

多模型比較下之學生休退風險預測研究



李馨茹、張祐銘、李銘宗、梁懷良

弘光科技大學 校務研究辦公室

摘要

挑戰
 高等教育在少子化背景下，學生休退學比例逐年攀升。

方法
 以XGBoost為預測模型，能更準確識別高風險學生。
 透過SHAP分析皆是關鍵重要因子。

研究結果
 早期預警工具有效辨識潛在休退風險學生，亦可作為學生預警與輔導決策之依據。

目標
 提前辨識高風險學生，有助強化輔導並提升留才率。

研究發現

變數建構

14個關鍵預測因子：

- 學術因素：學業成績、學業輔導、學業輔導次數、112學業成績
- 人口學背景因素：性別、年齡
- 經濟因素：家庭經濟狀況、家庭成員平均所得
- 生活適應因素：學校生活適應、生活適應

模型表現

模型(Model)	AUC	Recall
GLM_Selected	0.866	0.109
GLM_All	0.856	0.119
RF_Selected	0.866	0.706
RF_All	0.868	0.691
XGB_Selected	0.873	0.735
XGB_All	0.869	0.716

關鍵因子排序

低落的學業成績與輔導次數，是提升休退風險的主要原因。

SHAP摘要分析

SHAP摘要分析顯示，學業成績與輔導次數是影響休退風險的最重要因素。

SHAP關鍵因子決策依據

SHAP摘要分析顯示，學業成績與輔導次數是影響休退風險的最重要因素。

研究問題

我們的探尋：三個關鍵問題，一個明確目標

- 能否建立更精準的預警系統？
 建構學生休退風險預測模型，提升高退學學生之早期識別能力。
- 新技術真的優於傳統方法嗎？
 比較傳統統計模型與機器學習模型於預測效能上的差異。
- 我們能信任機器的預測結果嗎？
 結合 SHAP 模型解釋方法，提升預測結果之透明度與數據應用價值。

研究設計

Stage 1 變數篩選
 從各類資料中篩選預測因子，篩選工具：變量選擇高斯選擇、檢定方法、類別變量：LR Test、連續變量：Wald 檢定、篩選程序：Backward Stepwise、建立變數預測因子。

Stage 2 模型建構
 以篩選後的變數建構三種預測模型：Logistic Regression、Random Forest、XGBoost。

Stage 3 模型評估
 衡量模型的預測力與預測力：AUC (Area Under the Curve)、Recall (召回率)。

Stage 4 模型解釋
 創新機件決策邏輯與因果風險來源，導入SHAP (SHapley Additive exPlanations)方法深入解釋模型的運作邏輯。

資料來源

113學年度學生為研究對象

112學年度前置因素

總數：6,410
 休退學學生：681

結論與建議

我們的模型不僅能精準預測，更能提供可行動的洞察以支持學生成功

- 精準預測 (Accurate Prediction)**
 XGBoost 模型表現最佳 (AUC值 0.873)，能有效辨識出高達 73.5% 的高風險學生 (Recall)；為早期介入提供可靠基礎。
- 關鍵洞察 (Key Insights)**
 學業成績與輔導次數是影響休退風險的關鍵因素，為提升學業成績、增加輔導次數提供行動建議。
- 未來應用 (Future Application)**
 本模型可作為輔導老師的預警工具，幫助輔導人員精準識別高風險學生，並在適當的時機提供介入。