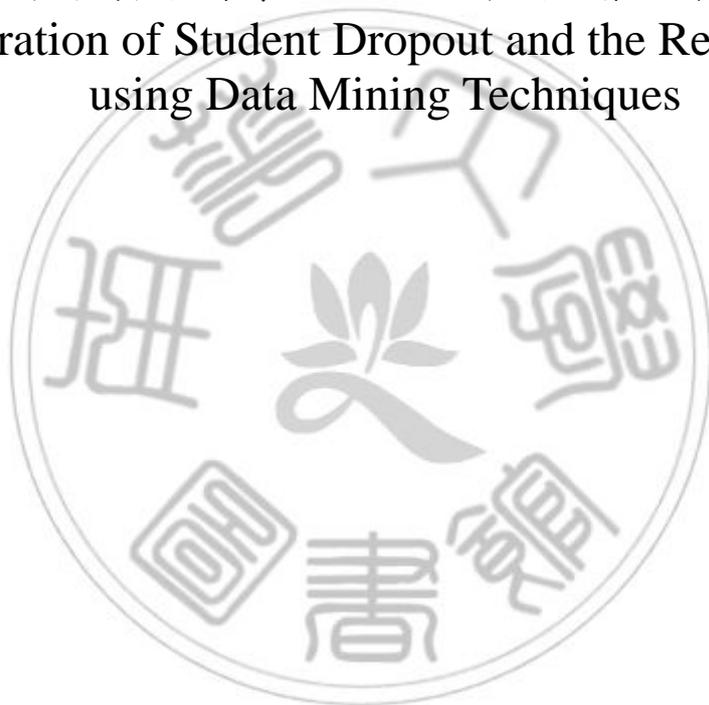


南 華 大 學

資訊管理學系

碩士論文

以資料採礦探討學生流失及其相關因素之研究
An Exploration of Student Dropout and the Related Factors
using Data Mining Techniques



研 究 生：陳芳君

指 導 教 授：邱宏彬

中 華 民 國 101 年 7 月

南 華 大 學

資訊管理研究所

碩 士 學 位 論 文

以資料採礦探討學生流失及其相關因素之研究

An Exploration of Student Dropout and the Related Factors

using Data Mining Techniques

研究生：陳芳君

經考試合格特此證明

口試委員：陳博言

李翔

邱宏樹

指導教授：邱宏樹

系主任(所長)：資訊管理學系 吳光閔 系主任

口試日期：中華民國 101 年 6 月 30 日

誌 謝

能夠完成本文要感謝的人很多，要不是有你們的協助是無法完成的。首先要感謝吳光閔所長幫忙，讓我能夠有完成論文的机会。

謝謝指導教授邱宏彬老師，他一直很有耐心和愛心不斷的鼓勵，讓我克服論文寫作的盲點，在論文出現瓶頸時不厭其煩的給予指導，讓學生深深感受到老師的教誨與包容，銘記在心。同時也要謝謝口試委員李翔詣教授和陳張宗榮教授寶貴的建議與指導，讓論文能更加完備，特此感謝。

感謝同學妃娜在接近完成的期間給我許多的協助和相互鼓勵；學妹思予幫忙減少了許多申請上的手續，讓我有更多的時間準備。

最後要感謝我的家人，家人是我能完成這篇論文的動力。謝謝我的爸爸、媽媽和公公的支持與協助；老公富仁在口試期間幫忙帶兩個小寶貝，讓我能全力以赴。

謹在此感謝許多陪我渡過這些日子幫助我的人，獻上我最深的感謝與祝福。

芳君

謹誌於南華大學資訊管理學系碩士班

2012/07

以資料採礦探討學生流失及其相關因素之研究

學生：陳芳君

指導教授：邱宏彬 博士

南 華 大 學 資 訊 管 理 學 系 碩 士 班

摘 要

有鑑於國內少子化的影響，高中職新生數劇減在民國 102 年呈現，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。本研究以一私立高職 97 至 99 學年度學生歷史學籍資料，運用資料採礦技術分析，找出流失學生之潛在共同因素，提供相關之建議以降低學生流失。

本研究共取得有效資料 1465 筆，結果顯示影響學生流失的主因是德行平均，1 年級學生轉換環境居多，2、3 年級則以轉學、休學為主。給個案學校的建議：1、對於 1 年級學生應校方適時給予輔導以免學生德行成績過低而被改變學習環境。2、在校生則是增加對學校的向心力，以減少學生流失。3、對於建教班學生學校可以給予更多的輔導和協助。4、可建立預警機制，將德行成績較低及出缺席異常之學生提早列入輔導。

關鍵詞：資料採礦、決策樹、學生流失

目 錄

論文口試合格證明.....	ii
誌謝.....	iii
中文摘要.....	iv
目 錄.....	v
表目錄.....	vi
圖目錄.....	vii
第一章、緒論.....	1
第一節 研究背景與動機.....	1
第二節 研究目的.....	2
第三節 研究範圍、限制.....	3
第四節 研究步驟.....	3
第五節 論文架構.....	4
第二章、文獻探討.....	6
第一節 顧客流失與學生流失定義.....	6
第二節 資料採礦定義.....	7
第三節 資料採礦的應用.....	9
第四節 決策樹.....	10
第五節 相關研究回顧.....	11
第三章、研究方法.....	12
第一節 研究對象.....	12
第二節 研究流程.....	12
第三節 資料處理.....	13
第四節 資料採礦工具.....	15
第四章、資料分析.....	17
第一節 資料特性分析.....	17
第二節 決策樹分析.....	22
第五章、結論與建議.....	40
第一節 研究結論.....	40
第二節 研究建議.....	41
參 考 文 獻.....	42
一、中文部份.....	42
二、西文部份.....	43

表 目 錄

表 1-1 各學年度國中學生數實際統計及預測值—按年級別分	1
表 1-2 高中職學生人數現況及推估一覽表	2
表 3-1 學生資料表.....	14
表 4-1 各科男、女生人數表	18
表 4-2 不同性別與全校人數比例表	19
表 4-3 各科不在學人數佔各科比例表	20
表 4-4 入學學年學期與不在學學年學期統計表	20
表 4-5 各學制不在學比例	22
表 4-6 決策樹分析之欄位	22
表 4-7 各模型分類矩陣比較表	29
表 4-8 各模型 3R比較表.....	31
表 4-9 dropout 4 決策樹第 1 子層不在學比例.....	34
表 4-10 dropout4 決策樹規則一不在學比例.....	34
表 4-11 dropout4 決策樹規則二不在學比例.....	35
表 4-12 dropout4 決策樹規則三不在學比例.....	36
表 4-13 dropout4 決策樹規則四不在學比例.....	37
表 4-14 97、98、99 學年度入學學生不在學原因	38

圖 目 錄

圖 3-1 研究流程圖.....	13
圖 4-1 全校男女生比例圖	17
圖 4-2 全校就學狀況圖	18
圖 4-3 不在學男、女生比例圖	19
圖 4-4 各學年學期不在學人數趨勢圖	21
圖 4-5 不在學原因比例圖	21
圖 4-6 dropout：參數為預設值之決策樹.....	24
圖 4-7 dropout1：COMPLEXITY_PENALTY=0.5 之決策樹.....	24
圖 4-8 dropout2：COMPLEXITY_PENALTY=0.1 之決策樹.....	25
圖 4-9 dropout3：COMPLEXITY_PENALTY=0.1、MINIMUM_SUPPORT=5 之決策樹.....	25
圖 4-10 dropout4：COMPLEXITY_PENALTY=0.5、MINIMUM_SUPPORT =5、SCORE_METHOD=1、SPLIT_METHOD=3 之決策樹.....	26
圖 4-11 dropout5：COMPLEXITY_PENALTY=0.1、MINIMUM_SUPPORT= 5、SCORE_METHOD=3、SPLIT_METHOD=3 之決策樹	27
圖 4-12 dropout6：COMPLEXITY_PENALTY=0.1、MINIMUM_SUPPORT= 5、SCORE_METHOD=1、SPLIT_METHOD=1 之決策樹	27
圖 4-13 dropout7：COMPLEXITY_PENALTY=0.1、MINIMUM_SUPPORT= 5、SCORE_METHOD=3、SPLIT_METHOD=1 之決策樹	27
圖 4-14 dropout8：COMPLEXITY_PENALTY=0.1、MINIMUM_SUPPORT= 5、SCORE_METHOD=4、SPLIT_METHOD=1 之決策樹	28
圖 4-15 dropout9：COMPLEXITY_PENALTY=0.1、SCORE_METHOD= 4、SPLIT_METHOD=1 之決策樹.....	28

圖 4-16 各採礦模型增益圖	32
圖 4-17 母體擴展 50%採礦圖例	33
圖 4-18 德行平均 <10 分及應屆 1 年級學生不在學原因比例圖	35
圖 4-19 德行平均 $\geq 10 < 40$ 分及 2 年級建教生不在學原因比例圖	36
圖 4-20 德行平均 ≥ 40 及 < 80 分及 2 年級學生不在學原因比例圖	37
圖 4-21 德行平均 > 80 分不在學原因比例圖	38

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，說明目前高中職學生人數之現況；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

依據教育部國民教育階段學生人數預測分析報告民國 100 年版中指出，國小新生數自 93 學年起產生大幅度負成長，國中新生數劇減現象則遞移至 99 學年，高中職則在 102 學年呈現。由表 1-1 中可看出 101 學年度國民中學畢業生由 31 萬多人降至 28 萬多人。

表 1-1 各學年度國中學生數實際統計及預測值—按年級別分

國民中學學生數					
學年度	總計	一年級	二年級	三年級	畢業生
95	952 344	317 762	319 638	314 944	314 010
96	953 277	317 072	317 298	318 907	317 975
97	951 976	318 239	316 818	316 919	316 080
98	948 534	313 942	318 052	316 540	315 798
99	919 802	288 229	313 815	317 758	316 630
100	872 009	271 085	287 871	313 053	311 946
101	842 642	284 724	270 750	287 168	286 158
102	828 396	273 926	284 373	270 097	269 151

(續下頁)

國民中學學生數					
學年度	總計	一年級	二年級	三年級	畢業生
103	798 749	241 460	273 597	283 692	282 684
104	742 765	228 651	241 169	272 945	271 978
105	682 055	213 084	228 376	240 595	239 736

資料來源：國民教育階段學生人數預測分析報告(民100)

依據教育部統計處資料，在表 1-2 中看出高中職學生人數推估 102 學年度較 101 學年度少約 3 萬人，103 學年度較 102 學年度更少約 4 萬人。面對少子化影響各高職學校招生情況也將日趨競爭。

表 1-2 高中職學生人數現況及推估一覽表

學年度	99		100	101	102	103	104	105
	實際學生數	占總學生數比率	推估學生數	推估學生數	推估學生數	推估學生數	推估學生數	推估學生數
總計	955,981	100.0	958,610	956,504	927,471	880,593	853,727	840,213
公立	461,293	48.3	463,233	461,782	448,076	425,532	412,120	405,613
私立	494,688	51.7	495,377	494,722	479,395	455,061	441,607	434,600

資料來源：教育部統計處（民100）

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，學校應針對在校學生學習與在校生活之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。

本研究擬運用資料採礦技術分析學生歷史學籍資料，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，提供相關之建議，以降低學生流失之情形。

第三節 研究範圍、限制

本研究之研究範圍以個案學校 97 學年度至 99 學年度入學日間部學生為主（不含綜合職能科學生），因分析來源為單一學校，結果恐難類推至所有高職學校，但仍可作為類似各校減少學生流失之參考；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；再者校方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，學校不提供書面或電子檔案，僅在拜訪校方取得資料時，由該校人員回答大致狀況。

第四節 研究步驟

本研究將以私立個案學校 97-99 學年度日間部入學新生之資料，研究探討影響個案學校日間部學生流失的相關因素，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 97-99 學年度入學日間部學生之資料，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 結論與建議

針對第四章所得結果進行整理，並提出相關結論及建議。

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的關係，是企業創造利潤的重要關鍵，如何預防客戶流失將十分重要。

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 Useama 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：

- 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(2009)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程；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

個案學校學生規模於民國 86 年達到為 3 千多人，100 學年度人數為 1 千 5 百多人。本研究以 97 學年度至 99 學年度入學日間部學生為主研究對象，科別為資料處理科、美容科、資訊科、餐飲科、汽車科等五科，由於研究樣本選取時間為 99 學年度第 2 學期，故其就學狀態為 99 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、轉出、死亡等，轉出分為轉出與轉出(改變學習環境)，高中職校因改為學年學分制，已無退學這項原因而以改變學習環境取代。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，擬探求學生流失之潛在影響因素，希望藉由發現學生流失之潛在共同因素，提供個案學校作參考，以減少學生流失之情形。

第二節 研究流程

本研究的主要研究變數為學制、性別、科系、縣市別、入學資格、入學年度、入學區分、學業平均、德行平均，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，找出學生流失的共同因素。本研究之研究流程如圖 3-1 所示，為研究動機、探討相關文獻、分析資料特性、決策樹分析、檢視原始資料、結論與建議。

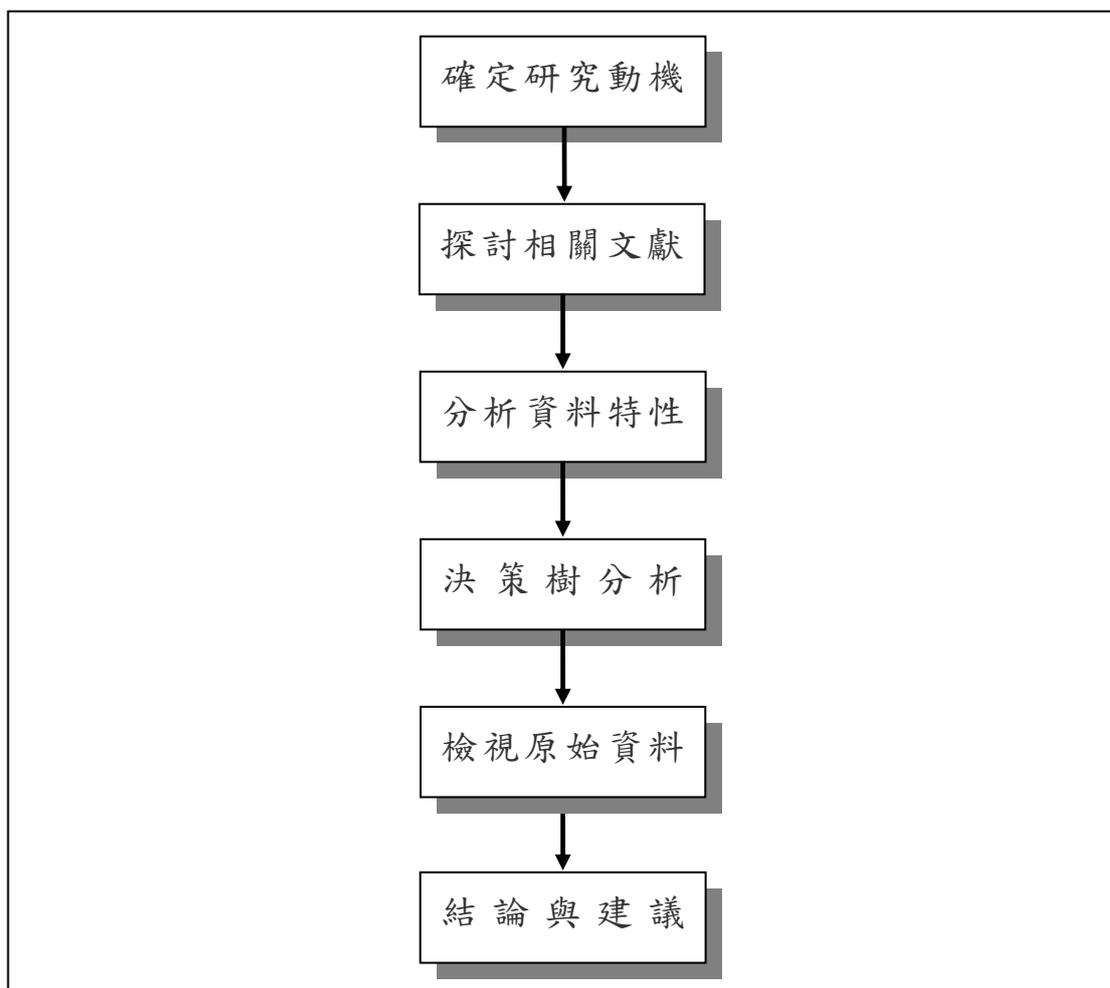


圖 3-1 研究流程圖

資料來源：本研究整理

第三節 資料處理

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、轉出、死亡之學生，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。

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，如入學國中轉為縣市別。另外將每學期學業成績及操性成績計算出學業平均及德行平均。

三、資料整合：

將教務處的學業成績與學務處的德行成績進行整合，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 所示

表 3-1 學生資料表

欄名	屬性	長度	說明
學號	文字	8	
學制	文字	8	日校、建教班、實用技能班
入學年度	文字	2	
性別	文字	1	男、女
科系	文字	12	
縣市別	文字	3	

(續下頁)

欄名	屬性	長度	說明
入學資格	文字	20	持國民中學修(結)業證明書 持國民中學畢業證書 公告招生降級轉科 轉學 公告招生(轉科)
入學區分	文字	6	應屆、非應屆
就學狀況	文字	3	在學、不在學
學業平均	數字	浮點	
德行平均	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 2008 R2 為資料採礦工具。SQL Server 2008 R2 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 2008 R2 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 2008 R2 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 2008 R2 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 2008 R2 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

本章共有四節，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 99 學年度第二學期日間部之全校人數(不含綜合職能科)，共取得 1465 筆其中女生為 616 人、男生為 849 人。全校男、女生比例如圖 4-1 所示。

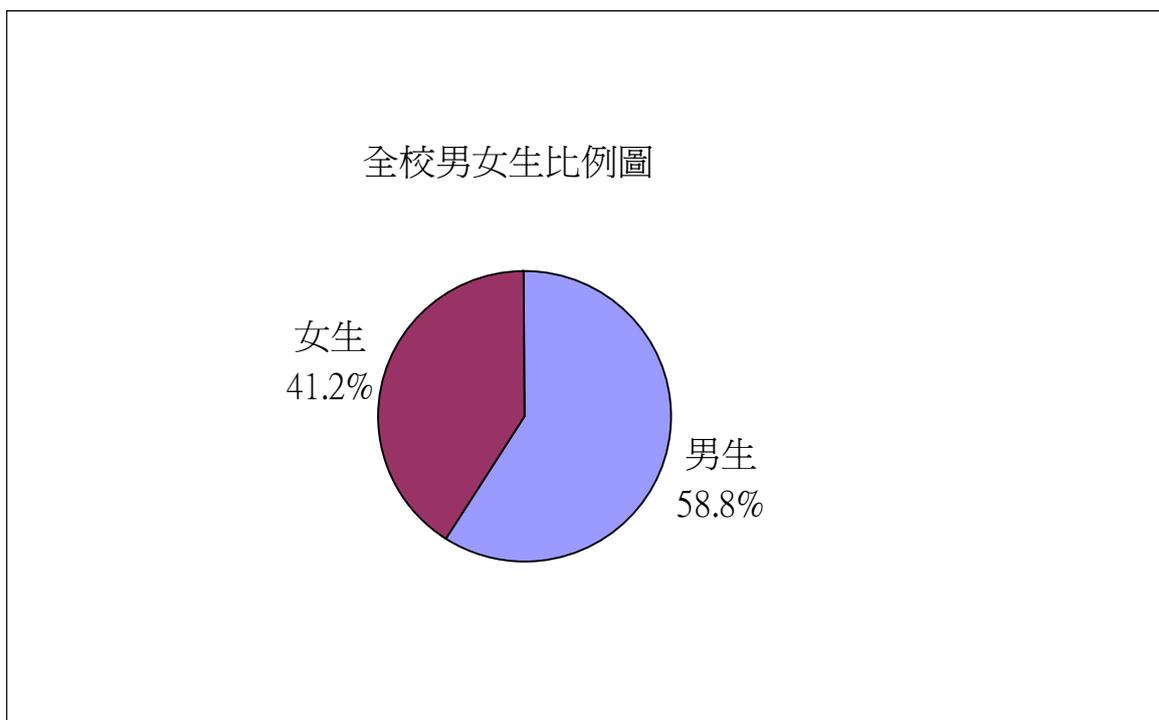


圖 4-1 全校男女生比例圖

資料來源：本研究整理

由於職業學校各科雖沒有性別限制，但由於工科與商科男女生就

讀比例本就有不同，如表 4-1 各科男、女生人數所示，故本研究不比較各科男、女生的流失比例，而是與全校男女生做比較。

表 4-1 各科男、女生人數表

科系	女	男	合計
汽車科	0	109	109
汽車修護科	3	48	51
美容科	69	22	91
美髮技術科	119	15	134
美顏科	47	2	49
資料處理科	120	146	266
資訊科	12	97	109
微電腦修護科	10	130	138
餐飲技術科	75	84	161
餐飲管理科	161	196	357
總計	616	740	1465

資料來源：本研究整理

貳、就學狀態分析：

本研究資料中就學狀況不在學有 218 人，在學有 1247 人，由圖 4-1 全校就學狀況表可看出不在學比例為 14.9%。

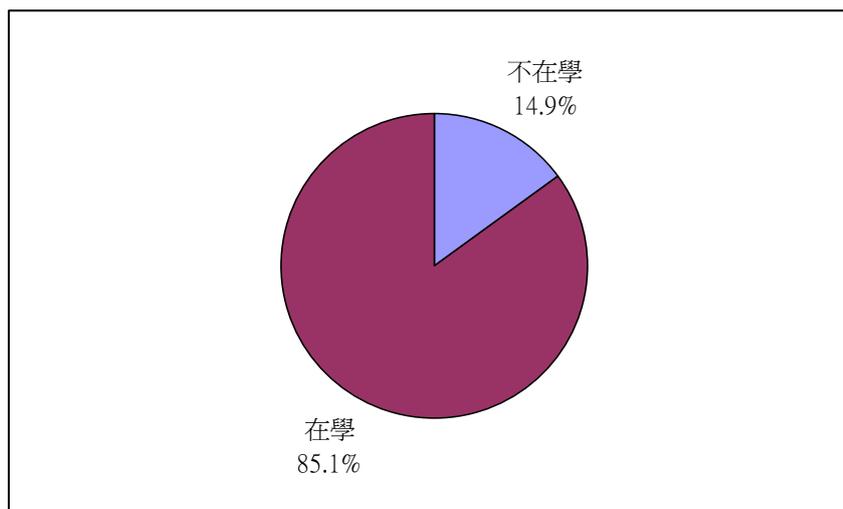


圖 4-2 全校就學狀況圖

資料來源：本研究整理

參、性別資料分析

由圖 4-3 不在學男女生比例圖中，可看出不在學的人數中男生不在學比例高過女生 28.4%。

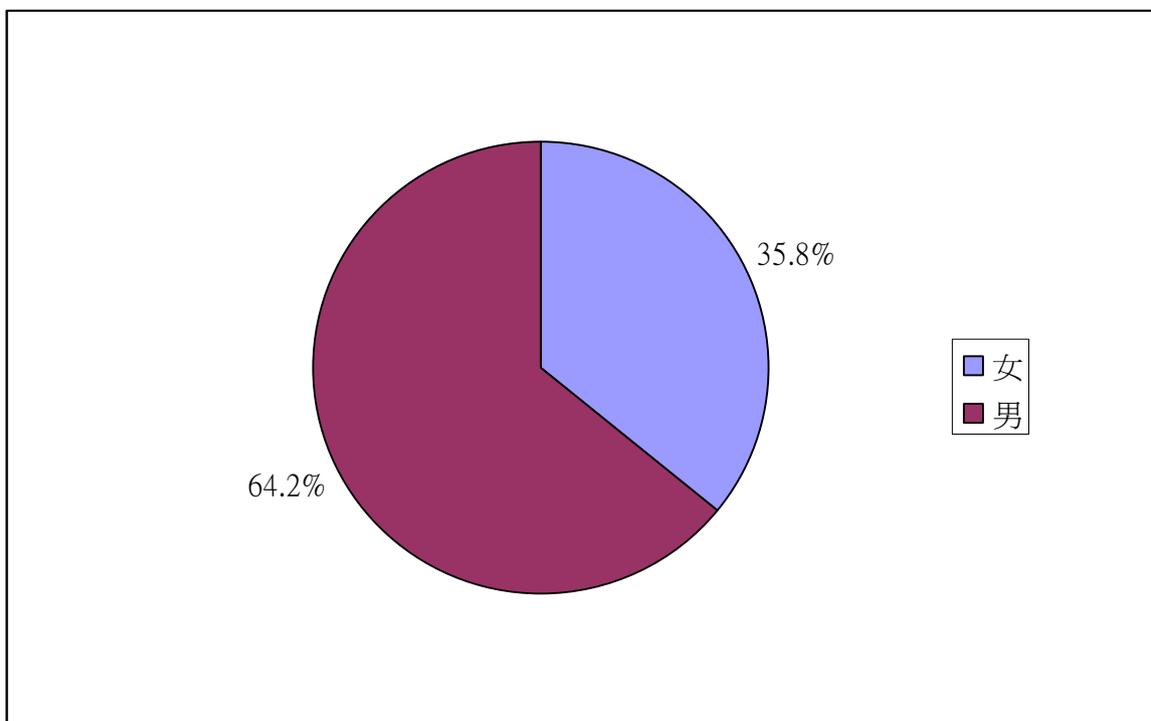


圖 4-3 不在學男、女生比例圖

資料來源：本研究整理

若與全校男、女生人數比較之後，不在學男生佔全校男生比例為 16.5%，而不在學女生佔全校女生比例為 12.7%，如表 4-2 所示。

表 4-2 不同性別與全校人數比例表

性別	不在學	全校	比例
女	78	616	12.7%
男	140	849	16.5%

資料來源：本研究整理

肆、各科資料分析

由表 4-3 各科不在學人數佔各科比例表中可看出，不在學比例最

高的是美容科的 25.3%，最低的是餐飲技術科。

表 4-3 各科不在學人數佔各科比例表

科系	人數	不在學	比例
汽車科	109	20	18.3%
汽車修護科	51	8	15.7%
美容科	91	23	25.3%
美髮技術科	134	27	20.1%
美顏科	49	3	6.1%
資料處理科	266	40	15.0%
資訊科	109	21	19.3%
微電腦修護科	138	14	10.1%
餐飲技術科	161	2	1.2%
餐飲管理科	357	60	16.8%

資料來源：本研究整理

伍、不在學資料分析

本研究資料因 100 學年度不在學資料只有第 1 學期，沒有第 2 學期資料，為顧及資料完整性故 100 學年度第 1 學期資料在分析時予以刪除。不在學資料是以學校每年 10 月底與 2 月底報到教育部中部辦公室之異動名冊為主。

在表 4-4 入學學年學期與不在學學年學期統計表中發現，入學的第一學年不在學率最高，第 2 年降低、第 3 年最低。

表 4-4 入學學年學期與不在學學年學期統計表

入學學年學期	不在學學年學期					
	971	972	981	982	991	992
97 學年第 1 學期	62	45	52	21	13	4
98 學年第 1 學期	-	-	48	49	42	21
99 學年第 1 學期	-	-	-	-	67	71

資料來源：本研究整理

由圖 4-4 可看出在每學年度的第 1 學期不在學人數相對較多，第

2 學期比較少。

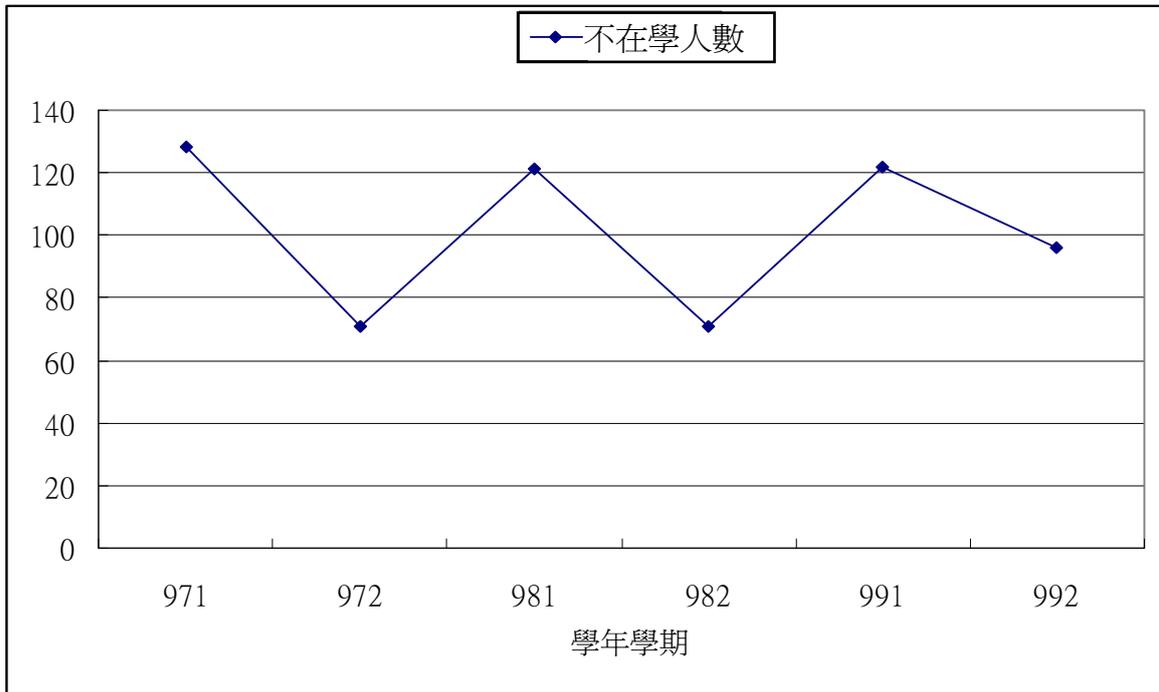


圖 4-4 各學年學期不在學人數趨勢圖

資料來源：本研究整理

由圖 4-5 不在學原因比例圖中可看出，以轉出（轉換環境）比例 52.3% 為最高，轉出（其他）18.3% 次之。

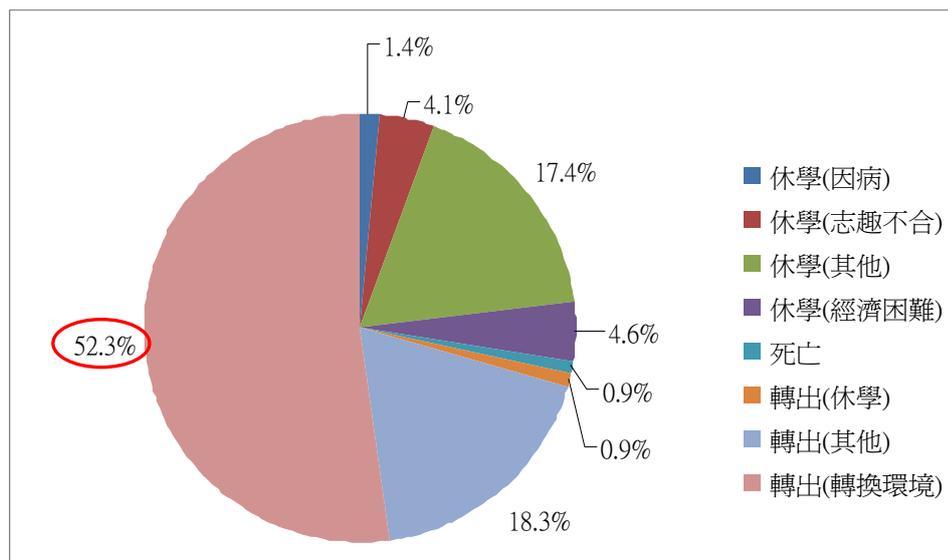


圖 4-5 不在學原因比例圖

資料來源：本研究整理

陸、各學制資料分析

表 4-5 各學制不在學比例中顯示，建教班同學不在學比例較其它學制還要高出將近 15%。

表 4-5 各學制不在學比例

學制	人數	不在學	比例
日校	548	63	11.5%
建教班	384	101	26.3%
實用技能班	533	54	10.1%

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，增加在學狀況欄分為在學與不在學，將異動名冊中轉出、休學列為不在學資料，共取得 218 筆，與在學資料比例約為 1：5.84。整理資料後將異動原因、異動學年度等不適合的欄位刪除，各學期學業成績和各學期德行成績加以整理成為學業平均與德行平均。經處理後所要分析的資料欄位說明如表 4-6：

表 4-6 決策樹分析之欄位

屬性	欄名	說明
Key	學號	
Input	學制	日校、建教班、實用技能班
Input	入學年度	
Input	性別	男、女
Input	科系	
Input	縣市別	
Input	入學資格	持國民中學修(結)業證明書、持國民中學畢業證書、公告招生降級轉科、轉學、公告招生(轉科)

(續下頁)

屬性	欄名	說明
Input	入學區分	應屆、非應屆
PredictOnly	就學狀況	在學、不在學
Input	學業平均	
Input	德行平均	

資料來源：本研究整理

貳、資料採礦作業：

資料採礦中面臨到的抽樣問題「稀有事件」，本研究將「不在學」設定為稀有事件，雖然不在學資料佔了全部資料的 14.9% 不是真的稀有，但與在學比例 85.1% 比較起來仍算「稀有事件」。

本次資料分為訓練組和測試組，訓練組佔總資料的 30% 即為 438 筆資料，其中不在學為 63 筆、在學為 375 筆，不在學比例約 14.4%，與全校不在學比例 14.9% 差不到 1%，與全校不在學比例相符。

接著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、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，再改變 COMPLEXITY_PENALTY 分別為 0.5、0.1 建立 dropout1、dropout2 模型，發現其分類矩陣沒改變，於是將 COMPLEXITY_PENALTY 維持在 0.1，再將

MINIMUM_SUPPORT 設為 5 降低最小案例數後，發現分類矩陣有改變。

(二)、將 COMPLEXITY_PENALTY 維持在 0.1、MINIMUM_SUPPORT 維持在 5 後，將 SORE_METHOD、SPLIT_METNOD 做不同的組合，產生不同的決策樹模型。圖 4-6 到圖 4-15 為各模型之樹狀圖。

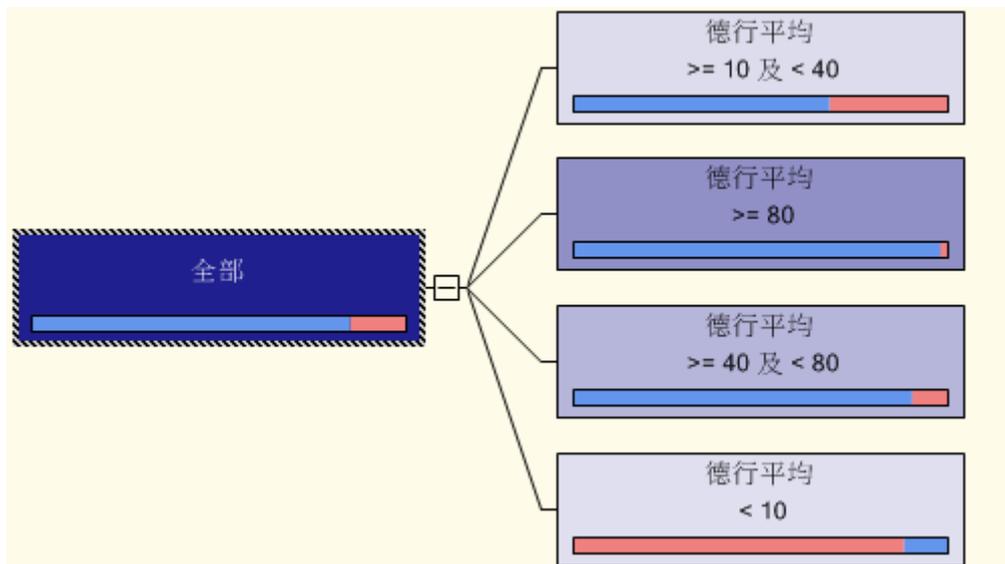


圖 4-6 dropout：參數為預設值之決策樹
資料來源：本研究整理

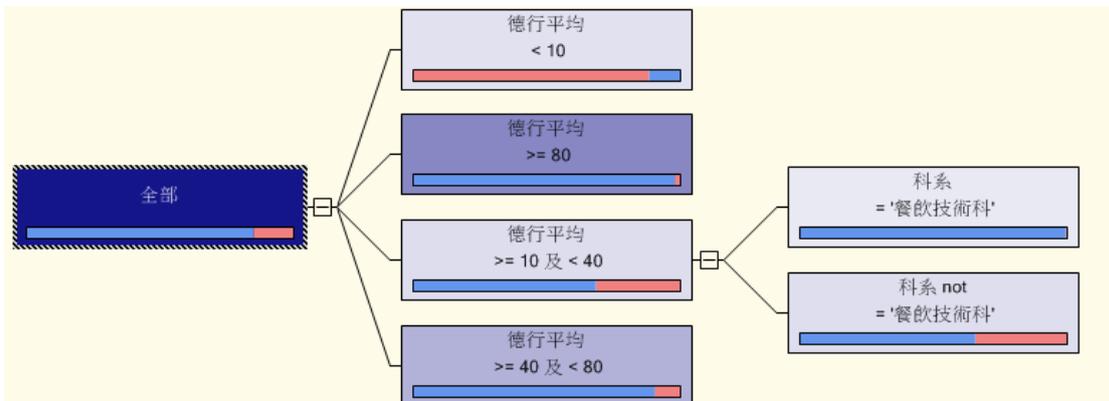


圖 4-7 dropout1：COMPLEXITY_PENALTY=0.5 之決策樹
資料來源：本研究整理

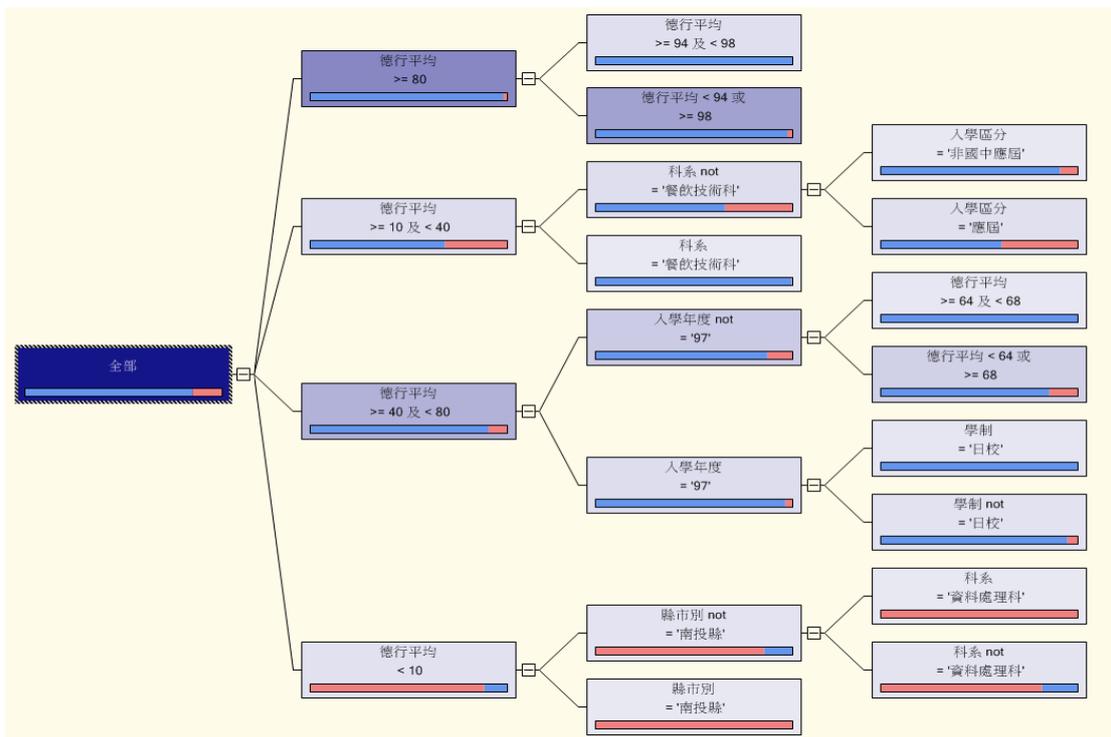


圖 4-8 dropout2 : COMPLEXITY_PENALTY=0.1 之決策樹

資料來源：本研究整理

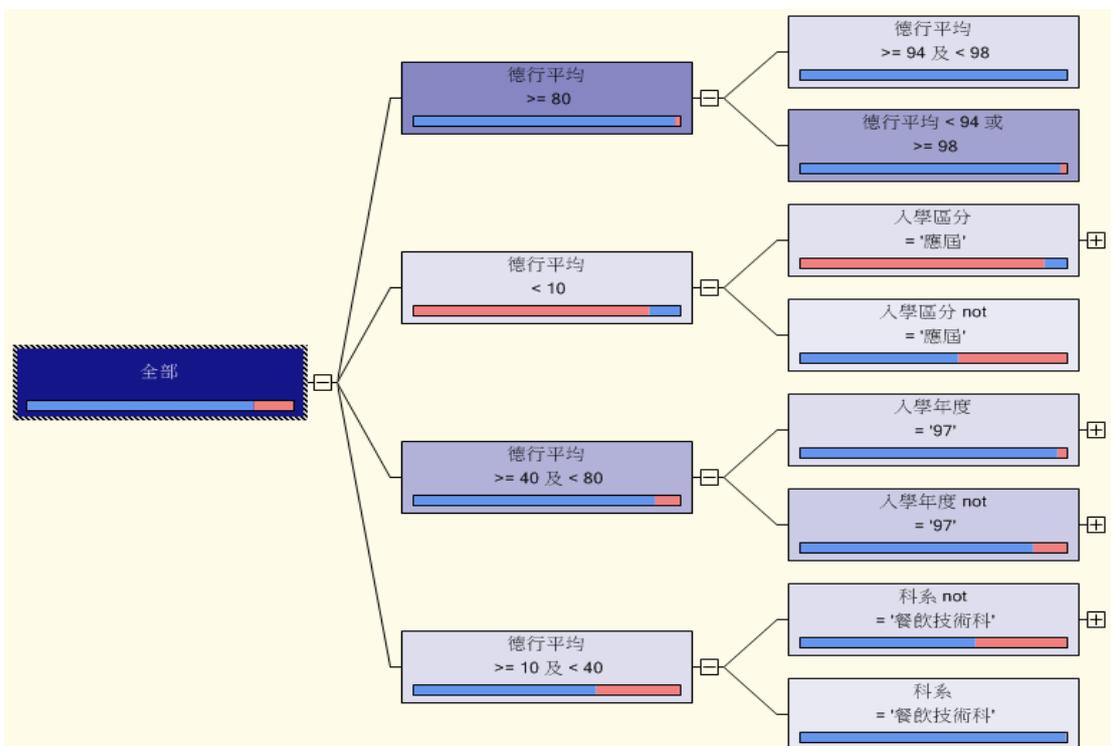


圖 4-9 dropout3 : COMPLEXITY_PENALTY=0.1、MINIMUM_SUPPORT=5 之決策樹

資料來源：本研究整理

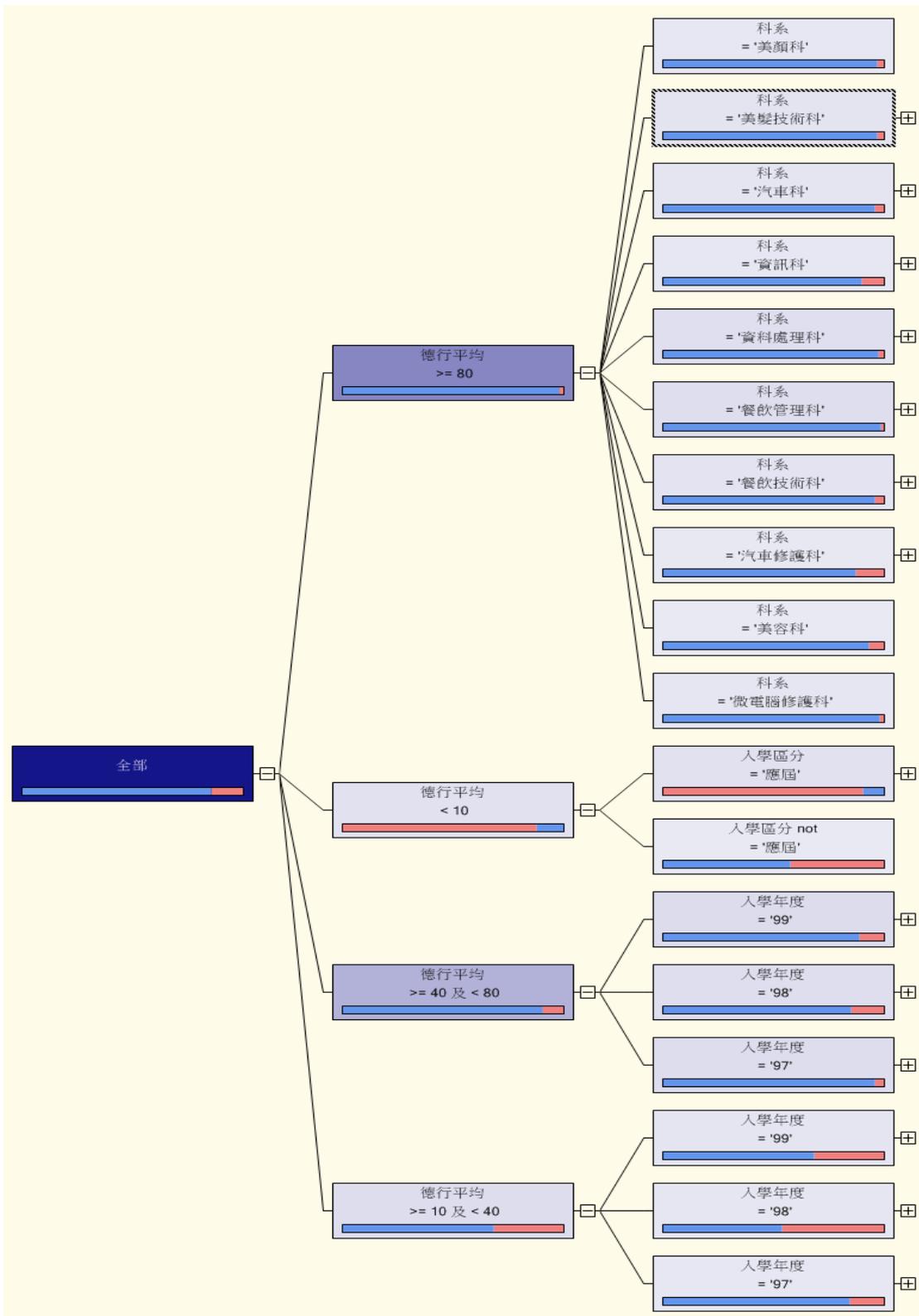


圖 4-10 dropout4 : COMPLEXITY_PENALTY=0.5、MINIMUM_SUPPORT=5、SCORE_METHOD=1、SPLIT_METHOD=3 之決策樹

資料來源：本研究整理

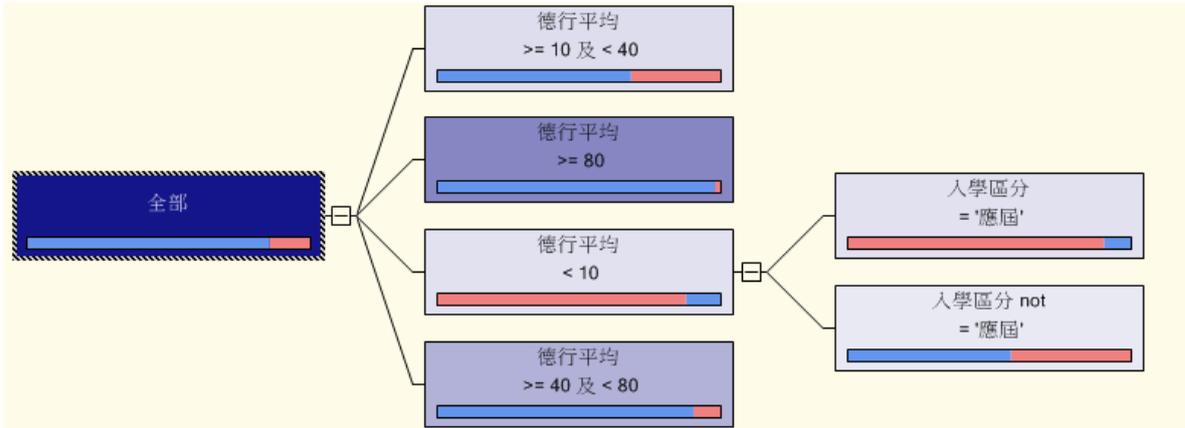


圖 4-11 dropout5: COMPLEXITY_PENALTY=0.1、MINIMUM_SUPPORT=5、SCORE_METHOD=3、SPLIT_METHOD=3 之決策樹

資料來源：本研究整理

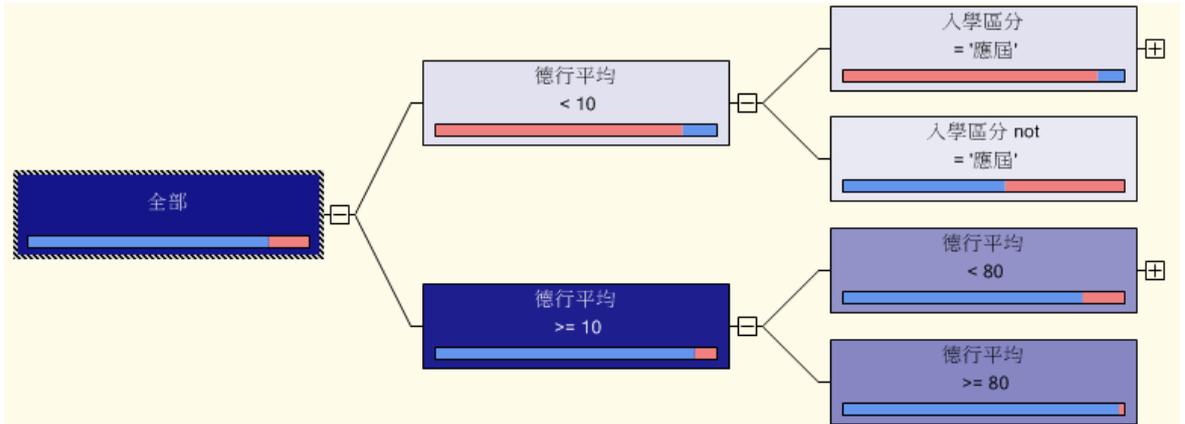


圖 4-12 dropout6: COMPLEXITY_PENALTY=0.1、MINIMUM_SUPPORT=5、SCORE_METHOD=1、SPLIT_METHOD=1 之決策樹

資料來源：本研究整理

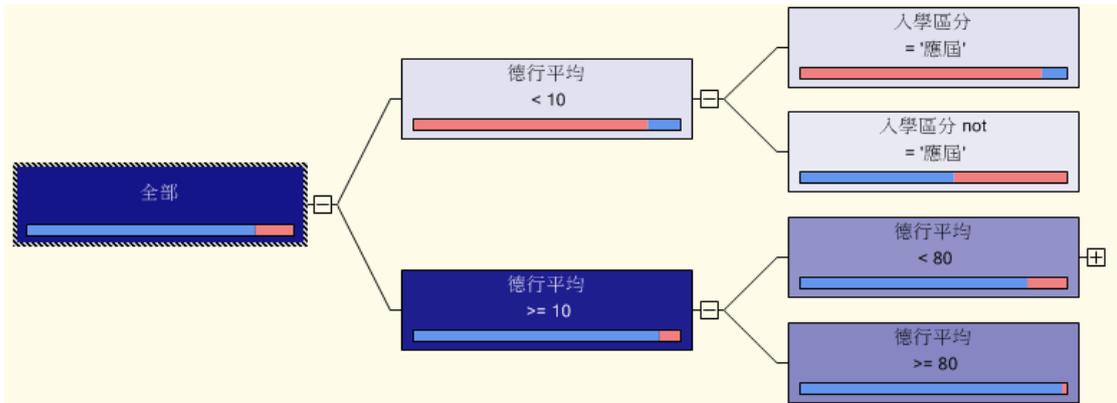


圖 4-13 dropout7: COMPLEXITY_PENALTY=0.1、MINIMUM_SUPPORT=5、SCORE_METHOD=3、SPLIT_METHOD=1 之決策樹

資料來源：本研究整理

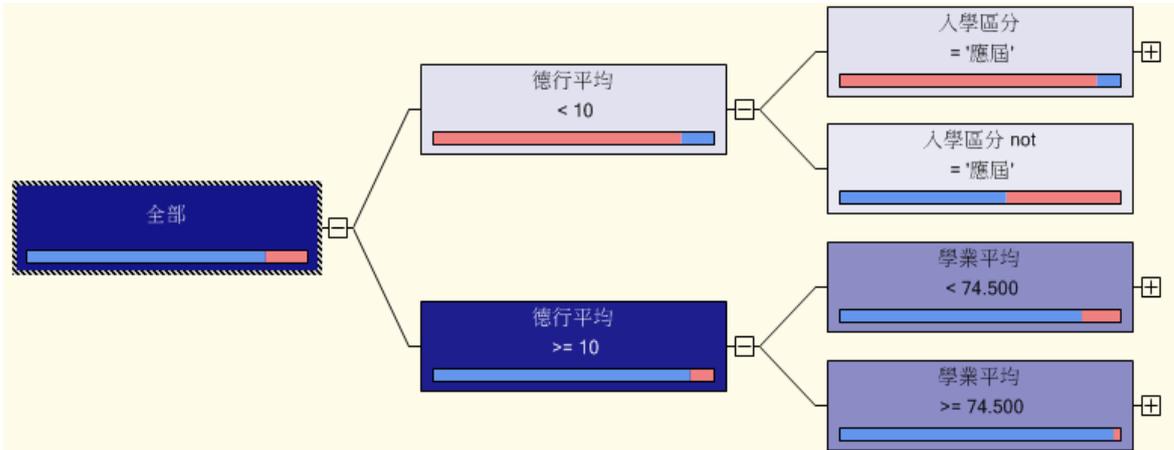


圖 4-14 dropout8：COMPLEXITY_PENALTY=0.1、MINIMUM_SUPPORT=5、SCORE_METHOD=4、SPLIT_METHOD=1 之決策樹

資料來源：本研究整理

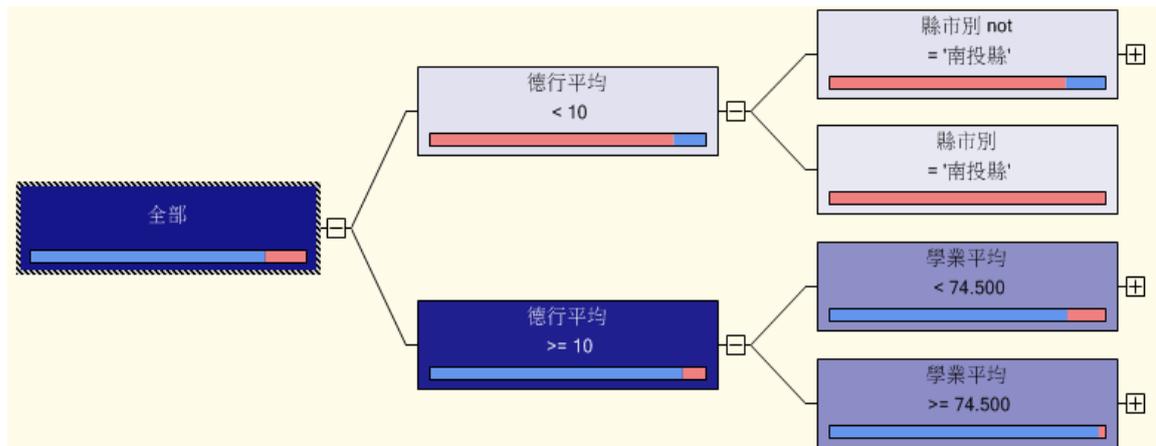


圖 4-15 dropout9：COMPLEXITY_PENALTY=0.1、SCORE_METHOD=4、SPLIT_METHOD=1 之決策樹

資料來源：本研究整理

四、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，雖然 dropout3 預測的數值不同但可能因資料只有 438 筆，正確與錯誤總數百分比和 dropout5、dropout6、dropout8 一樣。

表 4-7 各模型分類矩陣比較表

dropout 決策樹模型 演算法參數為預設值	預測的	不在學 (實際)	在學 (實際)
	不在學	33	3
	在學	29	373
	正確總數、百分比	406	92.69%
	錯誤總數、百分比	32	7.31%
dropout1 決策樹模型 COMPLEXITY_PENALTY=0.5	預測的	不在學 (實際)	在學 (實際)
	不在學	33	3
	在學	29	373
	正確總數、百分比	406	92.69%
	錯誤總數、百分比	32	7.31%
dropout2 決策樹模型 COMPLEXITY_PENALTY=0.1	預測的	不在學 (實際)	在學 (實際)
	不在學	33	3
	在學	29	373
	正確總數、百分比	406	92.69%
	錯誤總數、百分比	32	7.31%
dropout3 決策樹模型 COMPLEXITY_PENALTY=0.1 MINIMUM_SUPPORT=5	預測的	不在學 (實際)	在學 (實際)
	不在學	33	2
	在學	29	374
	正確總數、百分比	407	92.92%
	錯誤總數、百分比	31	7.08%
dropout4 決策樹模型 COMPLEXITY_PENALTY=0.5 MINIMUM_SUPPORT=5 SCORE_METHOD=1 SPLIT_METHOD=3	預測的	不在學 (實際)	在學 (實際)
	不在學	44	18
	在學	18	358
	正確總數、百分比	402	91.78%
	錯誤總數、百分比	36	8.22%

(續下頁)

dropout5 決策樹模型 COMPLEXITY_PENALTY=0.1 MINIMUM_SUPPORT=5 SCORE_METHOD=3 SPLIT_METHOD=3	預測的	不在學 (實際)	在學 (實際)
	不在學	33	2
	在學	29	374
	正確總數、百分比	407	92.92%
	錯誤總數、百分比	31	7.08%
dropout6 決策樹模型 COMPLEXITY_PENALTY=0.1 MINIMUM_SUPPORT=5 SCORE_METHOD=1 SPLIT_METHOD=1	預測的	不在學 (實際)	在學 (實際)
	不在學	36	5
	在學	26	371
	正確總數、百分比	407	92.92%
	錯誤總數、百分比	31	7.08%
dropout7 決策樹模型 COMPLEXITY_PENALTY=0.1 MINIMUM_SUPPORT=5 SCORE_METHOD=3 SPLIT_METHOD=1	預測的	不在學 (實際)	在學 (實際)
	不在學	41	18
	在學	21	358
	正確總數、百分比	399	91.10%
	錯誤總數、百分比	39	8.90%
dropout8 決策樹模型 COMPLEXITY_PENALTY=0.1 MINIMUM_SUPPORT=5 SCORE_METHOD=4 SPLIT_METHOD=1	預測的	不在學 (實際)	在學 (實際)
	不在學	33	2
	在學	29	374
	正確總數、百分比	407	92.92%
	錯誤總數、百分比	31	7.08%
dropout9 決策樹模型 COMPLEXITY_PENALTY=0.1 SCORE_METHOD=4 SPLIT_METHOD=1	預測的	不在學 (實際)	在學 (實際)
	不在學	33	3
	在學	29	373
	正確總數、百分比	406	92.69%
	錯誤總數、百分比	32	7.31%

資料來源：本研究整理

五、計算各模型 3R

計算各模型的 3R 來比較，但由表 4-8 各模型 3R 比較表中難以看出那一個模型比較好，故使用增益圖來比較。

表 4-8 各模型 3R 比較表

模型名稱	總體回應率	回應率 越高越好	回應率提升 越高越好	反查 越高越好	間距縮減 越低越好
dropout	14.16%	91.67%	647.58%	53.23%	8.22%
dropout1	14.16%	91.67%	647.58%	53.23%	8.22%
dropout2	14.16%	91.67%	647.58%	53.23%	8.22%
dropout3	14.16%	94.29%	666.08%	53.23%	7.99%
dropout4	14.16%	70.97%	501.35%	70.97%	14.16%
dropout5	14.16%	94.29%	666.08%	53.23%	7.99%
dropout6	14.16%	87.80%	620.30%	58.06%	9.36%
dropout7	14.16%	69.49%	490.92%	66.13%	13.47%
dropout8	14.16%	94.29%	666.08%	53.23%	7.99%
dropout9	14.16%	91.67%	647.58%	53.23%	8.22%

資料來源：本研究整理

六、建立各模型比較增益圖：

增益圖的縱軸百分比代是在資料中稀有事件的人數佔總體稀有事件人數的百分比。橫軸百分比代表資料採礦模型根據機率從高至低排序後的名單佔總體百分比。

增益圖中有一條隨機猜測模型的 45 度斜線，以增益圖的定義來說，代表當我們篩選一半的名單去檢視學生流失狀況時，就會剛好包含全體名單一半的學生流失數量。

正常的模型增益圖必定要比 45 度線向第二象限彎曲，且增益圖愈向上彎曲，表示模型效果愈好。「理想模型」的線代表完美預測的結果。所有模型的曲線必須介於「隨機猜測」以及「理

想模型」之間，愈接近「理想模型」愈好。(尹相志，2007)

縱軸表示目標母體百分比，這個值愈高就表示模型的預測力愈高。橫軸表示整體母體擴展，即 overall population%，如圖 4-16 各採礦模型增益圖中紅色的線 (dropout 9 決策樹模型) 對應在橫軸 50% 對應到縱軸的值為 91.94%，這代表資料採礦根據流失機率由最高至最低排序的前 50% 名單中抓到會流失的學生，就佔了總體會流失的學生的 91.94%，相對的，剩下的 50% 學生中，只佔了總體會流失學生 8.06%。在此，我們以母體擴展 50% 為基準作各模型之比較。

由圖 4-16 各採礦模型增益圖看來，各模型曲線除 dropout 1、dropout 5 和 dropout 9 明顯較低外，其餘曲線多少都有重疊，難以看出那一條曲線比較理想，所以進一步比較圖 4-17 母體擴展 50% 採礦圖例。

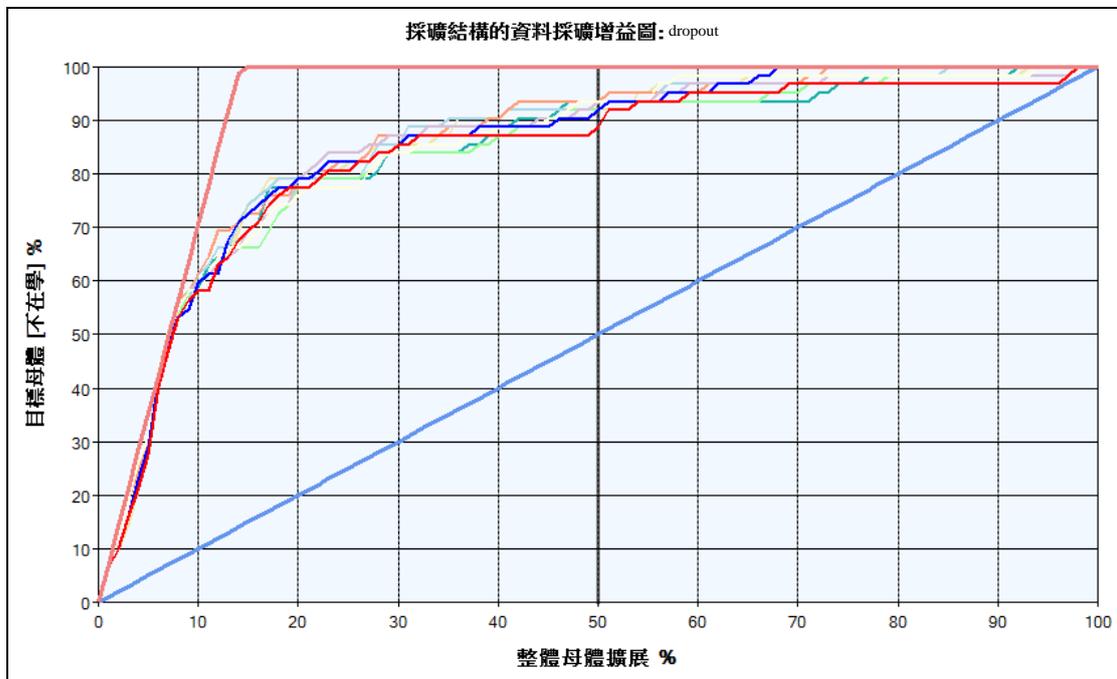


圖 4-16 各採礦模型增益圖

資料來源：本研究整理

由於「目標母體」值則僅就部分母體擴展而言，而「分數」指的是整體的面積，故以「分數」高低做為模型優劣的衡量，分數越高越好。所以依據圖 4-17 母體擴展 50%採礦圖例中顯示，以 dropout 4 的 0.93 分最高。

序列、模型	分數	目標母體	預測機率
dropout1	0.91	91.94%	9.93%
dropout	0.90	90.32%	9.93%
dropout2	0.92	91.94%	2.95%
drpooout3	0.92	91.94%	2.95%
drpooout4	0.93	93.55%	7.69%
dropout5	0.90	90.32%	9.79%
dropout6	0.92	90.32%	2.86%
dropout7	0.91	90.32%	7.39%
dropout8	0.92	90.32%	9.25%
dropout9	0.90	87.10%	9.25%
隨機猜測模型		46.00%	
dropout的理想模...		100.00%	

圖 4-17 母體擴展 50%採礦圖例
資料來源：本研究整理

七、採礦結果與圖表

本研究將 99 學年度入學學生定義為 1 年級生、98 學年度入學為 2 年級生、97 學年度為 3 年級生。依據各模型的樹狀圖來看(圖 4-6~圖 4-15)，發現無論各樹狀圖分層情形如何，第 1 子層皆為德行平均，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，以德行平均為最重要的因素。

依據圖 4-17 母體擴展 50%採礦圖例中顯示，以 dropout 4 的 0.93 分最高表現最好，故以下將依據 dropout 4 決策樹來分析。

在 dropout 4 決策樹中第 1 子層分為德行平均 < 10 分、> = 10 分及 < 40 分、> = 40 分及 < 80 分及 > = 80 分四個規則，如表 4-9 所示。以下依據四個規則中不在學比例高低依序說明與探討原因。

表 4-9 dropout 4 決策樹第 1 子層不在學比例

dropout4 第 1 子層規則	不在學人數	佔不在學比例
規則一：德行平均 < 10	74	48.37%
規則二：德行平均 > = 10 及 < 40	36	23.53%
規則三：德行平均 > = 40 及 < 80	32	20.92%
規則四：德行平均 > = 80	11	7.19%
總計	153	100.00%

資料來源：本研究整理

(一)、規則一：德行平均在 < 10 分以下不在學比例為 48.37%

由表 4-10 中可看出，第 2 子層德行平均 < 10 分不在學資料中又以應屆入學比例為 94.04%；第 3 子層德行平均 < 10 分及應屆入學不在學資料中，又以 1 年級生比例高達 87.5%。

表 4-10 dropout4 決策樹規則一不在學比例

dropout4 決策樹		不在學人數	比例
父層		153	
第 1 子層	德行平均 < 10	74	48.37%
第 2 子層	應屆	72	97.30%
第 3 子層	1 年級生	63	87.50%

資料來源：本研究整理

回頭檢視原始資料，德行平均 < 10 分、1 年級生不在學最大的原因是轉出（轉換環境），比例為 72.63%。其原

因為德行分數太低而被轉換環境，依據校方表示原因有送建教後未返校、不適應學校環境缺曠課過多、獎懲過多等因素。如圖 4-18 德行平均 <10 分及應屆 1 年級生不在學原因比例圖所示。

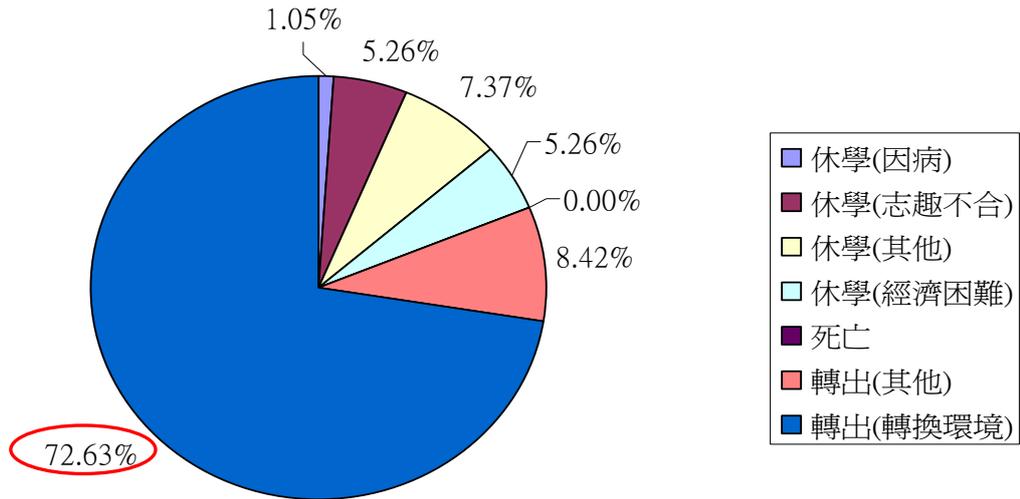


圖 4-18 德行平均 <10 分及應屆 1 年級學生不在學原因比例圖
資料來源：本研究整理

(二)、規則二：德行平均 ≥ 10 分及 < 40 分佔不在學比例 23.53%

在表 4-11 中，第 2 子層在德行平均 ≥ 10 分及 < 40 分的不在學資料中，以 2 年級生佔 50%；第 3 子層德行平均 ≥ 10 分及 < 40 分且 2 年級生不在學資料中，學制是建教班的比例為 77.78%。

表 4-11 dropout4 決策樹規則二不在學比例

dropout4 決策樹		不在學人數	比例
父層		153	
第 1 子層	德行平均 ≥ 10 及 < 40 分	36	23.53%
第 2 子層	2 年級生	18	50.00%
第 3 子層	建教	14	77.78%

資料來源：本研究整理

檢視原始資料發現不在學資料中，影響德行平均 ≥ 10 分及 < 40 分2年級建教班學生不在學的原因是休學(其他)占42.11%如圖4-19所示。

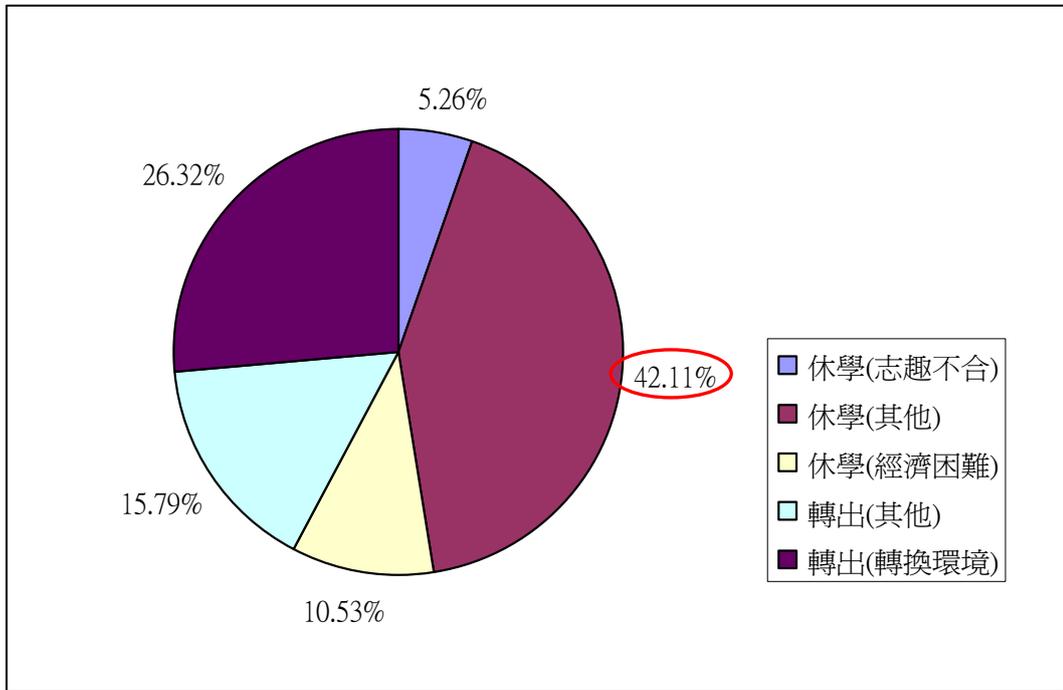


圖 4-19 德行平均 $\geq 10 < 40$ 分及2年級建教生不在學原因比例圖
資料來源：本研究整理

(三)、規則三：德行平均 ≥ 40 分及 < 80 分佔不在學比例20.92%

由表4-12中第2子層為2年級生不在學比例為65.63%，第3子層2年級生德行平均 ≥ 40 分及 < 80 分中，以德行平均 < 64 分不在學佔66.67%。

表 4-12 dropout4 決策樹規則三不在學比例

dropout4 決策樹		不在學人數	比例
父層		153	
第1子層	德行平均 ≥ 40 及 < 80 分	32	20.92%
第2子層	2年級生	21	65.63%
第3子層	德行 < 64 分	14	66.67%

資料來源：本研究整理

檢視原始資料，影響德行平均 ≥ 40 及 < 80 分且2年級生不在學比例最高的是因素為轉出(其他)，佔42.86%。如圖4-20德行平均 ≥ 40 及 < 80 分及2年級學生不在學原因比例圖所示。

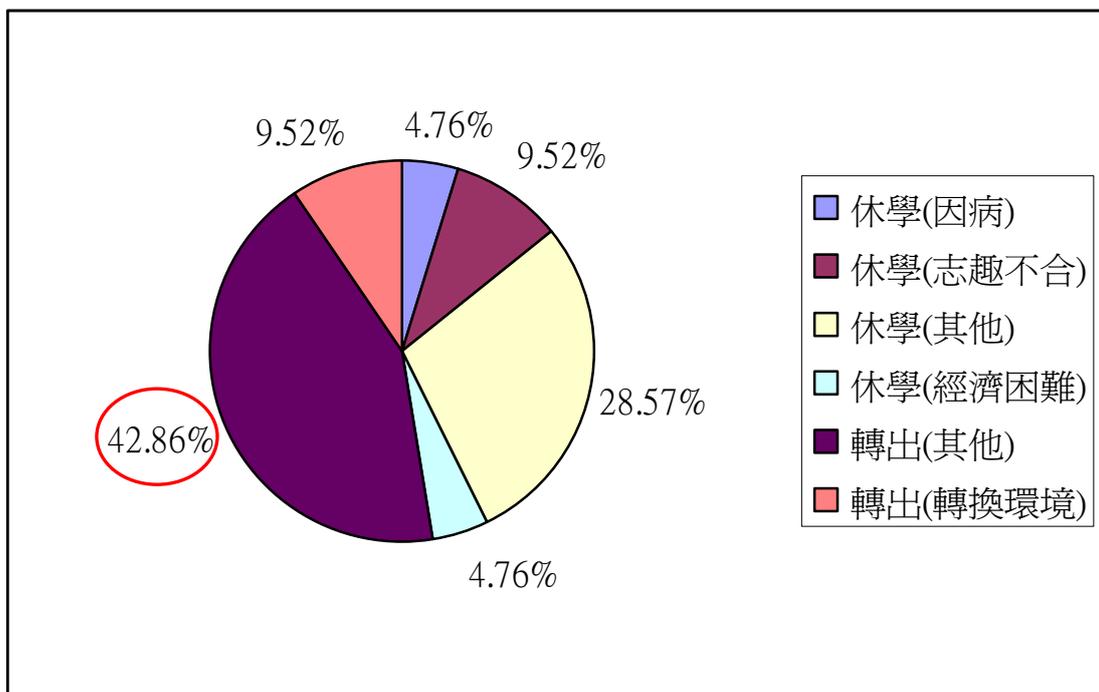


圖 4-20 德行平均 ≥ 40 及 < 80 分及2年級學生不在學原因比例圖
資料來源：本研究整理

(四)、德行平均 > 80 分不在學的比例為7.19%

由表4-13中看出，德行平均 > 80 分第2子層以科系為區分，因資料筆數較少，各科最多流失不到5人。影響不在學因素為休學(其他)佔40.00%，如圖4-21所示。

表 4-13 dropout4 決策樹規則四不在學比例

dropout4 決策樹		不在學人數	比例
父層		153	
第1子層	德行平均 < 80 分	11	7.19%
第2子層	各科系最多2人	2	18.18%

資料來源：本研究整理

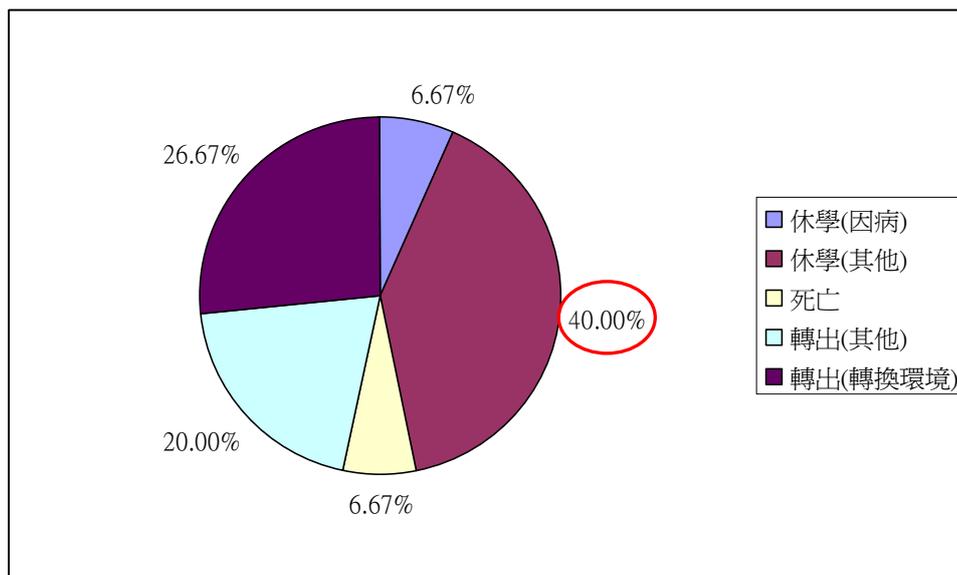


圖 4-21 德行平均>80 分不在學原因比例圖

資料來源：本研究整理

綜合 dropout 4 採礦結果與原始資料，1 年級學生德行平均>10 分不在學比例最高，其不在學主因為轉出(轉換環境)； ≥ 10 分及 < 40 分，其不在學主因為休學(其他)；德行平均 $\geq 40 < 80$ 分及 2 年級學生不在學主因為轉出(其他)；影響德行平均>80 分不在學主因為休學(其他)。

在 dropout4 決策樹中影響學生不在學原因除德行成績為主要因素外，還有入學年度。檢視原始資料 99 年入學為 1 年級生以轉出(轉換環境)人數最多；98 入學為 2 年級生轉出(轉換環境)人數較 1 年級為低；97 年入學為 3 年級生不在學人數最低，如表 4-14 所示。

表 4-14 97、98、99 學年度入學學生不在學原因

不在學原因	97 年入學	98 年入學	99 年入學	總計
休學(因病)	0	1	2	3
休學(志趣不合)	0	3	6	9
休學(其他)	7	17	14	38

(續下頁)

不在學原因	97年入學	98年入學	99年入學	總計
休學(經濟困難)	1	3	6	10
死亡	1	0	1	2
轉出(休學)	0	2	0	2
轉出(其他)	0	18	22	40
轉出(轉換環境)	8	18	88	114
總計	17	62	139	218

資料來源：本研究整理

第五章、結論與建議

第一節 研究結論

為了減少學生的流失率，本研究試圖以不在學的學生資料，找出影響學生不在學的潛在原因，以提供給學校老師找出需要重點輔導之學生，協助學生提早給予關心以預防學生流失。

壹、在資料特性分析中發現：

- 一、不在學比例最高的為 1 年級學生，年級越高不在學比例越小；每學年度的第 1 學期不在學人數相對較多，第 2 學期比較少。
- 二、男女生不在學比例中，男生不在學比女生不在學多，但因學校各科男女生人數本來就有不同，若是與全校男女生人數相比後，男生不在學比例沒有特別高。
- 三、各科不在學比例中，以美容科比例較高。不同學制不在學比例則以建教班較高，建教班學生 3 個月在公司工作 3 個月在學校讀書，除需適應學校外還要適應公司，比起其他學生需做更多的調適所以應該給予更多的協助。建教班不在學人數以美容科比例最高。
- 四、不在學原因最多的是轉出（轉換環境）為非自願性原因，是因為德行成績不及格(獎懲、缺曠過多、行為有重大缺失…等)或註冊未全部完成，大部份出現在 1 年級時；2、3 年級非自願性異動減少，以休學或轉出(其他)為主。

貳、在資料採礦模型中發現

影響學生不在學因素最重要的是德行平均。因高職目前採學年學分

制，不會因學業成績而被退學，因此德行成績反而成了是否在學的重要因素。

決策樹模型中發現德行平均 <10 分不在學學生以 1 年級學生轉出（轉換環境）居多；在德行平均 ≥ 10 及 < 40 分、 ≥ 40 及 < 80 分則是 2 年級學生轉出(其他)較多；德行平均 >80 分則是休學為主。

第二節 研究建議

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 97~99 學年度資料，資料範圍未能擴大，建議後續可增加研究樣本數量或許可發現更多影響學生流失的潛在因素。

參、給個案學校的建議：

一、對於 1 年級學生而言，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，適時給予輔導加強學生的認同感，以免學生德行成績過低而被改變學習環境。

二、對在校生則是了解學生在校學習、交友狀況並加強生活輔導，且增加對學校的向心力，以減少學生流失。

三、對於建教班學生方面，由於學生 3 個月在公司工作 3 個月在學校讀書，除了要適應學校也要適應公司，比起其他學生需做更多的適應，學校可以給予更多的輔導和協助。

四、可建立預警機制，將德行成績較低及出缺席異常之學生提早列入輔導，亦可讓學生家長對學校積極輔導的態度，獲得較為正面的肯定。

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=dncldr&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. 高中職免學費方案（節錄本）公布版，(民國 101 年 3 月 16 日)，取自 http://www.tpde.edu.tw/ap/news_view.aspx?sn=02ebaac0-a01d-4139-af30-65a801df2582
7. 國民教育階段學生人數預測分析報告民國 100 年版，(民 100 年 2 月)，取自 http://www.edu.tw/files/site_content/B0013/101basicstudent.pdf
8. 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。
9. 盧梅莉，「團體成員流失之探討」，諮商與輔導，76 期，34-36 頁，民 81。
10. 謝邦昌、鄭宇庭、蘇志雄，SQL Server 2008 R2 資料採礦與商業智慧，碁峰資訊，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 Roe 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.