

應用多層感知器於學生休退學預測—以朝陽科技大學學生為例

鄭煜輝¹，施慕馨²

¹ 朝陽科技大學資訊與通訊系

E-mail: yhcheng@cyut.edu.tw

² 朝陽科技大學資訊與通訊系

E-mail: s11230602@cyut.edu.tw

摘要

臺灣的少子化已導致多間學校面臨危機，而學生休退學也成為各校永續經營的核心議題。本研究使用多層感知器模型作為學生休退學預測之深度學習方法，旨在協助學校更有效掌握休退學的情況，從而提前部署輔導資源，提升資源配置的精準度。本次研究數據範圍定於朝陽科技大學 106 學年度至 110 學年度五年間的學生資料，其中 110 學年度第 2 學期資料用於模型測試，其餘 9 個學期資料則做為訓練集及驗證集，結果顯示，模型預測準確度達 96.05%，未來將持續匯入新學期資訊進行模型更新，並實際提供校方使用，作為輔導資源配置的重要依據。

關鍵字：多層感知器；深度學習；機器學習；休退學預測；精準教育。

Abstract

In recent years, Taiwan's Sub-replacement fertility has posed significant challenges to many schools, with student suspensions and dropouts becoming critical issues for the sustainable operation of educational institutions. This study aims to address these issues by developing a deep learning system using a Multilayer Perceptron (MLP) model to predict student suspensions and dropouts. The system helps schools to better understand and manage these issues, enabling them to allocate counseling resources more effectively and in a timely manner. The dataset used in this study covers five academic years (from 2017 to 2021) at Chaoyang University of Technology. Data from the second semester of the 2021 academic year was used for model testing, while the data from the remaining nine semesters were used for training and validation. The model achieved a prediction accuracy of 96.05%. In the future, new semester data will be continually integrated to update the model, providing the

university with a valuable tool for student counseling and resource allocation.

Keywords : Multilayer Perceptron; Deep Learning; Machine Learning; Suspension and Dropout Prediction; Precision Education.

1. 研究背景

近十年來，臺灣面臨嚴峻的少子情勢。根據國家發展委員會的統計數據[1]，2012 年新生兒數量為 22 萬 9,481 人，至 2022 年驟降至 13 萬 8,986 人，少子化儼然成為不容忽視的社會問題，更對教育產業帶來嚴重衝擊。隨著生源下降，各校缺額逐年增加，大專校院運用校務研究方法針對「招生」與「留生」提供實質的分析，作為學校制定策略之依據。休退學預測已有許多學者提出可行方法並發表文獻供後人作為研究參考。Jovial Niyogisubiz 等人於 2022 年提出「Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization」[2]，該研究以隨機森林、極端梯度提升機、梯度提升技術和前饋神經網路進行混合及堆疊，相較基礎模型在同樣的學生資料集下表現更佳。Cho, Choong Hee 等人於 2023 年發表「A Study on Dropout Prediction for University Students Using Machine Learning」[3]，使用了邏輯迴歸、決策樹、隨機森林、支持向量機、深度神經網路和光梯度提升機等多種機器學習方法進行訓練並比較，結果顯示光梯度提升機模型性能最佳。2023 年，Kiran Fahd 和 Shah J. Miah 發表「Effectiveness of data augmentation to predict students at risk using deep learning algorithms」[4]，使用多層感知器識別高休退學風險學生，並使用數據增強技術來補充數據實例和平衡數據集。

2. 研究方法

本研究以朝陽科技大學過去學生的 9 項相關資料進行模型訓練，以本學期學生在學資料預測下一學期該生在學狀態是為一般在學、休學或是退學，最後彙整休退學預測名單予校方，進行更有效並精確之輔導資源投入。綜合上述，可知分屬多元分類問題，且資料集類型無圖片或語音等多維資料，因此選擇多層感知器作為預測模型架構。

2.1. 資料預處理

研究資料欄位經評估後，彙整可能作為影響學生休退學之因子，包含該學期平均成績、系排名 PR、班排名 PR、請假(節)、曠課(節)、輔系修讀、雙主修修讀、學行優良獎獲取以及學業進步獎獲取等 9 項。依照本次研究設定範圍，篩除非必要資訊及含空值之樣本，使用 Excel 進行各項學生資訊串接，最終取得 138,141 筆學生資料。

2.2. 多層感知器模型

本研究使用多層感知器(Multilayer Perceptron，簡稱 MLP)架構建置模型，MLP 由輸入層、隱藏層和輸出層組成，每層包含多個神經元，並透過激活函數做更多元的處理，廣泛應用於分類、迴歸等任務，而本研究便是屬於分類之範疇。本次建置之模型包含 1 層輸入層、2 層隱藏層以及 1 層輸出層，輸出層依本研究篩出之休退學相關因子共設有 9 個節點、第一層隱藏層 64 個節點、第二層隱藏層 32 個節點、輸出層則依標籤類別數量設置 3 個節點。模型架構如圖 1。

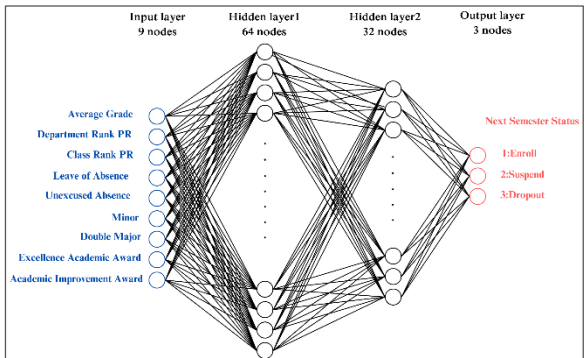


圖 1. 學生休退學預測模型架構圖

3. 研究結果

本研究將朝陽科技大學 106 學年度第 1 學期至 110 學年度第 1 學期共 9 學期 125,202 筆樣本做分割，80%作為訓練集、20%作為驗證集，並將 110 學年度第 2 學期 12,939 筆資料單獨作為測試集，訓練結果顯示，預測正確資料共 12,427 筆，此模型準確率達 96.05%，然從精確率、召回率以及 F1-score 可以看出在休學及退學的預測表現均不佳，這可能導致實際應用上的缺誤，未來擬透過平衡數據比例方法進行模型改善。混淆矩陣請見表 1，模型效能指標請見表 2。

表 1.混淆矩陣

Confusion Table		Real		
		1:Enroll	2:Suspend	3:Dropout
Predict	1:Enroll	12,416	180	319
	2:Suspend	0	0	0
	3:Dropout	9	4	11

表 2.效能指標

Performance	1:Enroll	2:Suspend	3:Dropout
Accuracy	96.05%		
Precision	96.14%	N/A	45.83%
Recall	99.93%	N/A	3.33%
F1 - score	97.48%	N/A	6.15%

4. 參考文獻

[1] "國家發展委員會人口推估查詢系統." https://pop-proj.ndc.gov.tw/Custom_Fast_Search.aspx?n=7&sms (accessed 01 October, 2024).

[2] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, and P. C. Nshimyumukiza, "Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100066, 2022.

[3] C. H. Cho, Y. W. Yu, and H. G. Kim, "A Study on Dropout Prediction for University Students Using Machine Learning," (in English), *Appl. Sci.*, Article vol. 13, no. 21, 2023, Art no. 12004, doi: 10.3390/app132112004.

[4] K. Fahd and S. J. Miah, "Effectiveness of data augmentation to predict students at risk using deep learning algorithms," (in English), *Soc. Netw. Analysis Min.*, Article vol. 13, no. 1, 2023, Art no. 113, doi: 10.1007/s13278-023-01117-5.