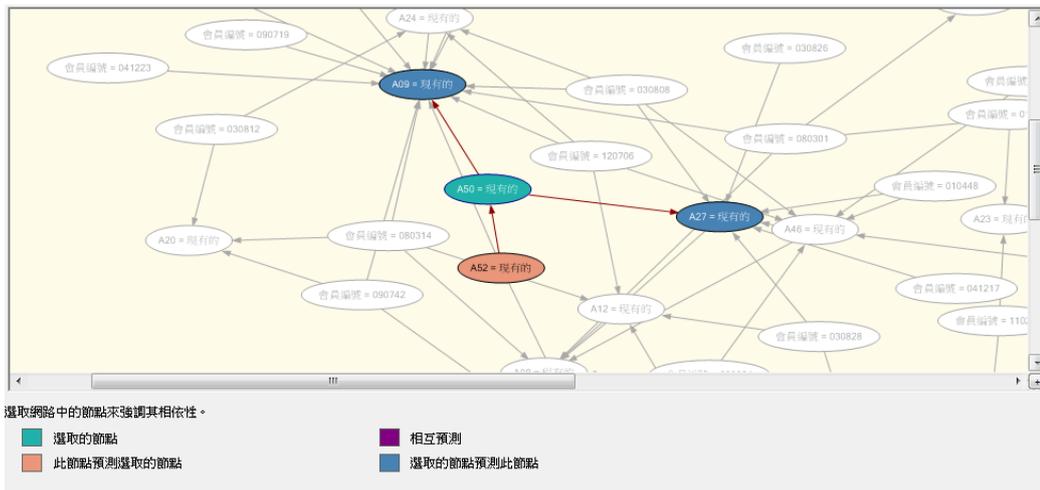


圖4-24 相依性網路圖點選A50

圖 4-24 以市集交易紀錄 102 年 12 月至 103 年 7 月為模型建立範圍，點選 A50 出現的關係資料。

# 第五章 研究結論與建議

## 第一節 結論

透過 SQL Server 資料探勘系統，利用關聯法則 Apriori 演算法，建立相依性網路圖，預測消費者購買行為，找出預測最佳模型，提高訂單預測的準確度，給予會員及消費者商品推薦。

本研究將 102 年 12 月至 103 年 7 月的交易資料與 103 年 8 月交易資料相比對驗證，連結度最高的廠商 (A50 與 A07、A27)，14 位有購買 A50 (樸實小舖)的消費者中，有 4 位有購買 A09 (穿龍豆腐坊)或 A27 (輝要無毒菜園)，數據資料約為 30%，應用於攤位擺設上，將攤位擺設在隔壁，可以提升 30%的提袋率。

## 第二節 後續研究建議

經由上述的研究結果，本研究針對致理幸福農學市集交易資料的分析，提供實務之建議，以下說明本研究的研究限制與後續建議。

- 一、資料限制：搜集市集的資料有限，所以本研究的分析結果，僅限於致理幸福農學市集這個場域使用，相同的研究方式可應用於其他場域。
- 二、資料探勘運用方法：本研究僅使用關聯法則 Apriori 演算法進行研究，未來可利用更多方法來進行探討，如：決策樹、類神經網路等演算法。

## 參考文獻

- [1] mayhem, S (2001), 《牛津地理學辭典》, 上海外語教育出版社。
- [2] 賴鳳霽(2010), 「農民市集的形塑與發展：以臺中合樸農學市集為例」, 國立臺灣師範大學地理學系碩士論文。
- [3] 美國農夫市集協會(The National Association of Farmer's Markets)
- [4] 陳嬾伊(2009), 「農夫市集發展歷程—以興大有機農夫市集為例」, 國立中興大學生物產業推廣暨經營學系碩士論文。
- [5] 林淑玲(2012), 「合樸農學市集的農學實踐與社群價值建構之研究」, 南華大學建築與景觀學系環境藝術碩士班碩士論文。
- [6] Fayyad, U, and Piatetsy-Shapiro, G, and Smith, P (1996), “From Data Mining to Knowledge Discovery: An Overview”, In advances in knowledge Discovery and Data Mining, 471-473.。
- [7] Uniminer Data Mining 資料採礦介紹, 取自：  
<http://www.uniminer.com/center01.htm>
- [8] 梁衍忠(2012), 「以資料探勘分析影響國民中小學學習成就因素之研究」, 國立東華大學課程設計與潛能開發學系教育博士班博士論文。
- [9] 許凱嵐(2012), 「以資料探勘技術建立國中學生升學落點之模型」, 國立高雄應用科技大學資訊管理系碩士在職專班碩士論文。
- [10] 宋堯正(2013), 「資料採礦分析法於解析客訴不良品之應用-以 TFT-LCD 製造廠某 C 公司為例」, 國立中央大學工業管理研究所碩士論文。
- [11] 江淑惠(2014), 「資料探勘技術於台灣製藥產業客戶價值分析-行銷策略與銷售人力績效特質之探討」, 銘傳大學企業管理學系博士論文。
- [12] William H. Inmon(1993), “Building the Data Warehouse”.
- [13] 郭家禎(2013), 「運用二階段分類技術挖掘潛在中小企業借貸戶之研究」,

銘傳大學資訊管理學系碩士班碩士論文。

- [14] 范瑋宸(2011),「以關聯法則分析網路流量為基礎偵測潛藏跳板之研究」, 國立高雄大學資訊管理學系碩士班碩士論文。
- [15] RakeshAgrawal , andRamakrishnanSrikant , “Fast Algorithms for Mining Association Rules”.
- [16] 常世杰(2014),「利用資料探勘 Apriori 演算法預測零售賣場之個人購物行為」, 國立高雄第一科技大學服務科學管理研究所碩士論文。
- [17] 張家綦(2010),「資料採礦應用於消費者網路團購因素探勘之研究」, 淡江大學管理科學研究所企業經營碩士在職專班碩士論文。
- [18] Kotler,Philip(1996) , “Marketing management-analysis , planning , implementation , and control” , 9th ed. N. J.:Prentice Hall International.
- [19] 洪薇婷(2012),「台北市非連鎖咖啡館顧客商店印象與購買行為之研究」, 國立台灣師範大學運動與休閒管理研究所碩士論文。
- [20] 楊和伸(2014),「消費者購買行為心理因素分析-以 LEXUS 為例」, 逢甲大學經營管理碩士在職專班碩士論文。
- [21] 李婉琦(2013),「運用資料探勘輔助網路行銷策略之研究」, 銘傳大學資訊管理學系碩士在職專班碩士論文。