

應用資料採礦技術於中途輟學線上預警系統之實現

翁秉逸 謝明哲*

國立臺東大學資訊管理學系

* hzm@nttu.edu.tw

摘要

本研究以臺東縣國中學生學籍資料為例，配合電子化學籍系統的資料收集、分析，取得 9,259 筆有效資料，應用 Microsoft SQL Server 2008 資料採礦技術建立三種預測模型進行測試，比較後得到最適合的預測模型為類神經網路模型，整體預測正確率為 75.6%，可找出 80.9% 實際中輟的學生，並依此模型實現可行的線上預警系統。藉由整合資料採礦技術與電子化學籍系統，將其應用於學習輔導相關資訊系統，將可提供有效的預警機制，積極進行事先的預防措施，投注更多資源給所需要關懷的學生，以降低學生中輟率。

關鍵詞：資料採礦、中途輟學、類神經網路。

Abstract

This research achieved 9,259 valid data by collecting in electronic school system of the student data in Junior High School in Taitung County. We utilized SQL Server 2008 of Microsoft soft package to run Decision Tree, Artificial Neural Network and Naive Bayes models. After comparative analysis of these models, Artificial Neural Network turned out to be the most suitable of the three. It is able to find out 80.9% actual dropout students, and could correctly identify the variables with accuracy rate of around 75.6%. And we will build a preventive system based on this model.

We aim to provide more resources to students in need of care, and the rate of dropping out can thus decline. This result of this study will not only contribute to providing an efficacious evaluation method for Taitung County Government so that they can actively prevent student dropout, but also predict student dropout by using Data Mining techniques and electronic school system.

Keywords: data mining, dropout, artificial neural network.

1. 前言

中途輟學在我國一直是個十分受重視的議題，也進行了許多行政通報措施以及建置尋回的機制，然而再多、再及時的通報處置，皆比不上事先的預測與預防來的有效與降低社會成本[1][2][3]，若能建置一套具有參考價值的預測機制，可讓許多在校內的隱性需關懷學生可以得到更多的照

顧與資源。

過去已有不少研究者針對中途輟學的成因、環境、演進等問題進行探討整理，章勝傑、陳金燕[4]的研究指出，整體而言，雖然在預測的準確性上仍有進步的空間，以學童國小學籍卡與輔導紀錄表上記錄的六年級資料來預測他們到國中的輟學行為已具可行性與實用性。而最近幾年，雲端技術蓬勃發展，臺東縣政府為了配合教育部建置教育雲的政策，也建立起全縣的雲端教育學籍、成績資料庫[5]，使相關資料樣本收集與分析及可性大為提升。

本研究以臺東縣國中學生之學籍、成績資料等變項來探討是否影響國中學生的中輟行為表現，進而預測發生的機率，茲將本研究目的列舉如下：

- (1) 運用採礦技術分析學生歷史學籍成績資料，挖掘學生發生中途輟學行為之潛在共同因素，並建立可提供學生中輟預測之最適採礦模型。
- (2) 應用最適採礦模型實現中途輟學線上預警原型系統，以為後續發展學習輔導即時決策支援系統之基礎。

2. 資料採礦模型建置與資料分析

本研究來源為求準確性及考量學生個資安全問題，向臺東縣政府教育處申請剔除敏感個資後的臺東縣 22 所國民中學 97~99 年入學之學生學籍與成績資料。考量學校實際使用狀況與系統限制，本研究篩選成績及學籍等相關欄位作為主要採礦資料。其中成績資料包含學生在學第一學期國語成績、數學成績、英語成績等各領域成績；學籍資料則包含入學年、曠課日數、是否為低收入戶、是否為原住民、是否為外籍或大陸配偶子女(新移民)、性別、在學期間記功、記過次數、是否有中輟紀錄等。

2.1 資料淨化、整理

資料採礦可視為資料庫知識發掘(Knowledge Discovery In Database, KDD)的一部份，且為核心的步驟之一，資料採礦前更需透過資料淨化與整理以利資料的利用[6]。由臺東縣政府教育處所提供 97~99 學年度入學學生(共 9,461 位)的原始學籍、成績資料，為增加變項數目以供實驗，另加入「領域成績平均(七科平均)」、「功過相抵後數量」兩欄位。為了減少資料敏感，將學校名稱改以代號顯示(A 校、B 校、C 校...)，以及將各欄位數值轉換成程式可運算數值如表 1 所示。

表 1 淨化、整理後資料欄位說明

| NO | 欄位名稱 | 值 | 說明 |
|----|----------|-----------------------|---------------------------|
| 1 | 學校 | 代號(A、B、C...) | |
| 2 | 流水號 | 不改變,可作為索引 | 僅為資料匯出時的流水號 |
| 3 | 性別 | 轉化 1 代表男, 2 代表女 | |
| 4 | 曠課日數 | 數值 | 學生曠課日數 |
| 5 | 是否有中輟紀錄 | YES 或 NO | 學生是否有連續三天曠課紀錄 |
| 6 | 國語 | 數值 | 學生第一學期國語成績 |
| 7 | 數學 | 數值 | 學生第一學期數學成績 |
| 8 | 英語 | 數值 | 學生第一學期英語成績 |
| 9 | 社會 | 數值 | 學生第一學期社會成績 |
| 10 | 自然 | 數值 | 學生第一學期自然成績 |
| 11 | 藝文 | 數值 | 學生第一學期藝文成績 |
| 12 | 健體 | 數值 | 學生第一學期國語成績 |
| 13 | 低收 | 轉化 1 代表是, 0 為否 | 學生是否為低收入戶註記 |
| 14 | 本人身障 | 轉化 1 代表是, 0 為否 | 學生本人是否為身障註記 |
| 15 | 家長身障 | 轉化 1 代表是, 0 為否 | 學生的家長是否為身障註記 |
| 16 | 原住民 | 轉化 1 代表是, 0 為否 | 學生是否為原住民註記 |
| 17 | 新移民 | 轉化 1 代表是, 0 為否 | 學生是否為新移民註記 |
| 18 | 記過 | 數值 | 學生第一學期記過次數 |
| 19 | 記功 | 數值 | 學生第一學期記功次數 |
| 20 | 成績平均(新增) | 數值 | 國語、數學、英語、社會、自然、藝文、健體科平均分數 |
| 21 | 功過相抵(新增) | 數值 | 記功數量-記過數量=功過相抵數量 |

扣除資料不全(轉學生、特殊班級學生),可得樣本數量共 9,259 筆資料,並以 EXCEL 亂數功能取出 30% 資料(2,777 筆)作為模型測試資料。

2.2 決策樹演算法模型分析

本研究案例總計 6,482 筆,為了比較不同 SCORE_METHOD 參數以取得最佳參數值[7],本研究進行三次不同參數之決策樹模型建置,並以目標母體為中輟等於「YES」,匯入之前保留之 30%測試資料得增益圖形如圖 1,其中「決策樹 A」模型代表 SCORE_METHOD 參數為 1、「決策樹 B」模型代表 SCORE_METHOD 參數為 3、「決策樹 C」模型代表 SCORE_METHOD 參數為 4,經比較增益圖於整體母體擴展 50%時,各參數模型目標母體的百分比及預測機率,得出最佳參數值為 4(決定用於計算分岔準則的方法為 Bayesian Dirichlet Equivalent Prior),其系統評分 AUC 值為 0.82(預測機率 2.84),其餘參數經反覆測試以預設值即為最佳模型參數[7],相依性網路顯示關聯性最強的連結依序為成績平均 > 曠課(日數) > 學校,並得出決策樹圖形如圖 2,本模型所得決策樹圖形共分出 12 個葉節點(規則路徑)。

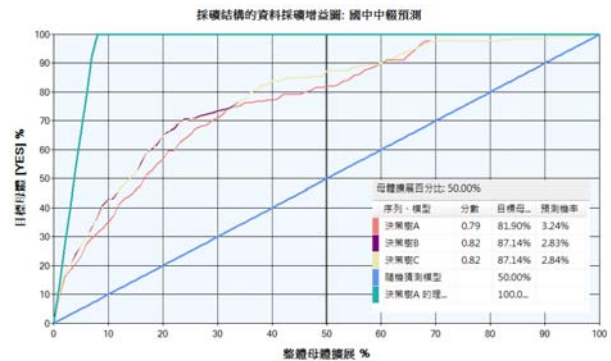


圖 1 決策數比較增益圖

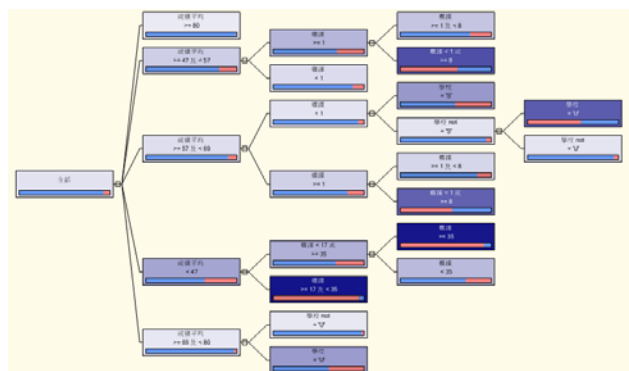


圖 2 決策樹圖形

為比較三種演算法模型,指定本研究保留 30% 學生測試資料庫進行案例測試得出模型分數(AUC 值)達 0.82,為區辨能力良好的模型[9]。為達預測目標 80% 實際中輟學生,將採礦模型預測準則設為預測中輟值為「YES」機率為 6.66% 以上則判斷預測為「有中輟」,進行實際案例測試,得表 2 分類矩陣,其中橫軸代表預測的計數,縱軸則是實際的計數。以本研究的測試資料庫顯示,決策樹模型所

預測中輟學生共 1127 位，實際中輟人數則為 176 位，準確率為 169/1127=15.6%，預測非中輟學生 1650 位，有 34 位學生實際曾經中輟，因此，為達預測 80% 實際中輟學生目標，本模型整體準確率為 64%。

表 2 決策樹模型預測分類矩陣表

| 預測 | 實際有中輟 | 實際無中輟 |
|-----|-------|-------|
| 有中輟 | 176 | 951 |
| 無中輟 | 34 | 1,616 |

決策樹模型預測正確率
 $= (176+1616)/(176+951+34+1616)=64\%$
 決策樹模型目標正確預測比率
 $= 176/(176+34)=83.8\%$
 決策樹模型預測中輟學生精準率
 $= 176/(176+951)=15.6\%$

比率意義說明如下：

- (1) 預測正確率：
預測出正確的名單佔整體預測名單的比率。
- (2) 目標正確預測比率：
預測為中輟生佔實際中輟生的比率。
- (3) 預測中輟學生精準率：
預測為中輟生且實際為中輟生佔預測為中輟生的比率。

2.3 貝式機率分類演算法模型分析

為建置貝式機率模型，經由反覆測試建置參數與類型[7]，改變各參數皆得同樣增益圖與結果，故各參數以預設值設定即可得最佳預測模型，本研究案例總計 6,482 筆，得出相依性網路(如圖 3)則顯示關聯性最強的連結依序為成績平均>曠課(日數)>國語(分數)>英語(分數)>數學(分數)>社會(分數)>自然(分數)>藝文(分數)>記過(次數)>健體(分數)>功過相抵>是否為原住民>記功次數>性別>是否低收入戶。

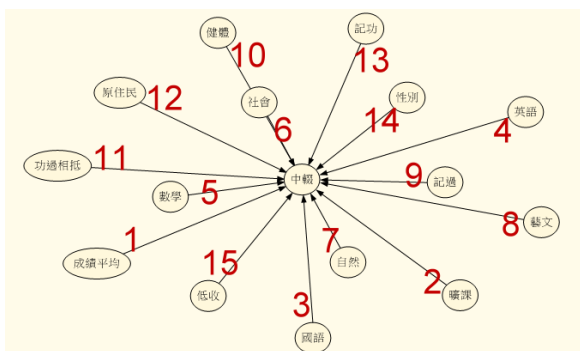


圖 3 貝式機率分類相依性網路圖

使用 SQL SERVER 2008 所提供採礦精確圖表功能，指定本研究保留 30% 學生測試資料庫進行案例測試得出貝式機率分類模型增益圖、為進行本研究比較模型目的，預測目標為 80% 實際中輟學生，

將採礦模型預測準則設為若預測中輟值為「YES」的機率限定 7.55% 以上則判斷預測為「有中輟」，並進行實際案例預測，可得統計結果數據如表 3 分類矩陣，其中橫軸代表預測的計數，縱軸則是實際的計數。以本研究的測試資料庫顯示，貝式分類模型所預測中輟學生共 918 位，實際中輟人數則為 169 位，準確率為 169/918=18.4%，預測非中輟學生 1860 位，有 42 位學生實際曾經中輟，因此，為達預測 80% 實際中輟學生目標，本模型整體準確率為 71.5%。

表 3 貝式機率分類模型預測分類矩陣表

| 預測 | 實際有中輟 | 實際無中輟 |
|-----|-------|-------|
| 有中輟 | 169 | 749 |
| 無中輟 | 42 | 1,818 |

貝式機率分類模型預測正確率
 $= (169+1818)/(169+749+42+1818)=71.5\%$
 貝式機率分類模型目標正確預測比率
 $= 169/(169+42)=80.0\%$
 貝式機率分類模型預測中輟學生精準率
 $= 169/(169+749)=18.4\%$

2.4 類神經網路模型建置分析

為取得較佳參數建立模型，本研究進行三次模型建置得增益圖形如圖 4，其中類神經 A 模型代表 HIDDEN_NODE_RATIO 參數為 0.0、類神經 B 模型代表 HIDDEN_NODE_RATIO 參數為 4.0、類神經 C 模型代表 HIDDEN_NODE_RATIO 參數為 8.0，經比較整體母體擴展為 50% 的增益圖形及預測機率得出最佳參數為 0.0(代表指定隱藏神經與輸入和輸出神經的比例為 0.0)，其 AUC 評分為 0.87，目標母體達 95.71%，其餘參數經反覆測試以預設值即為最佳模型參數[7]。

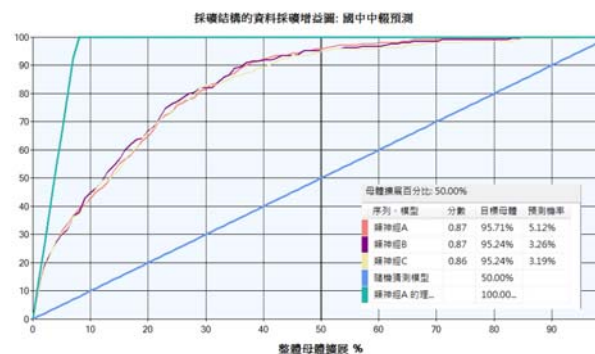


圖 4 類神經 A、B、C 模型參數之比較增益圖

本研究以保留 30% 測試資料驗證，得出預測變數機率比重如圖 5，從該圖中可顯示變數之間的交互作用，瞭解變數的重要程度，並將中輟喜好為 YES 的變數重要程度製成表 4 供參，影響輸出變數(中輟)較顯著的前十個變數依序為(曠課日數>學校>功過相抵>記過次數>藝文分數>成績平均>數學成

績>英語成績>國語成績>是否為低收入戶)。

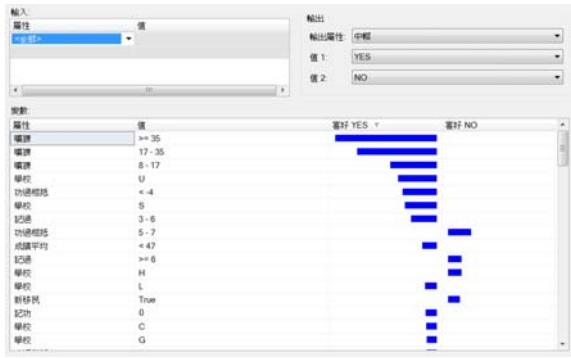


圖 5 類神經網路模型變數比重檢視圖

表 4 類神經網路模型變數喜好屬性表 (喜好值為 YES)

| 屬性 | 值 | 機率 |
|---------|---------|--------|
| 曠課 | >=35 | 96.81% |
| 曠課 | 17 ~ 35 | 87.60% |
| 曠課 | 8 ~ 17 | 62.24% |
| 學校 | U | 58.15% |
| 功過相抵 | <-4 | 57.64% |
| 學校 | S | 57.52% |
| 記過 | 3 ~ 6 | 54.31% |
| 成績平均 | <47 | 41.99% |
| 學校 | L | 41.85% |
| 記功 | 0 | 36.97% |
| 學校 | C | 34.85% |
| 學校 | G | 34.12% |
| 功過相抵 | -4 ~ -2 | 32.52% |
| 學校 | L | 29.13% |
| 曠課 | 1 ~ 8 | 28.51% |
| 藝文 | < 52 | 27.58% |
| 成績平均 | 47 ~ 57 | 26.12% |
| 英語 | <37 | 25.32% |
| 數學 | 39 ~ 52 | 13.55% |
| 英語 | 37 ~ 52 | 12.51% |
| 成績平均 | 57 ~ 69 | 10.53% |
| 國語 | <43 | 9.52% |
| 是否為低收入戶 | True | 8.14% |
| 性別 | 1 | 6.65% |
| 原住民 | True | 6.42% |
| 本人身障 | True | 6.27% |
| 家長身障 | False | 6.04% |
| 記功 | <1 | 6.01% |
| 健體 | >=84 | 5.06% |

為進行本研究比較模型目的，預測目標為 80% 實際中輟學生，將採礦模型預測準則設為若預測中輟值為「YES」的機率限定 10.14% 以上則判斷預測為「有中輟」，並進行測試案例預測，統計測試結

果得表 5 分類矩陣，其中橫軸代表預測的計數，縱軸則是實際的計數。以本研究的測試資料庫顯示，類神經網路模型所預測中輟學生共 805 位，實際中輟人數則為 170 位，準確率為 $170/850=21\%$ ，預測非中輟學生 1972 位，有 40 位學生實際曾經中輟，因此，本例整體準確率達 75%，實際所找出中輟學生比率為 $170/210=80.9\%$ 。

表 5 類神經網路模型預測分類矩陣表

| 預測 | 實際有中輟 | 實際無中輟 |
|-----|-------|-------|
| 有中輟 | 170 | 635 |
| 無中輟 | 40 | 1,932 |

類神經網路模型貝式機率分類模型預測正確率 $= (170+1932)/(170+635+40+1932)=75.6\%$
 類神經網路模型目標正確預測比率 $= 170/(169+42)=80.9\%$
 類神經網路模型預測中輟學生精準率 $= 170/(170+635)=21.1\%$

2.5 評估比較分析模型

資料採礦的用意，便是在於利用過去的歷史經驗資料來預測未來所發生的事件機率，而本研究所欲探採的目標為「中輟」事件，其本身並非屬於高機率，不容易被猜測到，自然更需要資料採礦的技術來進行預判。故本研究先以 97-99 年臺東縣學生學籍資料作為樣本訓練建構常見的採礦模型(決策樹、貝式機率分類、類神經網路模型)，繼而將另一份資料庫(預先保留的 30% 測試資料)導入三個模型之中進行驗證分析，並反覆測驗各種參數影響，判斷比較何種模型能夠取得較大的評估效益，提供有關單位參考。為達預測目的，目標屬性是為「中輟」，目標值設為「YES」，以 Microsoft SQL SERVER 2008 產生增益圖如圖 6，圖中 X 軸代表用來比較預測之測試資料百分比，Y 軸代表預測值的百分比，此圖中，最理想模型的線條以橘色顯示，藍色代表隨機猜測模型[8]。

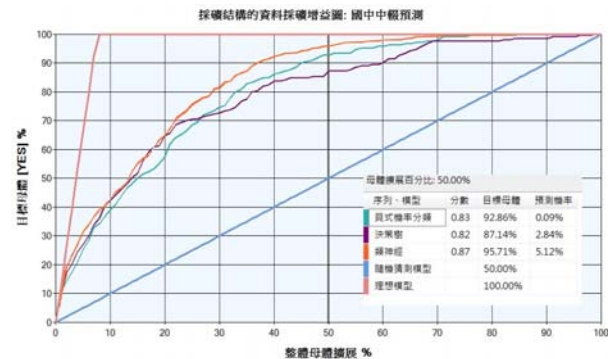


圖 6 三模型的增益圖比較

本圖可顯示「決策樹」於母體擴展 27% 以前優於「貝式分類」模型，27% 之後則為「貝式分類」模型較「決策樹」佳。然而，不論母體擴展比率為

何，類神經所提供的模型皆較佳，AUC 分數也最高，皆優於「決策樹」與「貝式分類」模型。當母體擴展百分比為 50%時，目標母體由高至低如下排序：類神經網路(95.71%)> 貝式分類(92.86%)> 決策樹(87.14%)。

經由預先保留的 30%測試資料導入三個模型之中進行驗證分析，實驗中貝式機率分類、決策樹、類神經網路皆有不錯的系統評分分數，分別為 0.83、0.82、0.87，皆高於 0.8，都可視為預測力良好的模型[9]。由此可見學籍資料可在中輟預測評估上扮演重要的角色，及提供評估的方向，建立預測的模型也是可行的。

另為比較三模型預測能力優劣，以目標找出 80%目標母體(中輟為 YES)為基準進行實際預測可得三模型比較表如表 6，其中類神經網路、貝式分類、決策樹找出實際中輟的學生數分別為 170、169、176 位，但整體正確率為 75.6%、71.5%、64.0%，雖然在系統評分上三模型皆有不錯的分數，但實際測試後所得的數據，類神經網路模型整體預測正確率最高，為本研究最適模型。

表 6 三模型比較表

| 模型名稱 | 預測 | AUC 分數 | 預測中輟精準率 | 整體預測正確率 | 優劣排序 |
|-------|-----|--------|---------|---------|------|
| 類神經網路 | 中輟 | 0.87 | 21.1% | 75.6% | 一 |
| | 未中輟 | | | | |
| 貝式分類 | 中輟 | 0.83 | 18.4% | 71.5% | 二 |
| | 未中輟 | | | | |
| 決策樹 | 中輟 | 0.82 | 15.6% | 64.0% | 三 |
| | 未中輟 | | | | |

3. 建立線上預警系統原型

3.1 系統架構

依據前一節評估與比較，本研究以「類神經網路」模型為最適模型，為了驗證此模型，並測試實際使用狀況，因此利用此最適模型建立測試平台架構如圖 7，當使用者(學校教職員)進入 SFS 校務系統之後，點選「教職員」目錄，選擇「班級學籍管理」進入學生基本資料畫面，系統便會傳送學生相關歷史學籍資料至本研究測試平台，考量系統元件相容性，本研究選擇開發 ASP.NET(C#)程式接收校務網站所送來的資訊，並以 MDX 語法(微軟為存取和管裡 Microsoft SQL Server Analysis Services 所開發出來的語言)[7]帶往 Microsoft SQL 2008 analysis Services 導入之前以學籍資料訓練得出的類神經網路採礦模型進行中輟變項預測，傳回預測結果為中輟為「YES」的機率數值，經過加工處理之後顯示

註記於校務系統之中，只要機率值大於 10.14%，代表該名學生被預測為中輟高風險，但畫面並不會直接顯示中輟字樣，而是以中輟機率百分比加上圖案作為識別，以免學生被貼標籤，並且於系統中顯示類神經網路模型運算出中輟為「YES」的機率以供使用者參考。

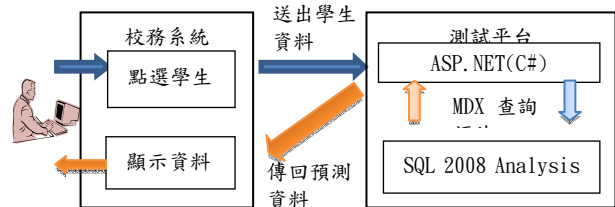


圖 7 線上預測系統架構

3.2 程式設計與實現

SFS 校務系統擁有學生的各項變數資料，平臺中的 SQL 2008 Analysis Services 則已有以歷史樣本資料訓練出來的類神經模型，接下來就是要靠一程式語言將系統中的資料變數透過 MDX 語法帶入 Analysis Services 進行預測並輸出結果[7]。本研究選擇使用 ASP.NET(C#)為接收傳遞的程式語言平台，理由及優點如下：

- (1) C#可跨平台使用，簡單易學。
- (2) 與 WEB 開發相結合：校務系統即為 WEB 形式，可直接以 WEB 開發使用者介面及後端傳遞程式，未來後續研究不論資料來源是何種作業系統與程式語言，皆能輕易結合。
- (3) 可直接使用 Microsoft 所開發元件 (Microsoft.AnalysisServices.AdomdClient)傳遞 MDX 語法與 SQL 2008 Analysis Services 溝通。

為了做為後續各級輔導單位發展可及時反應與處理之決策支援系統的基礎，本研究建立中途輟學線上預測系統原型，直接以臺東縣目前使用的 SFS 校務系統做為資料來源平台與使用者介面，當教職員完成學生相關變數資料輸入，系統即可傳送至預測平台進行資料判讀與評估，並傳回中輟機率以及示意畫面如圖 8 所示。

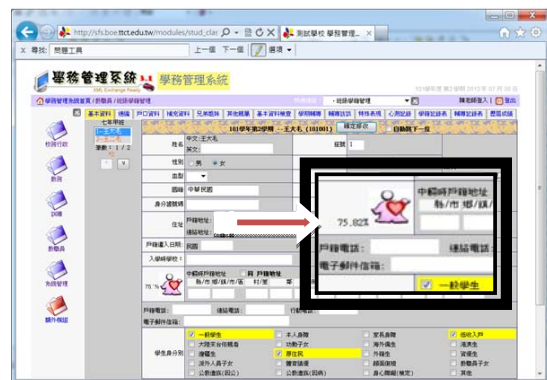


圖 8 測試平台結合校務系統實際畫面

4. 結論與建議

4.1 研究結論與貢獻

本研究針對臺東縣政府教育處所提供 97~99 學年度入學學生的原始學籍、成績資料欄位作為研究探討的母體，以常見的演算模型決策樹、貝式分類、類神經網路等三種演算方式來進行，搭配預先保留的 30% 測試資料，進行評估與分析，得出結論並說明如下：

- (1) 以影響的變數而言，從決策樹、貝式分類、類神經網路三種模型所取得較顯著的變數，可整理依序如下：
 1. 曠課日數：既然中輟的定義本來就與曠課日數息息相關，自然曠課日數較高，中輟的機率就愈高，各級教育單位應盡早完成線上登錄曠課系統，相關統計數字也可即時作為評估、及早介入輔導之用。
 2. 成績平均：學生的學習能力或許不能歸類於是否輟學的重要變數，各科領域成績高低也不代表其他領域的好壞。但是據教育部統計中輟的數據上，對學校生活不感興趣、不適應學校課程、考試壓力確為重要因素之一。
 3. 學校：部份偏鄉地區學校中輟率較高，顯示該地區學生需較多關懷與關注。
 4. 記過次數：從學生日常表現中，或許可發現個人因素而造成中輟的因子，如觸犯校規、行為偏差等。
 5. 功過相抵：為了避免銷過造成影響判斷數據，因此除了比較記過次數之外，另外加入功過相抵數量。
- (2) 本研究使用 ASP.NET (C#) 語法可成功以 WEB 介面連結前端資料庫(PHP+MYSQL)以及後端的預測平台，其他單位可循此模式建立自己本身的預測機制。
- (3) 資料採礦的本質，就是以過去的經驗來協助找出稀有的事件[10]，而本研究亦有此一特性，以類神經網路模型為例，從 2,777 筆測試資料中正確預測筆數為 2,102 筆，於本研究中整體預測力可達 75.6%，正確率佳，為最適模型，但預測中輟精準率僅 21.1%，顯示模型仍有進步空間。
- (4) 本論文主要貢獻為整合資料採礦技術(Data Mining)與電子化學籍系統，將其應用於政府相關資訊系統之評估，利用相關學籍資料變數，結合資料採礦方法，可做為臺東縣國中教育有關單位參考，以現有的資料做為資料採礦輸入變數，預測學生中輟機率。及早投入輔導人力與關懷給更需要的學生，進而降低全縣的中輟率。

4.2 建議

- (1) 本研究以電子化校務行政系統建檔的資料作為母體樣本，進行資料採礦所花費的準備、處理

時間相對較少，資料來源為各校教職員建立，可信度也較過去問卷調查為高，建議各縣市政府及早完成學籍教育資料庫，以利各項政策決策支援與學術研究。

- (2) 臺東全縣的訓練樣本資料或許不足以給某特定鄉鎮或單一學校較高機率的模型建置，建議後續研究者可以針對單一學校或單一地區(例：離島地區)的學生資料做採礦模型。
- (3) 未來研究建議，因本研究資料來源僅為目前臺東縣電子化校務系統可得資料，尚未考量其他變數，若未來得到其他因素或取得更新資料後，可再進一步資料採礦探究，待後續者繼續研究。

參考文獻

- [1] 教育部訓育委員會(2003)·各國中途輟學學生現況與輔導措施·臺北市：黃金博物館區。
- [2] 教育部 (2003)·教育部中輟生通報及復學系統·取自 <https://www.mlss.edu.tw/>。
- [3] 郭靜晃(2001)·中途輟學青少年之現況分析及輔導·臺北市：洪葉文化。
- [4] 章勝傑、陳金燕(2003)·從國小六年級學籍資料預測國中輟學行為的邏輯迴歸分析與區辨分析·臺東大學教育學報，頁 125-151。
- [5] 臺東縣教育網路中心(2008)·臺東縣 SFS 學籍系統平台·取自 http://sfs.boe.ttct.edu.tw/school_list/。
- [6] Fayyad, Gregory, and Smyth. (1996). From Data Mining to Knowledge Discovery in Databases. AI Magazine, 37-54.
- [7] Microsoft. (2010). Microsoft msdn:Analysis Services. Retrieved from <http://msdn.microsoft.com/en-us/library/bb522607.aspx>.
- [8] 邱宏彬、許依宸(2011)·資料採礦在學生流失偵測上之應用·南華大學資訊管理研究學術期刊，第十一期，頁 83-100。
- [9] David W. HosmerStanley Lemeshow, Rodney X. Sturdivant Jr. (2013). Applied Logistic Regression. New York: John Wiley and Sons.
- [10] Kleuissner. (1998). Data mining for the enterprise. In Proceedings of the Thirty-First Hawaii International Conference on, Volume 7, 295-301.