校務專題：預測退學高風險群

在推動精準教育的過程中，利用本校校務資料已建立退學高風險群的預測，為有效防止退學發生，在產生高風險名單後，通知系上與導師，進行學習輔導。此預測模型自 107 學年度起於校園中推動，並逐步導入教師系統；部份內容發表已獲得刊登於國際期刊。

(摘要連結、期刊全文連結)

精準有效輔導追蹤

導入名單

學務管理系統⇒導師學習關懷管理作業⇒退學高風險管理作業⇒匯入退學名單

輔導轉介

班導師可至教師資訊系統查看「退學高風險名單」，進行學習輔導，並填寫關懷表轉介相關單位

持續追蹤

追蹤退學高風險群二年級學習成效

精準教育—降低退學高風險學生之行政措施

退學高風險群預測名單

建立退學高風險群組（院或系）

設置溝通平台（導師系統或社群軟體）

落實輔導（課輔與轉系）

給予關懷（解除困境）

耐心傾聽（了解問題）
Precision education with statistical learning and deep learning: a case study in Taiwan

前言

高等教育普及之後，人才多元培育，提供社會及產業需求人力，有助社經發展。在如此競爭的環境裡，大學新生順利或經過一番奮鬥後進入大專校院，之後是否能適應大學生活？選讀校院系是否符合興趣？由教育部學校基本資料庫（via https://udb.moe.edu.tw）可得，101學年度大專校院休學總人數為49,337人增至106學年度的78,241人；退學總人數則由101學年度的84,719人增至106學年度的91,556人。在本研究將以機器學習和統計學習對同一資料進行退學預測，期待取得可行及可信的預測結果，予提升學生學習績效的成功率。

研究目的

(1) 利用統計預測模式與機器學習演算，以大學低年級學習行為預測未來高年級時學習失敗機率可行性。
(2) 探討機器學習演算及統計學習預測在預測學生學習績效的適用性。
(3) 確認影響學生學習績效的學習投入面變項以為強化補救教學介入。

研究方法

以二至四年級退學生共412人为研究組，在學學生共3,140人为對照組，進行學生學習失敗機率與影響因素之探討。自本校校務資料庫中撈取學生學習行為資料進行分析，包括學生個人背景資料、一年級學習表現、在學或退學學籍狀態等資料，並以統計學習與機器學習演算進行模型預測。

(1) 統計學習：依變數「退學」為二元(binary)型態的類別資料，假設退學事件發生勝算的對數和可能的風險變數為線性關係，利用邏輯斯迴歸(logistic regression analysis)進行分析。
(2) 機器學習演算：使用深度神經網路 (deep neural networks)、演算法為多層感知器 (multilayer perceptron) (Khashei et al, 2007)。

分析結果

利用單變項分析找出與退學相關變項後，以多變項邏輯斯迴歸分析和多層感知器 (multilayer perceptron) 演算法進行訓練，分別得到一預測模型，對於退學模式預測能力的正確率 (accuracy)、敏感度 (sensitivity) 與特異度 (specificity)，會隨不同退學預測機率的設定而不同，當設定預測機率10%以上者為退學生時，可得到最佳的敏感度與特異度，此時邏輯斯迴歸模式成功預測在學生與退學生的成功率68%，其中預測退學生的敏感度為61%；多層感知器演算法成功預測在學生與退學生的成功率77%，其中預測退學生的成功率為53%。

| 退學機率 | 預測機器 | 准確率 | 敏感度 | 特異度 |
|----------|------------|--------|--------|--------|
| 10%      | Multilayer perceptron | 77%    | 53%    | 80%    |
| LOGISTIC | Logistic regression | 68%    | 61%    | 69%    |
| 15%      | Multilayer perceptron | 83%    | 42%    | 88%    |
| LOGISTIC | Logistic regression | 80%    | 50%    | 84%    |
| 20%      | Multilayer perceptron | 87%    | 42%    | 93%    |
| LOGISTIC | Logistic regression | 84%    | 38%    | 90%    |
| 50%      | Multilayer perceptron | 90%    | 16%    | 99%    |
| LOGISTIC | Logistic regression | 88%    | 0%     | 100%   |

討論與結論

透過邏輯斯迴歸分析，找到影響學習失敗的因素，可用於未來關懷與輔導之用；並提供機器學習有效運算的資訊，再從中找到規則，進而做出預測。單只預測被退學的機率，機器學習可有90%的正確率，而邏輯斯迴歸亦有89%的正確率。但是，退學高敏感度是我們所期待的，若所需輔導改善成本不大，或許可以暫時不去考慮偽陽性者(false positives, 即在學新生被預測為退學的人)，取得退學高危險群預測名單後，加以輔導與改善學習，可降低學生學習績效不佳情形為精準教育