

南 華 大 學
資訊管理學系
碩士論文

以資料採礦探討多元入學學生之流失率與學業表現之研究
A Study on dropout rates and course performances of multiple-enrolled
students using data mining techniques



研究生：林青翰
指導教授：邱宏彬

中華民國 106 年 7 月 3 日

南 華 大 學

資訊管理學系

碩 士 學 位 論 文

以資料採礦探討多元入學學生之流失率與學業
表現之研究

研究生： 林素翰

經考試合格特此證明

口試委員：

陳張彥榮

邱宏彬

周志賢

指導教授：邱宏彬

系主任(所長)：洪錦建

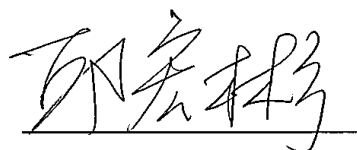
口試日期：中華民國 106 年 07 月 03 日

南華大學碩士班研究生
論文指導教授推薦函

資訊管理系碩士班 林青翰 君所提之論文
以資料採礦探討多元入學學生之流失率與學業
表現之研究

係由本人指導撰述，同意提付審查。

指導教授



106年 5月 3日

南華大學資訊管理學系碩士論文著作財產權同意書

立書人：_____林青翰_____之碩士畢業論文

中文題目：

以資料採礦探討多元入學學生之流失率與學業表現之研究

英文題目：

A Study on dropout rates and course performances of multiple-enrolled students using data mining techniques

指導教授：_____博士_____

學生與指導老師就本篇論文內容及資料其著作財產權歸屬如下：

- 共同享有著作權
- 共同享有著作權，學生願「拋棄」著作財產權
- 學生獨自享有著作財產權

學生：_____林青翰_____ (請親自簽名)

指導老師：_____邱宏彬_____ (請親自簽名)

中 華 民 國 1 0 6 年 0 7 月

以資料採礦探討多元入學學生之流失率與學業表現之研究

學生：林青翰

指導教授：邱宏彬 博士

南 華 大 學 資 訊 管 理 學 系

摘 要

有鑑於國內少子化的影響，大專校院學生數劇減在民國95年至105年間呈現負成長，招生日趨競爭，若能減少在校學生的流失，對學校而言則是一大助力。本研究以一大學102學年度入學大學日間部資管系學生歷史學籍資料，運用資料採礦技術決策樹分析，找出多元入學學生流失率與學業表現等因素，提供相關之建議以降低學生流失。

本研究共取得有效資料76筆，在資料特性分析發現學生學業表現以繁星推薦學生學業表現仍為各入學管道中，學業表現較為突出部分；反觀轉學考進入學校之學生學業表現仍為最弱；所以學業表現與學生入學方式有關。在學生流失率部份，與居住地區和入學方式有關。在資料採礦部分發現，以學業總成績為最重要的因素，入學方式與居住地區為其次因素，再其次為性別與班級。

關鍵詞：資料採礦、決策樹、學生流失

A Study on dropout rates and course performances of
multiple-enrolled students using data mining techniques

Student : Lin,Ching-Han

Advisors : Dr. Chiu,Hung-Pin

Department of Information Management
The M.I.M. Program
Nan-Hua University

ABSTRACT

In view of the impact of the domestic minority, college students in D.C. 2006 to 2016 years showed negative growth, enrollment increasingly competition, if the loss of students in school, the school is a big help. This study uses the data mining technology decision tree analysis to find out the factors such as the loss rate of the students and the academic performance of the students, and provide relevant suggestions to reduce the loss of students.

In this study, a total of 76 valid data were obtained. In the analysis of the data, it was found that the academic performance of the students was still recommended by the stars. The academic performance of the students was still the most prominent part of the students. The academic performance of the students was still the weakest. So academic performance and student enrollment. In the part of the student's wastage, related to the area of residence and admission. In the data mining part of the discovery, to academic total score for the most important factor, school enrollment and residential areas as the second factor, followed by gender and class.

Keyword: data mining, decision tree, student dropout

目錄

第一章、緒論	1
第一節 研究背景與動機.....	1
第二節 研究目的.....	1
第三節 研究範圍、限制.....	2
第四節 研究步驟.....	2
第五節 論文架構.....	4
第二章、文獻探討	5
第一節 顧客流失與學生流失定義.....	5
第二節 資料採礦定義.....	6
第三節 資料採礦的應用.....	8
第四節 決策樹.....	9
第五節 相關研究回顧.....	10
第三章、研究方法	11
第一節 研究對象.....	11
第二節 研究流程.....	11
第三節 資料前置處理.....	12
第四節 資料採礦工具.....	14
第四章、資料分析	16
第一節 資料特性分析.....	16
第二節 決策樹分析.....	23
第五章、結論與未來工作	29
第一節 研究結論與建議.....	29
第二節 研究未來工作.....	30
參考文獻	31
一、 中文部份.....	31
二、 西文部份.....	32

表目錄

表 3-1 學生資料表	13
表 4-1 102 級資管系學生男、女生人數表	16
表 4-2 不同就學狀況比例表	18
表 4-3 入學方式學業表現一覽表	19
表 4-4 居住地區就讀人數與流失人數表	20
表 4-5 入學方式流失人數統計與比例表	21
表 4-6 居住地區、入學方式、就讀人數與流失人數統計表	22
表 4-7 決策樹分析之欄位	23
表 4-8 各模型分類矩陣比較表	28



圖目錄

圖 3-1 研究流程圖	12
圖 4-1 102 級資管系男女生比例圖	16
圖 4-2 學生流失狀態圖	17
圖 4-3 就學狀況比例圖	18
圖 4-4 入學方式與學業成績分析圖	19
圖 4-5 決策樹模型 1	24
圖 4-6 決策樹模型 2	25
圖 4-7 決策樹模型 3	25
圖 4-8 決策樹模型 4	26
圖 4-9 決策樹模型 5	26
圖 4-10 決策樹模型 6	27



第一章、緒論

本章說明本研究之研究背景與動機、研究目的、以及研究範圍與限制。第一節為研究背景與動機，說明目前大專校院學生人數之現況；第二節為研究目的；第三節為研究範圍與限制；第四節為研究步驟；第五節論文架構。

第一節 研究背景與動機

依據教育部統計處公告大專校院概況中顯示學生人數民國 106 年版中報告指出，大專校院學生數自 95 學年起至 105 學年度產生大幅度負成長，總計約降 4,552 人（教育部統計處 106 年大專校院概況 http://stats.moe.gov.tw/files/important/OVERVIEW_U01.pdf）在各校招生情況日趨嚴峻的情況下，如果對於已招收的在校生能夠減少學生休退學、降低在校生流失留住學生，對於學校應是一大助力。

第二節 研究目的

為因應招生情況日益嚴峻，加上教育部逐年增加學生多元入學方案，學校應針對入學後學生在校學習與在校生活各種情形多加觀察注意，加強輔導關心學生使學生能適應學校環境，避免學生的流失，如此亦可讓學生家長對學校積極輔導的作法有正向的回應。

由於導師而言，管理班上眾多學生已屬不易，且學校各人員及任課教師之日常行政工作及教學活動已相當忙碌，如何協助及早發現潛在可能流失學生，而多加關心輔導以減少其流失率，則是一重要關鍵。另一方面可提供系所在面對各多元管道招生學生學業表現上有所助益。

本研究擬運用資料採礦技術分析學生歷史學籍資料，找出多元入學管道之流失學生其影響因素，作為在學學生可能流失之預測，針對資料採礦之結果，提供相關之建議，以探討多元入學學生流失與學業表現之情形。

第三節 研究範圍、限制

本研究之研究範圍以研究學校 102 學年度入學之大學日間部資管系學生為主，因分析來源為單一學校科系，結果恐難類推至所有大專學校，但仍可作為類似各校減少學生流失與各招收管道學生學業表現之參考；此外，本研究以學校校務系統中的資料庫之資料為分析來源，對於學生之情緒性因素、家庭、經濟、交友等情形較難以取得與分析；再者因個資限制為保護學生隱私有關德行成績相關之獎懲記錄、缺曠課資料，校方無法提供書面或電子檔案，僅在工作中可取得資料。

第四節 研究步驟

本研究將以研究學校 102 學年度大學日間部資管系入學學生之個人資料，研究探討個案學校多元入學學生流失與學業表現，研究工具為資料採礦中之決策樹，並依循下列步驟來完成本研究。研究步驟包含：研究動機及目的確認、文獻探討、擬定研究方法及架構、資料採礦與分析、結論與建議，其說明如下：

壹、確認研究動機及目的：

確認本研究之動機與目的，以確認資料蒐集及研究進行方向。

貳、資料蒐集與回顧文獻：

蒐集相關文獻瞭解研究之相關知識，並尋求適合之技術及工具。

參、擬定研究方法及流程：

針對所研究之問題及運用工具，擬定適合的研究方法及流程。

肆、取得分析資料：

本研究以個案學校 102 學年度入學資管系學生之資料，作為研究使用的資料庫。為能正確地使用資料採礦技術，在進行資料庫的分析之前，先行檢視資料中各資料欄位之意義與價值，了解各項資料原始意義及與使用特性和限制條件，以便能正確地運用分析方法與資料性質，進行資料採礦工作。

伍、資料前置處理：

將不具分析價值之資料欄位加以刪減、修改錯誤的資料格式及利用適當方法將原有資料轉換成具分析價值之資料，以避免不適合的資料對分析結果造成不利的影響。

陸、資料採礦分析：

運用決策樹進行相關資料之分析，取得學生流失預測模型。

柒、結論：

找出流失學生之潛在共同因素，作為在學學生可能流失之預測，針對資料採礦之結果，提供可行之建議，以降低學生流失之情形。

第五節 論文架構

本研究共分為五章，各章節結構說明如下：

第一章 緒論

說明本研究之研究背景與動機、研究目的、研究步驟、研究範圍與限制、研究步驟以及本研究之論文架構。

第二章 文獻探討

就相關文獻分別探討學生流失及資料採礦之定義，並針對本研究所運用之統計資料分析與資料採礦方法決策樹的加以深入探討。

第三章 研究方法

本章節針對研究資料來源作概略介紹，包括研究對象、研究架構、資料處理以及資料採礦工具的介紹。

第四章 資料分析

先對研究資料作基本敘述統計，再針對所需要進一步研究的部分進行整理後做資料採礦分析，並比較結果。

第五章 結論與未來工作

針對第四章所得結果進行整理，並提出相關結論及未來工作。

第二章、文獻探討

本章共分為五節，第一節為學生流失定義，第二節為資料採礦定義，第三節為介紹資料採礦的運用，第四節介紹決策樹，第五節針對兩篇資料採礦在學生流失上的應用作相關探討。

第一節 顧客流失與學生流失定義

對企業而言，顧客快速流失為企業獲利下降的警訊，將顧客流失率降到最低，企業本身才達到最佳顧客保留率(Strouse, 1999)。透過成本的節省、顧客重複購買與購買量的增加與口碑效果，若能減少 5%的流失率，可使獲利率增加25%到85%(Reichheld and Sasser, 1990)，而企業如能有效延緩顧客流失的速度，預估可增加15%的長期收益(Bolton, 1998)。

維持既有客戶的企業成本遠低於開發新客戶群的成本(Heskett et al.,1989)，就利潤面考量，既有客戶對於企業獲利之貢獻更遠多過新開發客戶群，口耳相傳的正面宣傳效果替企業提高新顧客流量(Keaveney,1995)。顧客忠誠度可為企業帶來實質利益，企業一年的顧客保留率增加 5%，平均可為公司帶來多達 75%的總利潤，甚或是 100%的企業利潤(Reichheld et al.,1990)。由此可見，促使顧客流失情形的減少均能夠創造出高收益和低成本的利益(Fonell and Wernerfelt, 1987)。與客戶建立緊密且長久持續的關係，是企業創造利潤的重要關鍵，如何預防客戶流失將特別重要。

雖然學校性質不似企業，然而隨招生競爭日趨強烈，顧客關係管理 (Customer Relationship Management, CRM) 亦逐漸被運用在學校方面。從 CRM 的概念來看，中途離校生即等於客戶流失，若能運用顧客關係

管理的方式經營與學生的關係，從整體來看對學校經營應有所助益。

「流失」一詞普遍被運用在地質學及團體輔導的領域中，在英文中，有幾個關於成員流失的用字，例如："drop out"、"withdraw"、"premature termination" 及 "casualty" 等，但其內涵上有些許的差異(盧梅莉，民 81)。王智弘(民 77、民 83)曾論及「流失」一詞的定義：對於團體成員的流失在英語上的用字是"drop out"也就是「中途退出」的意思，而另外常用的字還包括"premature termination"（未成熟的終結）與"casualty"（成員的傷亡）等，一般對團體成員流失的定義是：「在事先約定的團體過程結束之前，團體成員不顧諮商員或團體領導員的勸告而中途離開團體」，不過對於"casualty"一字，Garfield & Bergin (1978) 則建議要與"premature termination"一字加以區別，認為：「並非所有離開團體的人都可稱之為『成員之傷亡』，有些離開的成員並非因受到團體經驗的傷害而離開，而可能只是因為對團體有不適當的期望或團體領導的不夠敏感，而感到團體不能滿足其需要而已」。也就是說"casualty"一詞更專指那些受到團體經驗傷害而中途流失的成員(引自王智弘，民 83)。

根據上述對「流失」一詞用語的相關討論，以"drop out"也就是「中途退出」的解釋，所謂「流失」(drop out) 係指團體成員在完成團體訓練或輔導的過程中，團體成員中途退出團體組織的情形，較能合乎本研究的意涵。所以在本研究中對流失的定義為：『曾經註冊繳費，擁有本校學籍資料，因故無法完成學業中途退出，未取得畢業證書之學生』。

第二節 資料採礦定義

Data mining 經中華資料採礦協會 (Chung-Hua Data Mining Society) 譯為「資料採礦」。依中華資料採礦協會 (2002) 指出資料採礦最早由

Usama Fayyad(1991)提出，其目的為從龐大的維修資料中，找出規則。

資料採礦是一大量自動化的過程，其運用統計分析來從大量的資料集合中，發現有用的、不明顯的和先前未知的特徵或資料趨勢 (Frawley et al., 1992)。其實，資料採礦並非是一種技術或者是一套軟體。事實上，它是一種結合數種專業技術的應用。並且不是無所不能，它只是從大量的資料中發掘出各種假設 (Hypothesis)；但是，它並不會幫忙檢查，也不會幫忙確認假設。同時，無法協助判斷這些假設對使用者的價值為何。(謝邦昌、鄭宇庭、蘇志雄，2011)。Kleissner (1998)則表示，資料採礦是去發現公司資料中所隱含的知識並讓企業的管理者能夠瞭解，而來支援決策的分析過程。

藉由資料採礦的技術，可以增進對顧客需求和行為的瞭解，並有助於企業提供客製化的服務，強化與顧客之間的連結、溝通與互動 (Cheng et al., 2005)，亦即可發掘大量關於顧客特徵和購買模式而有益於行銷的知識 (Shaw et al., 2001)，多數公司運用資料採礦作為策略的基礎，協助其打敗競爭者、確認新顧客以及降低成本 (Davis, 1999)。

一般來說，大致上常用的資料採礦技術主要有下述六類(許依宸，民國 98)：

- 壹、分類 (Classification)：預測類別變數的過程，我們稱之為「分類」。依據已知的資料及其分類的屬性，建立出資料的分類模型；接著，利用此分類模型來預測新資料的類別。例如：顧客的購物習性分類模型..等。
- 貳、推估 (Estimation)：運用處理過後連續性數值的結果，給定一些輸入資料以推估未知的連續性變數的值。例如：金融商品價格之預測...等。

- 參、群集化 (Cluster)：群集意為物以類聚。即依資料本身的自我相似性 (self-similarity) 而群集在一起。群集 (Clusters) 的意義則要經由事後之闡釋方能得知。
- 肆、關聯法則 (Association Rule)：指的是從歷史資料中，找出哪些事件總是相伴發生。
- 伍、序列 (Sequential)：即在同質分組中透過序列來找出事物「先後」發生的順序，這樣的規則被稱為時序規則 (Sequential Pattern)。
- 陸、描述 (Description)：指的是在資料採礦的過程中，除了分析的預測模型以外，更重要的是在分析與處理資料的過程中，透過資料視覺化以及觀察來找出許多有意義的規則。

第三節 資料採礦的應用

其實，資料採礦並非不只是一種技術或者是一套軟體。而是一種結合數種專業技術的應用 (謝邦昌等人, 2011)，資料採礦技術因依不同的特性而適用於不同之領域，例如：

- 壹、醫療生技業-預防醫學分析、院內感染分析、臨床病徵分析、基因圖譜比對、基因定序、演化分析。
- 貳、零售業者而言-分析購買行為、瞭解顧客消費模式強化客戶關係，達到留住顧客目的。
- 參、金融保險業-信用評等、客製化金融服務、客戶資產管理、呆帳分析、保險潛在客戶名單分析、分析購買行為、偵測信用卡詐騙行為、股匯市行情預測。
- 肆、教育業-學生來源分析、課程規劃、學習評量、適性化教學、學生異動分析。

伍、零售製造業-分店設點區位分析、銷售產品組合、庫存管理、即時輔助購買決策、連續銷售、促銷商品組合、DM 名單、庫存分析、排程技術生產。

陸、銀行業者而言-瞭解信用卡發放可能產生之弊端，找出最有利潤、忠誠度佳的顧客、呆帳分析。

柒、保險業者而言-分析保戶要求理賠之模式，並可加強稽核，以防止詐財之發生。

第四節 決策樹

所謂的決策樹，指的是從一個或多個預測變數中，針對類別應變數的階級，預測案例或物件的關係（會員數）是資料採礦其中一項主要的技巧（謝邦昌等人，2011）。相較於其他的演算法，決策樹的確是比較容易讓人解讀及理解變數的因果關係，而模型的預測成效只是資料採礦的一環而已。對於企業決策者來說，更重要的是能否從規則的內容中獲得啟發（尹相志，2007）。

決策樹是功能強大且相當受歡迎的分類和預測工具。這項以樹狀圖為基礎的方法，其吸引人之處在於決策樹具有規則，和類神經網路不同。規則可以用文字來表達，讓人類了解，或是轉化為 SQL 之類的資料庫語言，讓落在特定類別的資料紀錄可以被搜尋。

決策樹是一項建立分類模式(classification models)的方式之一，針對給定的資料利用歸納的方式產生樹狀結構的模式。為了要將輸入的資料分類，決策樹的每一個節點即為一個判斷式，判斷式針對一個變數去判斷輸入的資料大於或等於或小於某個數值，每一個節點因而可以將輸入的資料分成若干類。

決策樹修剪技術可區分為兩種，一是事前修剪（Prepruning）即在訓練模型的同時，將該節點設為「葉節點（Leaf Node）因此該節點即停止生長；另一種指的是事後修剪（Postpruning），就像園藝中在修剪盆栽一樣，將已產生的決策樹多餘的規則修剪掉（尹相志，2007）。

決策樹的階層數不宜過多或是過少。如果過多階層數，則表示其分割過多，所產生出的規則也會失去其原始功用；反之，若決策樹的階層過少，則意謂著分割過程太早結束，而此時所建構的模型也未必會產生良好的分類規則。

第五節 相關研究回顧

許依宸（民 98）資料採礦在學生流失偵測上之應用，探討影響學生流失的因素，以學號、性別、科系、入學方式、居住地區、歷年學業平均成績、平均每學期不及格學分數等欄位，利用決策樹及類神經網路兩種資料採礦技術來找出探討影響學生流失的因素。其研究發現決策樹及類神經網路都以歷年學業平均成績為分類的重要因素，顯示成績確為學生流失與否的主要因素。

王建華（民 93）探討中途離校生之出缺席狀況及學業成績關聯，找出技職五專學生其行為特徵並比較與人格測驗之分析結果，其使用變項為德行成績（出席記錄、個人因素）、學業成績，運用決策樹 C4.5 分析缺曠課情形（包含請假原因及節次，找出請假缺課最嚴重的節次）及學業成績[分學期、分數（0 分、及格、不及格）、科系]，其研究發現中途離校生缺課情形嚴重且成績較差。

第三章、研究方法

本章節提出本研究之架構並說明資料處理之方式，以作為後續研究之基礎。首先在第一節說明研究對象之概況；第二節提出本研究之流租；第三節說明資料處理方式；第四節針對資料採礦工具進行介紹。

第一節 研究對象

本研究以 102 學年度入學大學日間部資管系學生為主研究對象，由於研究樣本選取時間為 105 學年度第 2 學期，故其就學狀態為 102 學年度第 1 學期至 105 學年度第 2 學期時之就學狀態。個案學校學生休學是採填寫申請表方式，而學生填寫休學申請單中所填寫之內容，學生多數僅填寫個人因素，故無法獲得更多更明確之訊息；而校務行政系統上登載之學生離校原因包括-休學、退學及轉學等，轉學分為轉入與轉出。由校務行政系統上並無法確實得知學生離校之實際原因，故本研究採用資料採礦方式，擬探求學生流失率，希望藉由發現各多元入學管道之學生學習表現與學生流失率是否相關，提供個案學校作參考。

第二節 研究流程

本研究的主要研究變數為入學管道、性別、居住地區、學業成績，主要為校務行政系統上之學生資料。然後用決策樹來分析資料，找出各入學管道學生流失率關聯。本研究之研究流程如圖 3-1 所示，確定研究動機、探討相關文獻、分析資料特性、資料前置處理、資料特性分析、決策樹分析、結論與未來工作。

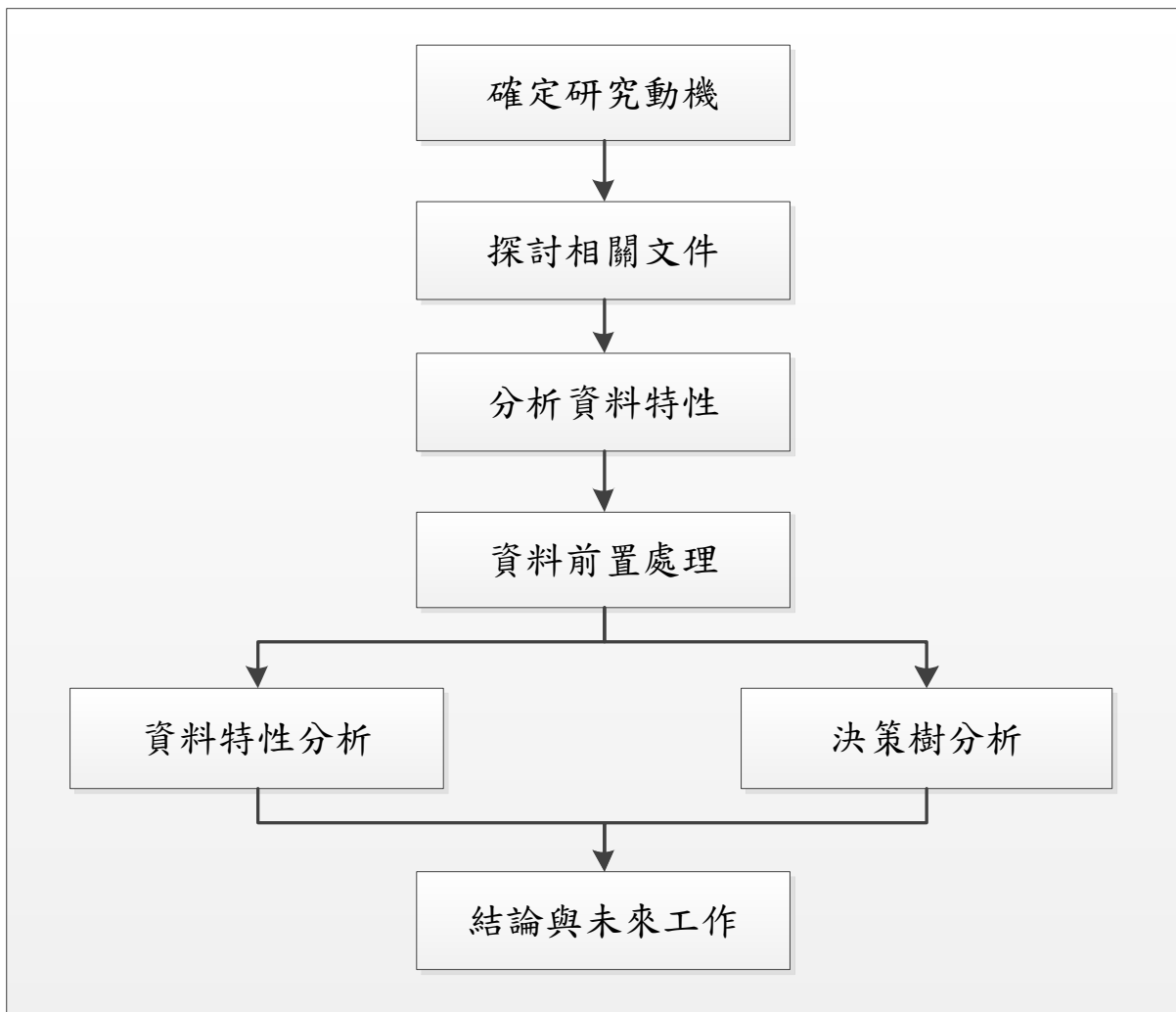


圖 3-1 研究流程圖
資料來源：本研究整理

第三節 資料前置處理

本研究將流失界定為：「曾經註冊繳費，擁有學籍資料之本校學生，因故無法完成學業，中途退出，未取得畢業證書之學生」。依本研究流失之定義，將所取得的學生資料中異動原因為休學、退學、轉學(轉出)之學生，其就學狀態為不在學。

了解每個資料欄位所代表的含意與其使用之價值後，由於考慮到資料庫中有部分欄位不具分析價值，如電話、身份證號碼、家長姓名等，故將不具分析價值的欄位予以刪減，以避免造成分析時的負擔與影響資料採礦分析結果之準確性。

本節就資料之前處理及後續分析之使用指標分別說明如下：

壹、資料前置處理主要包括：

一、資料淨化：

主要是確認資料的完整性及正確性。將用不到的資料欄位予以刪除，如學生姓名、身份證字號、電話、地址等；同時刪除在就學狀況進進出出之狀況學生刪除。

二、資料轉換：

為使資料內容更容易資料採礦之進行，將部分資料進行轉換，如居住地址轉為縣市別；地址資料中之臺與台統一改為台字；桃園縣與桃園市縣市合併後，統一更改為桃園市。另將每學期學業成績計算出學業平均。

三、資料整合：

將大學四年的 8 學期學業成績與最後的總成績進行整合，以利進行相關後續分析。

貳、資料處理後學生資料表之欄位說明如表 3-1 學生資料表表 3-1 所示

表 3-1 學生資料表

欄名	屬性	長度	說明
學號	文字	8	102XXXXX

性別	文字	1	男、女
居住地區	文字	3	XX 縣、XX 市
各學期學業成績	數字	浮點	
班級	文字	1	A、B
入學方式	文字	4	考試分發、個人申請 繁星推薦、轉學考
就學狀態	文字	2	休學、在學、退學、轉學
學期總成績	數字	浮點	

資料來源：本研究整理

第四節 資料採礦工具

本研究採用 SQL Server 為資料採礦工具。SQL Server 不僅擁有有決策樹與群集演算法之外，更有類神經網路、線性迴歸、羅吉斯迴歸、貝氏機率分類、關聯規則、時序群集以及時間序列等演算法資料採礦工具，也有文字探勘工具，且其豐富的視覺化呈現不僅能讓分析者深入瞭解模型規則的內容，更可透過互動的機制，讓使用者深入瞭解潛藏在模型中的趨勢。SQL Server 資料庫引擎除了提供關聯式和結構式資料外也能在資料庫中儲存半結構化與非結構化的文件，讓應用範圍更廣。

SQL Server 中為 IT 部門提供一個更符合成本效益、可擴展且易於管理的高性能資料庫平台。它可管理大型多資料庫環境與改進資料庫的能力，能確保簡化開發和部署數據驅動的應用程序。SQL Server 更可將功能透過雲端延伸到 SQL Server 平台上，透過分佈式的數據服務，提供企業更多的可用性與商機（謝邦昌等人，2011）。

SQL Server 的優勢：

壹、易於使用，不需要任何專門知識任何分析人員、決策者、資料庫管理員皆能操作。

貳、簡單而豐富的 API。

參、演算法允許的整個資料集運行，不需創建樣本集在所有資料上都有效。

肆、控制成本，同時兼顧效能、可用性、延展性和安全性。



第四章、資料分析

本章共有二節，第一節資料特性分析，針對所取得之資料進行描述性統計，說明所取得資料之概況，第二節決策樹分析，針對所取得之資料進行決策樹分析。

第一節 資料特性分析

壹、資料說明：

本研究所得之資料為 102 學年度第 1 學期至 105 學年度第 2 學期大學日間部資管系 102 級學生資料，共取得 76 筆，其中男生為 65 人、女生為 11 人。102 級資管系男、女生比例如圖 4-1 所示。

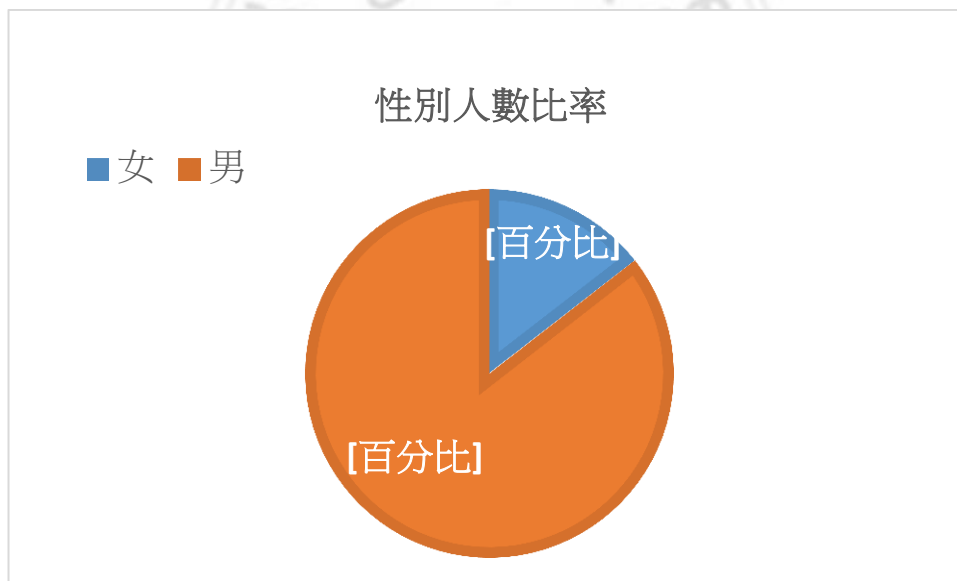


圖 4-1 102 級資管系男女生比例圖

資料來源：本研究整理

表 4-1 102 級資管系學生男、女生人數表

級別	男	女	合計
102 級	65	11	76

資料來源：本研究整理

貳、流失狀態分析：

本研究資料中流失狀態未流失有 60 人、流失 16 人，由圖 4-2 學生流失狀態表可看出流失比例為 21%。

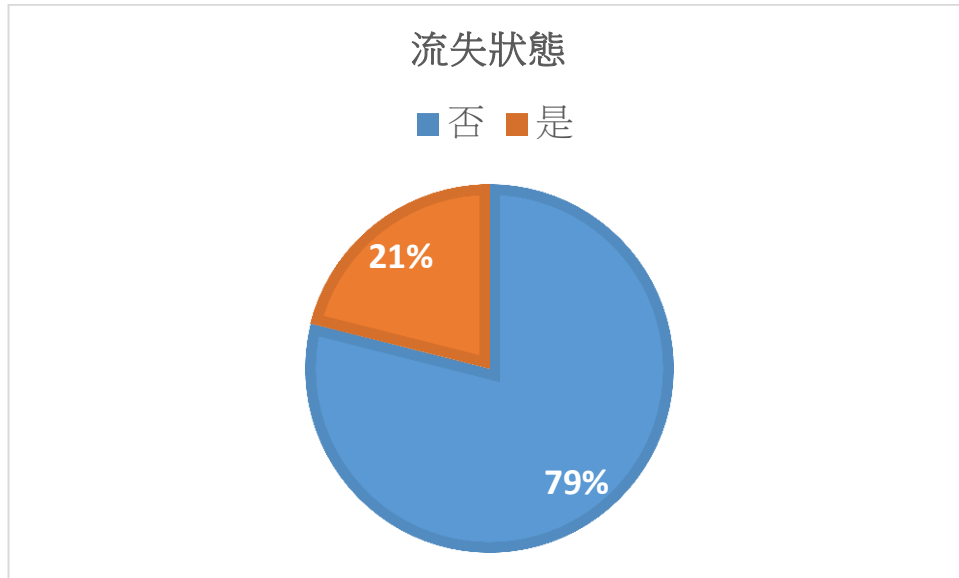


圖 4-2 學生流失狀態圖
資料來源：本研究整理

參、就學狀態分析

由圖 4-3 就學狀況比例圖中，可看出不在學的人數比例中轉學比例 12.1%最高。

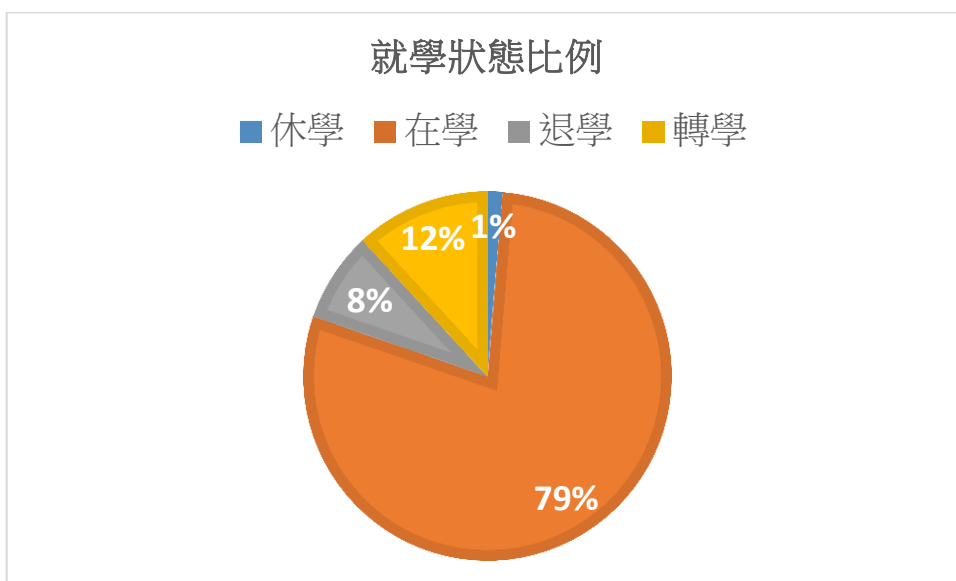


圖 4-3 就學狀況比例圖

資料來源：本研究整理

若以就學狀況來說，綜合四種情況而言，在學人數比例 79%，尚高於不在學(休學、退學、轉學)比例 21%，如表 4-2 所示。

表 4-2 不同就學狀況比例表

就學狀況	休學	在學	退學	轉學
人數	1	60	6	9
比例	1%	79%	8%	12%

資料來源：本研究整理

肆、各入學方式與學業成績表現分析

由表 4-3 及圖 4-4 看出，繁星推薦學生學業表現仍為各入學管道中，學業表現較為突出部分；反觀轉學考進入學校之學生學業表現仍為最弱。

表 4-3 入學方式學業表現一覽表

入學方式	102 平均成績	103 平均成績	104 平均成績	105 平均成績
考試分發	76.78888889	76.54769231	77.44685714	79.66671429
個人申請	71.57277778	70.12217391	76.475	78.65571429
繁星推薦	78.435	80.68	81.12	81.585
轉學考	NULL	73	60.95166667	64.97166667

資料來源：本研究整理

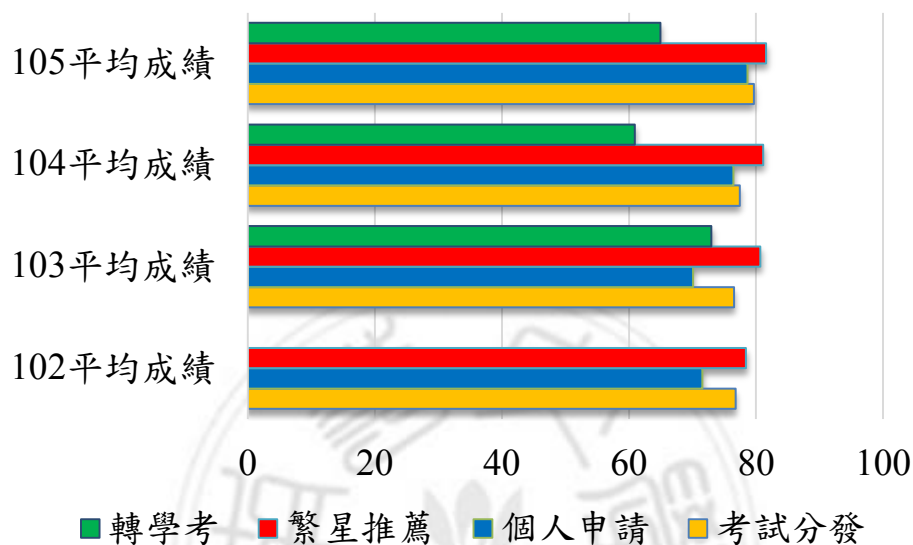


圖 4-4 入學方式與學業成績分析圖

伍、居住地區就讀人數與流失人數資料分析

由表 4-4 可看出，以高雄市就讀人數 17 人最多，但流失人數僅有 1 人，推估表示於高雄市地區仍占有較大比例招生市場；反觀於嘉義縣地區雖就讀人數非為最少，但在流失比例上卻是最高，其流失人數與居住地區相關。

表 4-4 居住地區就讀人數與流失人數表

居住地區	就讀人數	流失人數
台中市	5	1
台北市	3	1
台南市	13	2
花蓮縣	3	1
屏東縣	8	2
苗栗縣	1	0
桃園市	4	1
高雄市	17	1
基隆市	1	0
雲林縣	4	1
新北市	5	2
新竹市	1	0
嘉義市	1	0
嘉義縣	5	3
彰化縣	5	1

資料來源：本研究整理

陸、入學方式與流失人數資料分析

由表 4-5 可看出，在繁星推薦及轉學考中，學生的流失人數為 0，基本上多為穩定；而在考試分發及個人申請上則流失比例均為 22.22% 左右，顯示出流失學生多為此兩類學生，其流失人數與入學方式相關。

表 4-5 入學方式流失人數統計與比例表

入學方式	人數	流失人數	比例
考試分發	45	10	22.22%
個人申請	27	6	22.22%
繁星推薦	1	0	0.00%
轉學考	3	0	0.00%

資料來源：本研究整理

柒、居住地區、入學方式、就讀人數與流失人數分析

由表 4-6 可看出，以高雄市地區考試分發 11 人，流失 1 人，學生就讀狀況最為穩定，推估此地區將來可作為招生安排較為頻繁且有效地區；而以就讀人數與流失人數來看，以嘉義縣個人申請 3 人，流失 2 人，其流失比例最為嚴重；其流失人數與居住地區、入學方式相關。

表 4-6 居住地區、入學方式、就讀人數與流失人數統計表

居住地區	入學方式	人數	流失人數
台中市	考試分發	4	0
台中市	個人申請	1	1
台北市	考試分發	3	1
台南市	考試分發	5	0
台南市	個人申請	8	2
花蓮縣	考試分發	2	1
花蓮縣	個人申請	1	0
屏東縣	考試分發	5	2
屏東縣	個人申請	2	0
屏東縣	繁星推薦	1	0
苗栗縣	轉學考	1	0
桃園市	考試分發	1	0
桃園市	個人申請	3	1
高雄市	考試分發	11	1
高雄市	個人申請	6	0
基隆市	考試分發	1	0
雲林縣	考試分發	2	1
雲林縣	個人申請	1	0
雲林縣	轉學考	1	0
新北市	考試分發	5	2
新竹市	個人申請	1	0
嘉義市	考試分發	1	0
嘉義縣	考試分發	1	1
嘉義縣	個人申請	3	2
嘉義縣	轉學考	1	0
彰化縣	考試分發	4	1
彰化縣	個人申請	1	0

資料來源：本研究整理

第二節 決策樹分析

壹、資料說明

將資料先行預處理，經處理後所要分析的資料欄位說明如表 4-6。

表 4-7 決策樹分析之欄位

屬性	欄名	說明
Key	學號	
Input	性別	男、女
Input	班級	A、B
Input	入學方式	考試分發、個人申請 繁星推薦、轉學考
Input	居住地區	
Input	學業總成績	
PredictOnly	流失	是、否

資料來源：本研究整理

貳、資料採礦作業：

以下說明進行資料採礦作業步驟如下：

- 一、在 Microsoft SQL Server Management Studio 上建一資料庫，再到 SQL Server Business Intelligence Development Studio 中選 Integration Services 專案，將資料匯入資料表。
- 二、在 SQL Server Business Intelligence Development Studio 中新增一個 Analysis Services 專案，將資料來源設為剛匯入的資料庫，然後新增資料採礦結構，資料結構設為決策樹。
- 三、訓練模型資料量 80%約 61 筆；測試資料量 20%約 15 筆。
- 四、模型參數解釋如下：

(1) Complexity Penalty；複雜性懲處

- (1.1) 參數值介於 0-1 之間，不可等於 0 或 1
- (1.2) 當值接近 1 時，代表提供較大的懲處
- (1.3) 決策樹成長受到較多限制，產生分岔較少樹狀規則
- (1.4) 反之決策樹規則越龐雜。

(2) Minimum Support：符合規則節點最小案例數

(2.1) 較此值低時，規則將被系統自動刪減

(2.2) 此值小於 1 之小數時，代表最小案例數目為總案例的百分比(如值為 0.01 時，代表每個規則案例數至少必須占總體案例的 1%)

(2.3) 值>1 整數時，代表最小案例數為指定的絕對數目

五、先以預設參數建立一個採礦模型後，再新增採礦模型更改參數。

(一) 預設的參數建立 dropout 決策樹模型後，先設定參數 COMPLEXITY_PENALTY 為 0.5 建立 dropout1 模型，MINIMUM_SUPPORT 設為 5 最小案例數後，發現分類在屬性第 2 子層學業總成績下，已無其他屬性可分，如

圖 4-5。

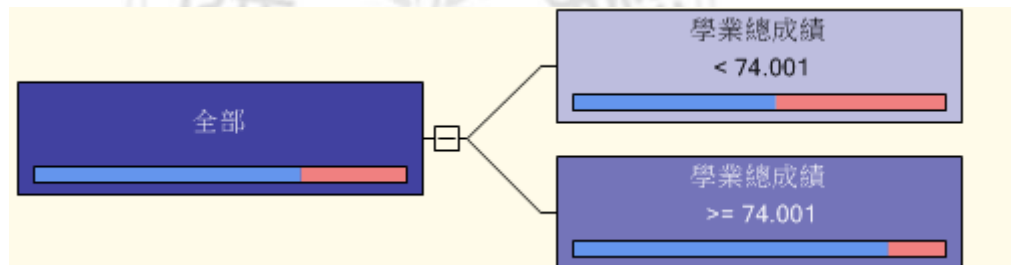


圖 4-5 決策樹模型 1

參數: Complexity Penalty: 0.5, Minimum Support: 5

資料來源：本研究整理

(二) 接著設定參數 COMPLEXITY_PENALTY 為 0.5 建立 dropout2 模型，MINIMUM_SUPPORT 設為 3 降低最小案例數後，發現分類在屬性第 2 子層學業總成績下，已無其他屬性可分，如

圖 4-6。

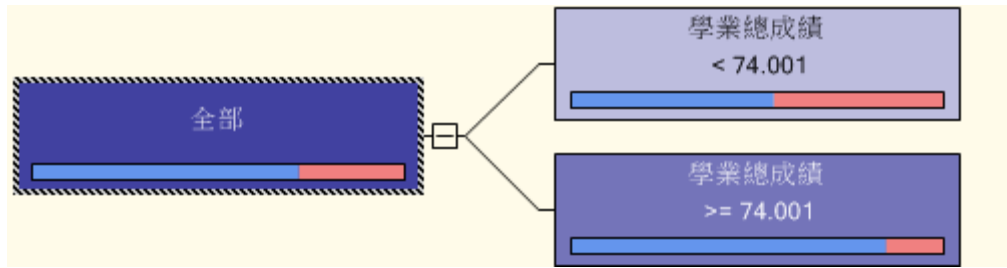


圖 4-6 決策樹模型 2

參數: Complexity Penalty: 0.5, Minimum Support: 3

資料來源: 本研究整理

(三) 將 COMPLEXITY_PENALTY 調整為 0.1，建立 dropout3 模型、MINIMUM_SUPPORT 維持在 3 後，發現分類在屬性第 2 子層學業總成績下，尚有第 3 子層入學方式、第 4 子層居住地區及第 5 子層性別屬性可分，如

圖 4-7。

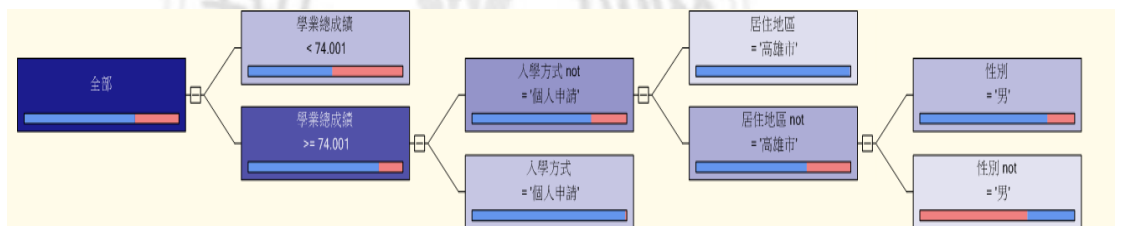


圖 4-7 決策樹模型 3

參數: Complexity Penalty: 0.1, Minimum Support: 3

資料來源: 本研究整理

(四) 接著再將 COMPLEXITY_PENALTY 調整為 0.05，建立 dropout4 模型、MINIMUM_SUPPORT 維持在 3 後，發現分類在屬性第 2 子層學業總成績下，尚有第 3 子層入學方式、第 4 子層居住地區及第 5 子層性別屬性可分，如

圖 4-8。

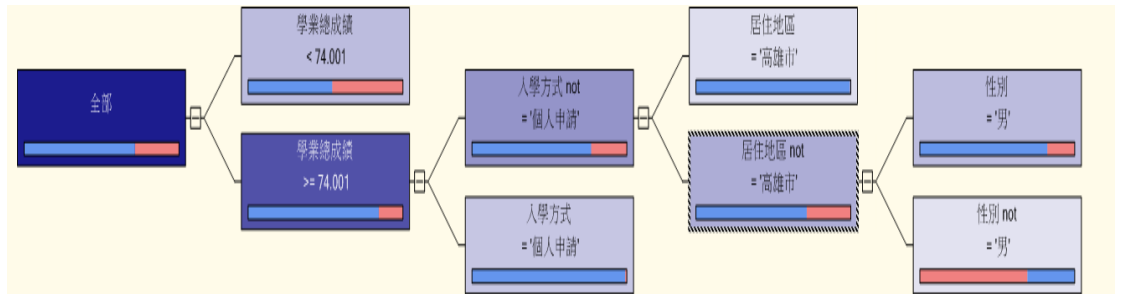


圖 4-8 決策樹模型 4

參數: Complexity Penalty: 0.05, Minimum Support: 3

資料來源: 本研究整理

(五) 持續將 COMPLEXITY_PENALTY 固定為 0.05，建立 Droupou5 模型、MINIMUM_SUPPORT 調整為 5 後，發現分類在屬性第 2 子層學業總成績下，尚有第 3 子層入學方式、第 4 子層居住地區可分，如

圖 4-9。

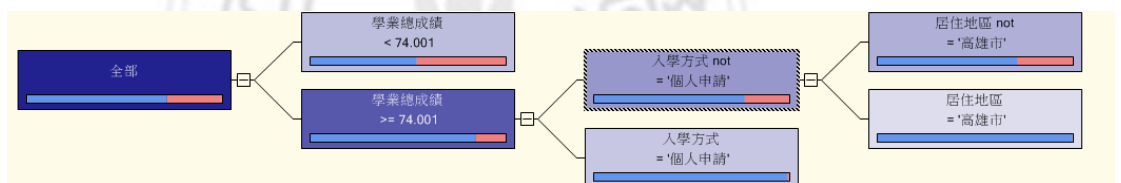


圖 4-9 決策樹模型 5

參數: Complexity Penalty: 0.05, Minimum Support: 5

資料來源: 本研究整理

(六) 再持續將 COMPLEXITY_PENALTY 固定為 0.05，建 Droupou6 模型、MINIMUM_SUPPORT 調整為 10 後，發現分類在屬性第 2 子層學業總成績下，尚有第 3 子層居住地區及第 4 子層班級屬性可分，如圖 4-10。

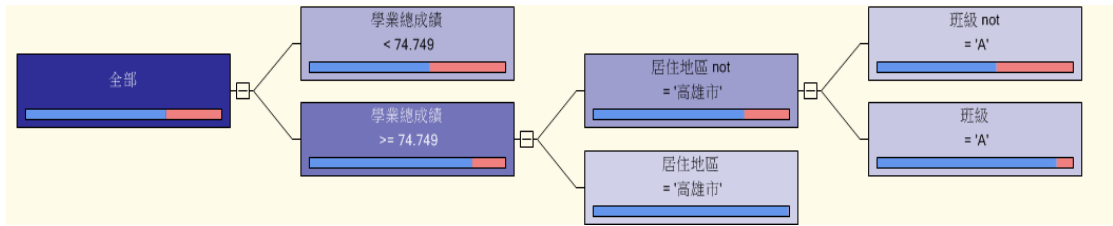


圖 4-10 決策樹模型 6

參數: Complexity Penalty: 0.05, Minimum Support: 10

資料來源: 本研究整理

(七) 以測試資料進行測試，建立測試資料各模型之分類矩陣並比較，由各分類矩陣之比較發現，在預測與實際之分類矩陣中 dropout3 與 dropout4 之兩者否定值較小，與建立模型較為吻合，如表 4-7 所示。



表 4-8 各模型分類矩陣比較表

Dropout1 決策樹模型 COMPLEXITY_PENALTY=0. 5 MINIMUN_SUPPORT=5	預測的	否 (實際)	是 (實際)
	否	13	2
	是	0	0
Dropout2 決策樹模型 COMPLEXITY_PENALTY=0. 5 MINIMUN_SUPPORT=3	預測的	否 (實際)	是 (實際)
	否	13	2
	是	0	0
dropout3 決策樹模型 COMPLEXITY_PENALTY=0. 1 MINIMUN_SUPPORT=3	預測的	否 (實際)	是 (實際)
	否	11	2
	是	2	0
dropout4 決策樹模型 COMPLEXITY_PENALTY=0. 05 MINIMUN_SUPPORT=3	預測的	否 (實際)	是 (實際)
	否	11	2
	是	2	0
Dropout5 決策樹模型 COMPLEXITY_PENALTY=0. 05 MINIMUN_SUPPORT=5	預測的	否 (實際)	是 (實際)
	否	13	2
	是	0	0
Dropout6 決策樹模型 COMPLEXITY_PENALTY=0. 05 MINIMUN_SUPPORT=10	預測的	否 (實際)	是 (實際)
	否	13	2
	是	0	0

資料來源：本研究整理

六、採礦結果與圖表

本研究依據各模型的樹狀圖來看 (圖 4-5~圖 4-10)，發現無論各樹狀圖分層情形如何，第1子層皆為學業總成績，第2子層才有其他因素加入，故本研究推論在所有影響就學狀況的因素中，以學業總成績為最重要的因素；入學方式與居住地區為其次因素；再其次為性別與班級。

第五章、結論與未來工作

第一節 研究結論與建議

為了探討學生的流失率與入學方式的相關性，本研究試圖在學的學生資料，找出不同的入學方式學生其學業成績表現與學生流失的關聯，以提供給學校老師於教學過程中注意學生的學習表現；並協助學生的學習過程以預防學生流失。

壹、在資料特性分析中發現：

- 一、在多元入學學生學生學業表現上，繁星推薦學生學業表現仍為各入學管道中，學業表現較為突出部分；反觀轉學考進入學校之學生學業表現仍為最弱。
- 二、在多元入學學生之流失率分析中發現以下狀況：
 - (1) 在就學狀況中，在休學、退學及轉學方面流失率，以轉學 12.1% 最高。
 - (2) 在居住地區就讀人數與流失人數部份，以高雄市就讀人數最多，且流失人數比例偏小；反觀於嘉義縣地區雖就讀人數非為最少，但在流失比例上卻是最高，其影響流失率程度與居住地區有相關。
 - (3) 在入學方式與流失人數部份，在繁星推薦及轉學考中，學生的流失人數為 0，基本上多為穩定；而在考試分發及個人申請上則流失比例較為偏高，顯示出流失學生多為此兩類學生，其影響流失率程度與入學方式有相關。
 - (4) 在居住地區、入學方式、就讀人數與流失人數部份，以高雄市地區考試分發最多人，流失人數偏低，學生就讀狀況最為穩定；而以嘉義縣個人申請人數及流失人數比例最高，推估其影響流

失率程度與居住地區及入學方式有相關。

貳、在資料採礦模型中發現影響學生流失率最重要的是學業成績總成績，其次是居住地區與入學方式，再其次是性別與班級。

第二節 研究未來工作

壹、對於其他影響學生流失的相關因素如：家庭狀況、經濟狀況、交友情形等本研究未加入考量，建議後續可加入進行相關研究。

貳、限於本研究只取得 102 級入學資管系學生資料，資料範圍未能擴大，建議後續可增加研究樣本數量以推論與預測影響學生流失率因素。

參、針對本研究給研究學校的建議：

- 一、對於轉學生而言，剛進入學校環境比較陌生難免有適應不良的地方，校方應隨時注意學生狀況，適時給予學習輔導。
- 二、善加利用課業輔導及預警制度，將學業表現較低學生列入輔導，進而提高學生學習興趣，降低學生流失率。
- 三、另外有關多元入學與學生流失關係，可以作為招生時，各管道招生員額設定及重點經營地區，進而提高註冊率。(如研究中發現繁星推薦較不易流失及高雄地區來源學生較為穩定)

參考文獻

一、 中文部份

1. 中華資料採礦協會<http://www.cdms.org.tw/xoops2/html/modules/news/>
2. 尹相志，SQL Server 2005 Data Mining 資料採礦與 Office 2007 資料採礦增益集，悅知文化，台北，2007。
3. 王建華(民 93)，「資料挖掘技術再技職院校中途離校生輔導之應用--以醒吾技術學院為例」，國防管理學院國防資訊研究所碩士論文。取自<http://ndltd.ncl.edu.tw/cgi-bin/gs32/gsweb.cgi?o=dnclcdr&s=id=%22092NDMC1654030%22.&searchmode=basic>
4. 王智弘，「探討團體成員流失問題之原因、影響及因應對策」，輔導月刊，24 卷 6 期，37-40 頁，民 77。
5. 王智弘，「團體成員流失問題之探討」，輔導季刊，30 卷 4 期，27-36 頁，民 83。
6. 許依宸(民 98)，「資料採礦在學生流失偵測上之應用」，南華大學資訊管理學系碩士班碩士論文。
7. 陳芳君(民 101)，「以資料採礦探討學生流失及其相關因素之研究」，南華大學資訊管理學系碩士班碩士論文。
8. 盧梅莉「團體成員流失之探討」諮商與輔導，76 期，34-36 頁，民 81。
9. 謝邦昌、鄭宇庭、蘇志雄，SQL Server 2008 R2 資料採礦與商業智慧，基峰資訊，初版，台北市，2011。

二、西文部份

1. Berson, A., Smith, S. and Thearling, K., "Building Data Mining Applications for CRM", Customer Retention, New York, McGraw-Hill, 2000.
2. Bolton, Ruth N,"A Dynamic Model of the Duration of the Customer's Relationship with A Continuous Service Provider:The Role of Satisfaction," Marketing Science, Vol. 17,No. 1,pp.45~65, 1998.
3. Davids, M., "How to avoid the 10 Biggest Mistake in CRM", Journal of Business Strategy,Vol.4,pp.22-26,1999.
4. Fonell, Clases, and Birger Wernerfelt "Defensive Marketing Strategy by Customer Complaint Management: A Theoretical Analysis", Journal of Marketing Research, 24, pp. 337-346, November 1987.
5. Frawley, W., Piatetsky-Shapiro, G. and Matheus, C., "Knowledge Discovery in Databases: An Overview", AI Magazine, pp. 213-228 , Fall 1992.
6. Heskett, James L. W.Earl Sasser Jr., and Leonard A. Schlesinger, The Service Profit Chain, Free Press, New York, 1997.
7. Keaveney, Susan M, "Customer Switching in Service Industries:An Exploratory Study", Journal of Marketing,Vol.59,pp.71~82. April 1995.
8. Kleissner, C., "Data mining for the enterprise", Proceedings of the Thirty-First Hawaii International Conference, pp.295-304, 1998.
9. Reichheld, Fredirick F. & Sasser, W.E. Jr., Zero Defection: Quality Comes to Services, Harvard Business Review, Vol. 68, No.5, pp. 105~111, 1990.
10. Shaw, M. & Subramaniam, C, "Knowledge management and data mining for marketing", Decision Support Systems, pp. 127~137, 2001, 31.
11. Strouse, Karen G, Marketing Telecommunications Services New Approaches for A Changing Environment, Boston:Artech House.1999.