# 學業壓力與憂鬱風險之相關性分析

## 摘要

本研究以 27 901 筆來自大學生的問卷資料為基礎，探討學業壓力（Academic Pressure）與憂鬱風險（Depression）之間的關聯性。採用描述性統計、群組比較、卡方檢定、主成分分析（PCA）以及兩種機器學習預測模型（Logistic Regression、Random Forest）。結果顯示學業壓力與憂鬱風險呈現中度正相關（r ＝ 0.475），不同壓力層級的憂鬱比例從低壓力組 19.44％、中壓力組 52.03％ 到高壓力組 81.63％；卡方檢定顯示壓力層級間差異顯著（χ² ＝ 5740.656，df ＝ 2，p ＜ 0.001）。Logistic Regression 模型準確率 73.4％、AUC 0.805，Random Forest 準確率 69.0％、AUC 0.751；兩者均將學業壓力視為最關鍵特徵。研究建議學校針對高壓力學生配置更多心理健康資源，並發展壓力管理課程。

## 1 緒論

1.1 研究背景與動機

* 隨著競爭加劇與課業負擔提升，大學生的心理健康議題日益受到重視。學業壓力已被證實與焦慮、憂鬱等負面情緒有關，但在台灣大學生族群中的實證研究尚不足。 1.2 研究目的
* 本文旨在量化學業壓力與憂鬱風險的關聯強度，並透過統計檢定與機器學習模型，評估壓力指數對憂鬱風險的預測能力，為校園心理輔導提供決策依據。

## 2 研究方法與資料

2.1 資料來源與樣本概況

* 資料集來源：匿名線上問卷，最終讀取 CSV 檔 /content/student\_depression\_dataset.csv，共 27 901 筆有效樣本，18 個欄位。
* 主要欄位：
  + Academic Pressure（學業壓力，0–5）
  + Depression（憂鬱風險，0 ＝無、1 ＝有）
  + 其他數值變項：Age、CGPA、Study Satisfaction 等
  + 2.2 描述性統計

| 變項 | count | mean | std | min | 25% | 50% | 75% | max |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | 27 609 | 25.82 | 4.91 | 18 | 21 | 26 | 30 | 59 |
| Academic Pressure | 27 901 | 3.14 | 1.38 | 0 | 2 | 3 | 4 | 5 |
| CGPA | 27 901 | 7.66 | 1.47 | 0 | 6.29 | 7.77 | 8.92 | 10 |
| Study Satisfaction | 27 901 | 2.94 | 1.36 | 0 | 2 | 3 | 4 | 5 |

2.3 資料清洗流程

* 保留原始副本後，將 Depression 類別型映射為 0/1。
* 填補 Degree 缺值並歸類為「高中及以下／大學／碩士／博士」四級序數。
* 刪除重複值與 Academic Pressure、Depression 之缺失，最終樣本數 27 873。
* 數值變項以中位數填補缺漏，並以 Z-score (|z|<3) 去除離群觀測。

## 3 探索性分析與視覺化

3.1 分層群體憂鬱風險

* 將學業壓力切分為「低壓力／中壓力／高壓力」三組，樣本數分別為 4 799、11 631、11 443。
* 群組憂鬱比例：
  + 低壓力組：19.44％
  + 中壓力組：52.03％
  + 高壓力組：81.63％
  + 3.2 條形圖演示（示意）
* 左圖：三組憂鬱比例比較；右圖：學業壓力連續值五等分後之憂鬱比例。
* 由圖可見壓力越高，憂鬱風險明顯攀升。

## 3.3 預留區塊：XAMPP + MySQL + Grafana 資料視覺化整合（未來擴充）

為強化資料分析系統的擴展性與實務應用，本研究預留以下架構與區塊，以整合 MySQL 資料庫與 Grafana 資料視覺化平台：

### (1) 架構說明

* **資料庫系統：** 使用 XAMPP 套件內建之 MySQL Server（port: 3306）
* **視覺化平台：** Grafana（連接資料來源為 MySQL，port: 3000）
* **資料輸入方式：**
  + 將 student\_depression\_dataset.csv 處理後匯入 MySQL 資料表 student\_depression
  + 使用 Python pymysql 或 sqlalchemy 上傳清洗後 DataFrame 至資料庫
* **資料查詢樣板：**
* SELECT Academic\_Pressure\_Category,  
   COUNT(\*) AS total,  
   ROUND(SUM(Depression) / COUNT(\*), 4) AS depression\_rate  
  FROM student\_depression  
  GROUP BY Academic\_Pressure\_Category;

### (2) 預期 Grafana 整合圖表（待實作）

| 視覺化圖表類型 | 說明 |
| --- | --- |
| 條形圖（Bar Chart） | 不同學業壓力等級下的憂鬱比例（Depression Rate） |
| 熱力圖（Heatmap） | 學歷 × 憂鬱程度 × CGPA 的交互分布 |
| 折線圖（Time Series） | 若未來整合時間欄位，可顯示學期壓力波動與心理狀況變化趨勢 |
| Dashboard | 將各圖表集中於一頁，供決策參考或學校輔導單位持續監控學業壓力風險狀態 |

### (3) 預留程式接口（Python → MySQL）

# 預留程式碼 - 將 df 寫入本地 XAMPP MySQL  
import pymysql  
from sqlalchemy import create\_engine  
  
engine = create\_engine("mysql+pymysql://root:your\_password@localhost:3306/depression\_db?charset=utf8mb4")  
df.to\_sql('student\_depression', con=engine, if\_exists='replace', index=False)

### (4) 預期貢獻

* 本模組未來整合後，將提供學校輔導單位一個即時可查、視覺化動態儀表板，協助監測高風險學生群體
* 同時培養學生實務資料庫整合與儀表板製作能力，符合資料分析跨系統應用趨勢

✅ **備註**：未來完成程式後，請補上：

1. MySQL 資料表建構 SQL（CREATE TABLE 語法）
2. Grafana 所使用的資料查詢與 Panel 配置說明（可附圖）

## 4 推論性分析

4.1 卡方檢定

| 組別 | 有憂鬱 (1) | 無憂鬱 (0) |
| --- | --- | --- |
| 低壓力 | 933 | 3 866 |
| 中壓力 | 6 052 | 5 579 |
| 高壓力 | 9 341 | 2 102 |

* 卡方值 χ² ＝ 5 740.656，df ＝ 2，p ＜ 0.001
* 結論：不同壓力層級間憂鬱風險差異顯著。
* 4.2 主成分分析（PCA）
* PC1 解釋變異比例 73.7％。
* 前五大載荷量排序：Age (0.999)、Academic Pressure (0.026)、degree\_ord4 (0.017)、Study Satisfaction (0.003)…
* 學業壓力在 PC1 中排名第 2，顯示其對整體變異的貢獻度具次要影響力。
* 4.3 與其他變項相關性

| 相關對象 | Academic Pressure\_Value |
| --- | --- |
| CGPA | –0.025 |
| Age | –0.077 |
| Study Satisfaction | –0.112 |

* 顯示學業壓力與 CGPA、學習滿意度、年齡略呈負相關。

## 5 預測模型與比較

5.1 Logistic Regression

* 準確率 73.4％、AUC 0.805
* 混淆矩陣：
  + Precision(有憂鬱)=0.80, Recall=0.73
  + Precision(無憂鬱)=0.66, Recall=0.74
  + 5.2 Random Forest
* 準確率 69.0％、AUC 0.751
* 5.3 特徵重要性
* 兩模型均將學業壓力列為第一重要特徵，彰顯其預測價值。
* 建議採用 Logistic Regression 作為初步風險篩檢工具，Random Forest 可作為穩健性驗證。

## 6 結論與建議

6.1 主要發現

1. 學業壓力與憂鬱風險中度正相關（r ＝ 0.475）
2. 壓力越高，憂鬱比例從 19.44％ 直線上升至 81.63％
3. 兩種分類模型均將學業壓力視為關鍵預測變項

6.2 實務建議

* 高壓力學生應優先安排心理諮商與壓力管理工作坊
* 可在學期中期設立自我壓力評估機制，及早介入
* 未來研究可納入社交支持、睡眠品質等交互作用變項
* 6.3 未來研究
* 探討多元心理健康指標如焦慮、睡眠、飲食習慣等
* 結合長期追蹤資料，以觀察壓力變化軌跡與憂鬱發展

### 口頭報告時間分配建議（共 15 分鐘）

1. 緒論與動機：2 分鐘
2. 資料來源與清洗：2 分鐘
3. 探索性分析：3 分鐘
4. 推論性檢定與 PCA：3 分鐘
5. 預測模型成果：3 分鐘
6. 結論與建議：2 分鐘