

112.1學期第1次校務研究業務報告會議

報告人：黃俐瑜

報告時間：30分鐘



報告大綱

1

專案議題報告：建立學生休退學預警模式

2

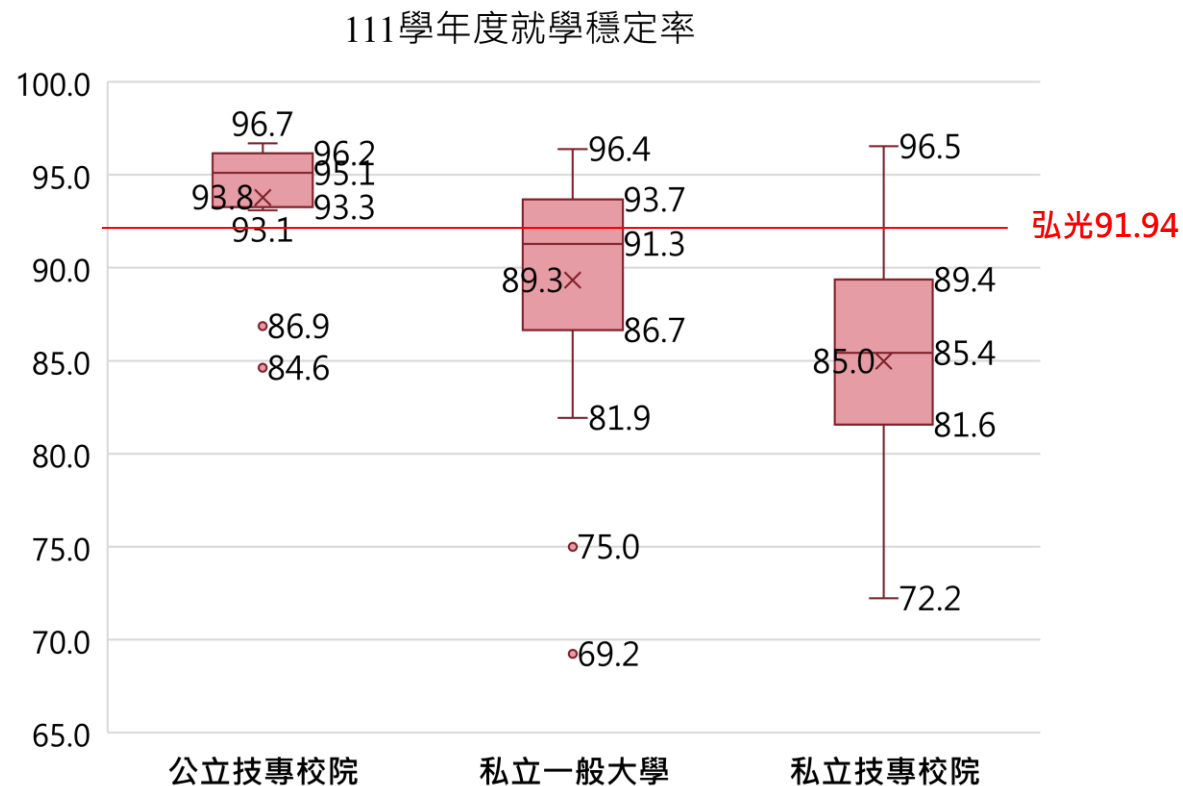
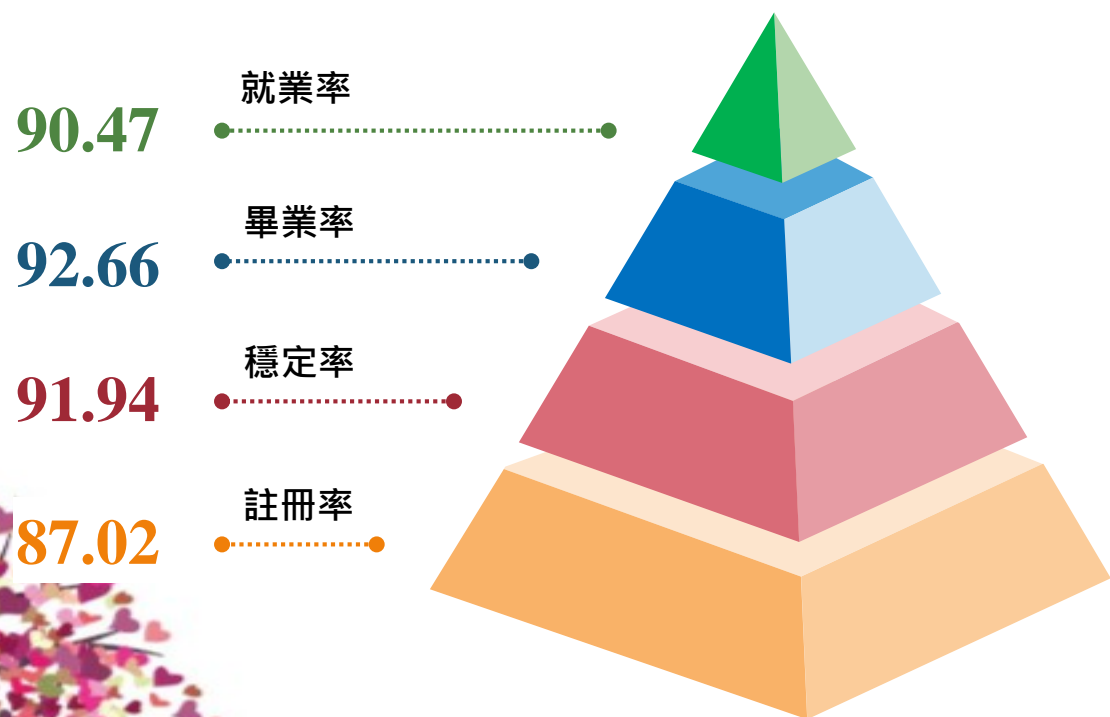
臨時動議

專案議題報告： 建立學生休退學預警模式



目的：重新定位與改善學校計畫和政策

☆永續發展目標 SDGs 目標4：確保有教無類、公平以及高品質的教育，及提倡終身學習。



文獻：文獻蒐集與研究流程

■ 中文

- 大學休退率與必修課程之相關研究(嶺東科大)
- 基於人工智慧技術預測學生的休退狀況(朝陽科大)
- 應用多重支持之廣義關聯分類法建構大學休退預測系統(高雄大學)
- 應用堆疊是降造自動編碼器建構學生退學預測(台北科大)
- 大學休退學影響因素之分析(清華大學)

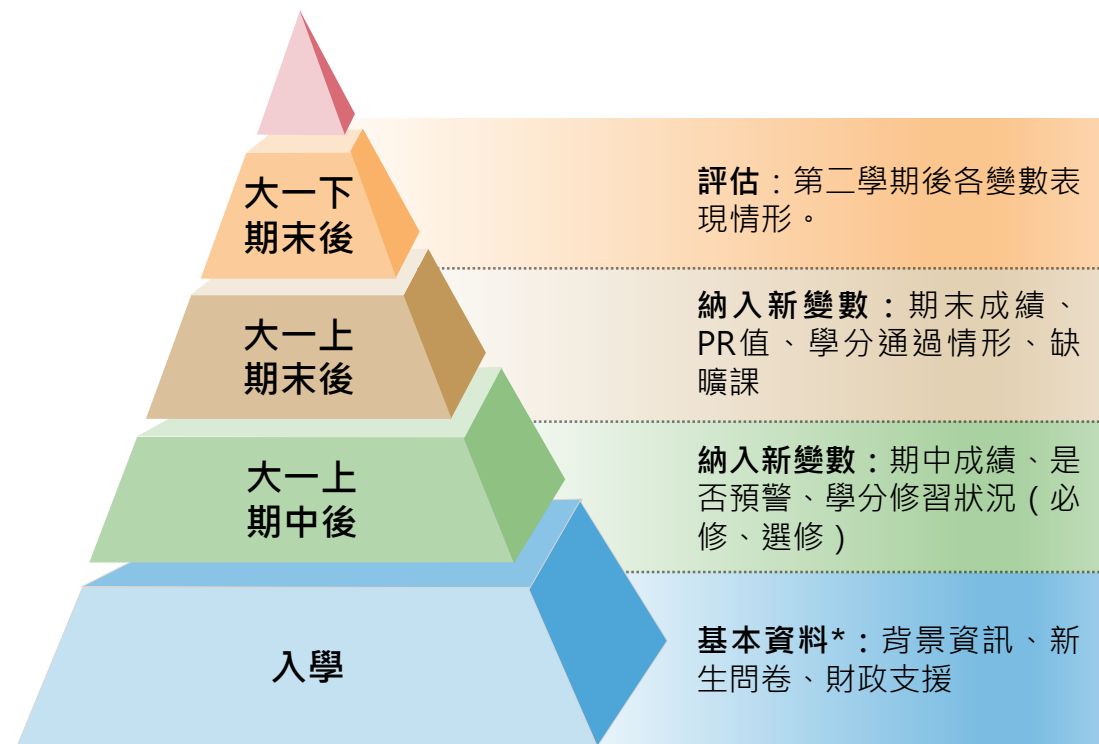
■ 英文

- Precision education with statistical learning and deep learning: a case study in Taiwan
- The Data Revolution Comes to Higher Education: Identifying Students at Risk of Dropout in Chile
- Early Dropout Prediction using Data Mining: A Case Study with High School Students



文獻：運用時機與模型定義

- ◎ **就學穩定**：為第三學期(即第二年的第一學期) 於同所大學進行註冊，亦指新生隔年乃留在原校之情形。
- ◎ **調查時間**：108.1~111.2學年度日四技學生(N=7,691)。



***基本資料*：**

背景資訊：性別、年齡、縣市、入學管道、高中職群類別、科別符合情形；新生問卷：父母收入、志願序；財政支援：減免身分、減免金額、校內工讀

文獻：統計方法與優劣勢

流程	說明	方法	優勢	劣勢	呈現
特徵篩選	減少冗餘的特徵影響預測性。	逐步回歸(step)	減少找尋最佳子集的時間。	1.可能選入過多變數。 2.造成模型過度擬合。	
		遞迴特徵消除(RFE)	以交叉驗證的方式找到最佳解。	1.計算成本過高。(40分) 2.難以處理嘈雜數據。	(隨機森林法)
		L1、L2正規化	提供特徵與目標函數之間的連結。	可能導致淡化部分模型的權重。	√(ridge法)
模型建置	依據問題的種類建置預測模型。	線性回歸(OLS)	建模速度快、可解釋性高。	不適合非線性數據，對異常值過於敏感。	
		羅吉斯迴歸	可用於觀測樣本發生機率。	難以處理大量多特徵之變數。	√
		決策樹(CART)	適用於有缺失或無相關特徵的數據。	1.造成模型過度擬合。 2.忽略數據間的關聯。	
評估指標	衡量模型整體表現的指標。	混淆矩陣	有助於了解分類的所有訊息。	難以處理發生機率極小的不平衡數據。	√(ROC/AUC 曲線)
模型測試	評估模型訓練結果的好壞。	交叉驗證	數據有限情況下可避免模型過度擬合。	計算成本過高。	√

方法：模型初擬與研究流程

以108~111學年度資料進行模型的建置。
(日四技)

資料蒐集

- 1.L2正規化找出特徵與目標函數間的連結。
- 2.建模70驗證30。
- 3.隨機種子2023。

模組設定

以預警資料來瞭解干預措施的有效性。

成效評估

資料清理

- 1.高中職群類對應系所招收群類。
- 2.類別變數的轉換。
(二分法)
- 3.休退時間點的計算。

統計分析

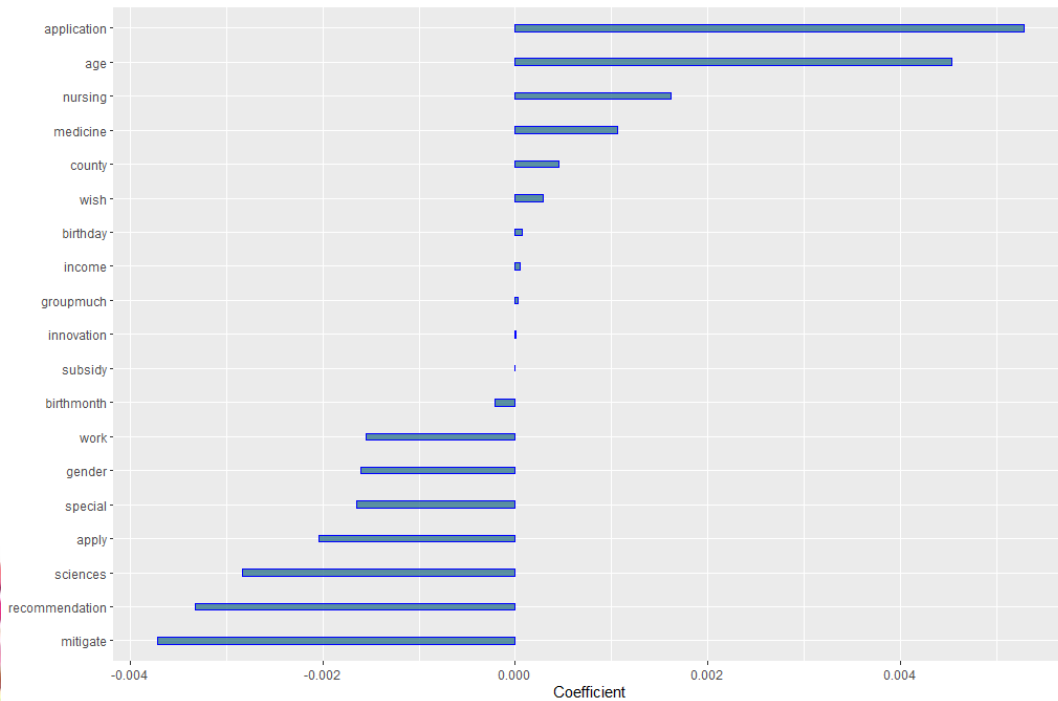
- 1.羅吉斯模型
- 2.混淆矩陣
3. ROC曲線
- 4.差異性分析

方法：研究資料與分析流程



- 108~111學年度日四技學生。
- 背景資料、期中學習資料、期末學習資料。
- 階段性納入變數。
- 特徵篩選：L2正規化。
- 模型建置：羅吉斯迴歸。
- 評估指標：混淆矩陣(ROC曲線)。
- 模型測試：交叉驗證。
- 數值型：變異數分析
- 類別型：卡方檢定
- 小提琴圖

方法：分析成果的解釋與意涵 - L2正規化



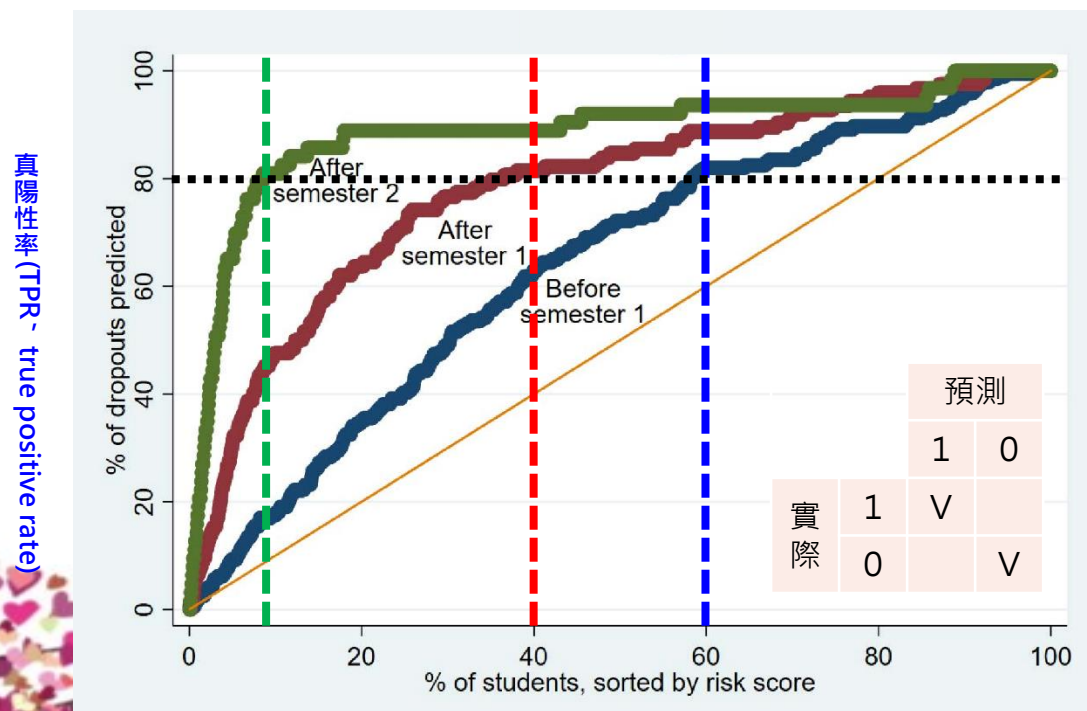
解釋：

1. 在目標函數中加入L2懲罰函數，針對過多參數的模型進行懲罰。
2. 當懲罰效果越大，係數會越接近於0。
3. L2的優點是能將具有相關性的變數推向彼此，並保留所有的變數(不重要的變數其權重會趨近於0)，進而清楚的識別出模型中真正的訊號(signals)。

意涵：

1. 減少過度配適的情形發生。
2. 瞭解各變數的權重情形與彼此間的關聯性。

方法：分析成果的解釋與意涵 - ROC曲線



偽陽性率 (FPR - false positive rate)

解釋：

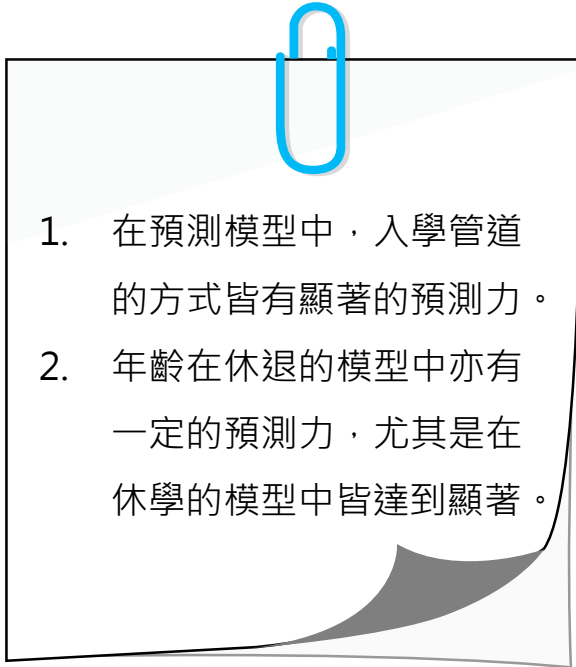
1. 錯誤預測風險的機率(X)。準確預測風險的機率(Y)。
2. 曲線顯示：預期成本(X)與實質效益(Y)間的關聯性。
3. 對角線參考線代表隨機預測。

意涵：

1. 與其針對高風險群體進行干涉，不如廣泛且低成本的進行干涉，如郵件、活動、專案課程或是廣收相關專業等。
2. 入學：雖然目標不明確，但可如前述所進行低成本的干涉。
3. 一下：雖AUC從有64%提高到89%，但時間上可能有1成的學生已輟學。(文獻：The Data Revolution Comes to Higher Education: Identifying Students at Risk of Dropout in Chile)

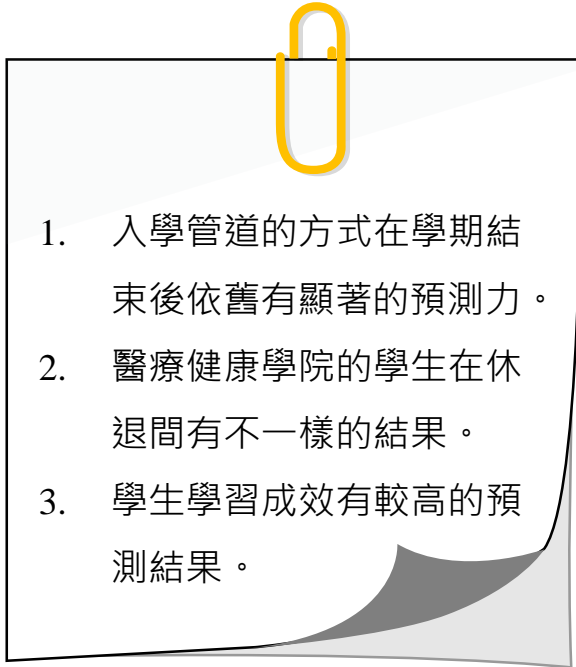
模擬：模型建置成果(入學) - Summary

模型	參數	Estimate	Std. Error	z value	Pr(> z)	ACU	差異	
休學	(Intercept)	-7.460	0.948	-7.873	0.000 ***	0.6148		
	性別	-0.247	0.226	-1.091	0.275			
	年齡	0.192	0.046	4.127	0.000 ***			>19
	群類相符	0.224	0.277	0.811	0.418			
	推甄入學	-0.779	0.227	-3.441	0.001 ***			否
	申請入學	-0.428	0.413	-1.035	0.301			
	技優入學	-0.571	0.441	-1.295	0.195			
退學	醫療學院	0.238	0.223	1.064	0.287			
	(Intercept)	-10.025	2.881	-3.479	0.001 ***	0.7522		
	性別	-0.882	0.789	-1.118	0.264			
	年齡	0.161	0.151	1.069	0.285			
	申請入學	2.163	0.819	2.639	0.008 **			是
	登記分發	1.371	0.734	1.867	0.062 .			
是否減免	1.053	1.058	0.995	0.320				

- 
1. 在預測模型中，入學管道的方式皆有顯著的預測力。
 2. 年齡在休退的模型中亦有一定的預測力，尤其是在休學的模型中皆達到顯著。

模擬：模型建置成果(期末) - Summary

模型	參數	Estimate	Std. Error	z value	Pr(> z)	ACU	差異
休學	(Intercept)	-6.491	1.199	-5.414	0.000***	0.8203	
	年齡	0.263	0.061	4.334	0.000***		>19
	推甄入學	-0.670	0.315	-2.125	0.034*		否
	登記分發	-0.413	0.310	-1.330	0.184		
	醫療學院	0.536	0.274	1.957	0.050.		
	學期成績	0.048	0.011	4.340	0.000***		<65.7
	學分通過率	-0.072	0.008	-9.575	0.000***		<79.2
退學	(Intercept)	-9.608	2.051	-4.685	0.000***	0.8813	
	性別	0.454	0.339	1.338	0.181		男
	年齡	0.200	0.107	1.861	0.063.		
	推甄入學	-0.793	0.365	-2.173	0.030*		否
	醫療學院	-0.963	0.501	-1.920	0.055.		否
	是否預警	0.641	0.064	10.071	0.000***		>3

- 
1. 入學管道的方式在學期結束後依舊有顯著的預測力。
 2. 醫療健康學院的學生在休退間有不一樣的結果。
 3. 學生學習成效有較高的預測結果。

模擬：模型建置成果 - ROC圖

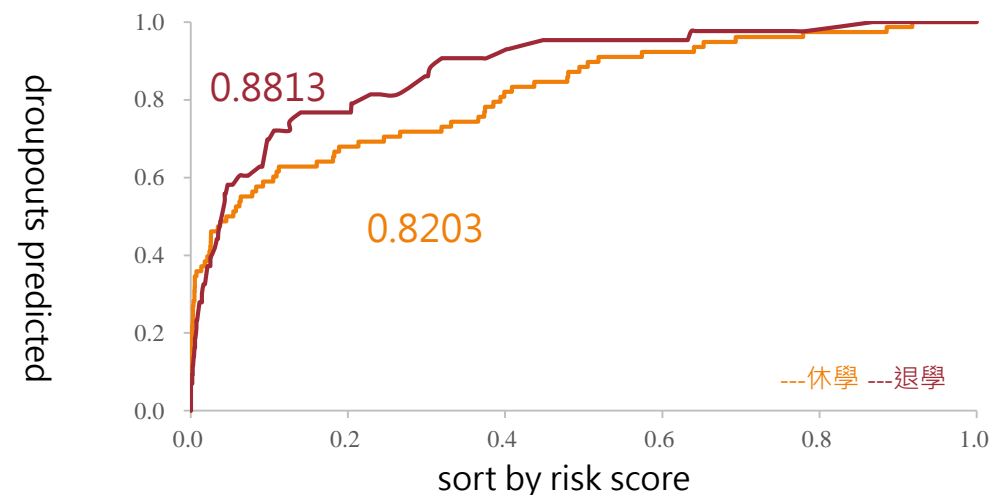
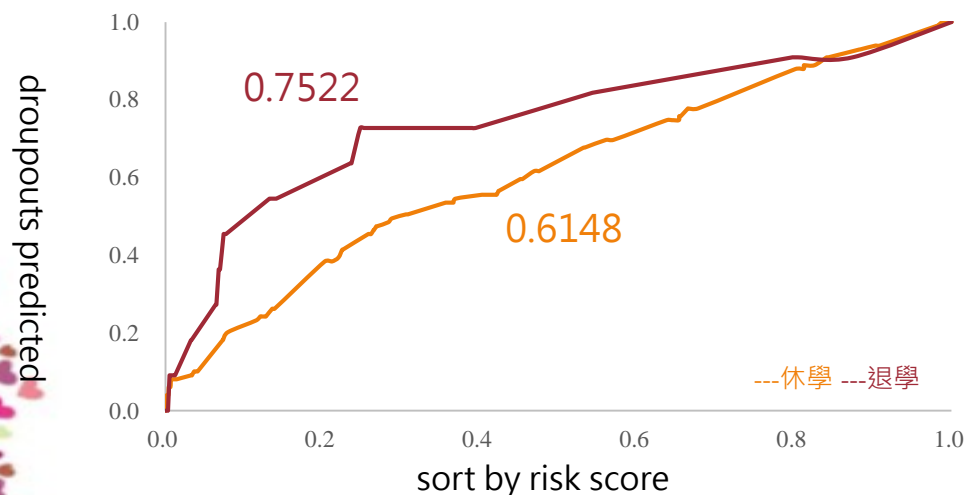
雖然大部分的個體可預測出其結果，但在精準度來看僅有期末後的退學預測成果較優。

入學即休退

		休學預測		退學預測	
		1	0	1	0
實際	1	0	99	0	11
	0	0	6,005	0	6,104

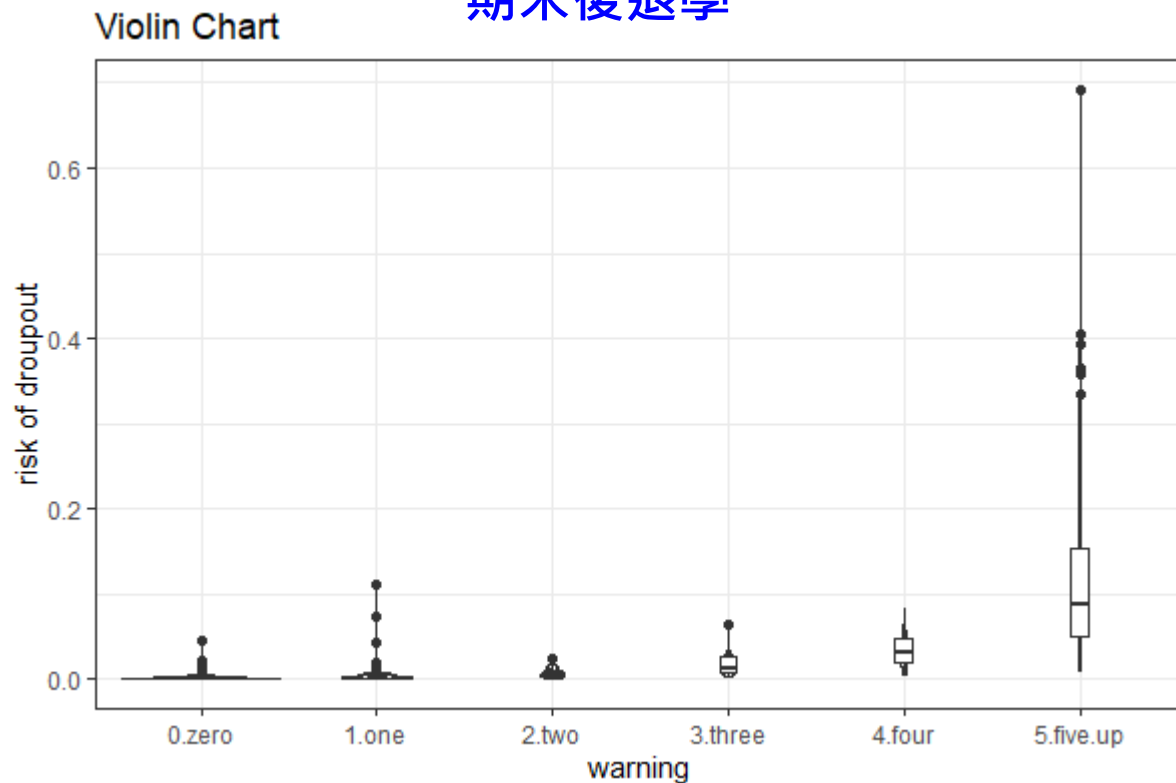
期末後休退

		休學預測		退學預測	
		1	0	1	0
實際	1	0	43	11	67
	0	1	6,004	0	5,923



評估：干預措施的成效評估

期末後退學



1. 學期末後預警科目越多者，期退學的風險也越高。
2. 另一方面，在預警科目達到3科以上後，干預措施的成效未達預期。

結論：總結研究結果

01

精準

精準針對**高風險族群實施干預計畫**，即使參與的人數高達50%以上。

02

預測

學生學習情形具備高預測性，可評估出高輟學的學生加以預防。

03

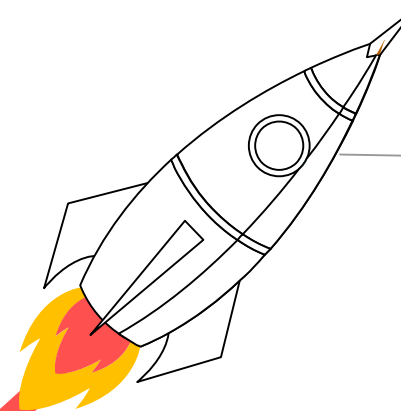
成本

依據風險值進行**優先序**，可降低成本。

04

時效

開學前兩周就應針對學生參與活動、出缺勤等進行干涉評估。



干涉措施



建議：後續研究規劃

研究方向



技術精進



優化更新



臨時動議



簡報完畢，恭請指導

