# 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 69.3％、AUC 0.750；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 資料視覺化整合系統（XAMPP + MySQL + Grafana）

本研究成功實現完整的資料視覺化整合系統，將 Python 分析結果透過 MySQL 資料庫串接至 Grafana 視覺化平台，形成三層架構的動態監控系統：

### (1) 系統架構與資料流

* **資料前處理層**：Python 資料科學生態系（pandas, scikit-learn）
  + 特徵工程、模型訓練與預測結果產出
  + 透過模組化設計支援持續資料更新與重新分析
* **資料持久層**：MySQL 關聯式資料庫（XAMPP, port: 3306）
  + 集中存儲多維度分析結果與原始資料
  + 支援索引優化與複雜查詢，確保視覺化效能
* **視覺化展示層**：Grafana 儀表板（port: 3000）
  + 基於資料庫查詢建立即時更新的互動式圖表
  + 支援多維度篩選與下鑽分析

### (2) 資料流轉與自動化

整體系統採用單向資料流設計，確保分析結果的一致性與可追溯性：

┌───────────────┐ ┌───────────────┐ ┌───────────────┐  
│ Python │ │ MySQL │ │ Grafana │  
│ 分析環境 │ ──> │ 資料庫 │ ──> │ 視覺化平台 │  
└───────────────┘ └───────────────┘ └───────────────┘  
 │ │ │  
 資料清洗 資料持久化 互動式展示  
 特徵工程 預計算視圖 多維度篩選  
 模型訓練 索引優化 警報機制  
 預測評估 查詢加速 分享功能

資料庫寫入採用完全自動化方式，透過 SQLAlchemy ORM 實現：

# 自動化資料庫寫入流程  
from sqlalchemy import create\_engine, text  
  
# 1. 建立資料庫連線  
engine = create\_engine("mysql+pymysql://root@localhost:3306/depression\_db?charset=utf8mb4")  
  
# 2. 自動建立資料表結構（若不存在）  
with open('db/create\_table.sql', 'r') as sql\_file:  
 with engine.connect() as conn:  
 conn.execute(text(sql\_file.read()))  
 conn.commit()  
  
# 3. 匯出資料至資料庫  
df.to\_sql('student\_depression', con=engine, if\_exists='replace', index=False)

### (3) 視覺化圖表與應用

系統整合四種核心視覺化圖表，各自針對不同分析維度：

| 視覺化類型 | 分析目標 | 技術實現 | 應用場景 |
| --- | --- | --- | --- |
| **壓力層級條形圖** | 呈現三組壓力層級的憂鬱比例對比 | SQL + Grafana Bar Gauge | 快速識別高風險群體 |
| **壓力-憂鬱熱力圖** | 展示壓力值與憂鬱狀態的二維分布 | SQL + Grafana Heatmap | 識別臨界閾值 |
| **學歷-壓力-憂鬱線圖** | 比較不同學歷層級的壓力與憂鬱關係 | SQL + Grafana Time Series | 差異化干預策略 |
| **性別統計表格** | 展示不同性別的壓力與憂鬱指標 | SQL + Grafana Table | 群體差異分析 |

視覺化部分特別針對校園心理健康監測設計，支援：

1. **下鑽分析**：從整體趨勢到個體層級的多層次探索
2. **警報機制**：當憂鬱比例超過設定閾值時自動通知
3. **互動式篩選**：支援多維度條件篩選，如學歷、性別等
4. **定期更新**：支援新資料導入後的自動更新視圖

### (4) 資料庫設計與優化

為提升系統效能與擴展性，資料庫設計包含：

* **合理的索引策略**：
* -- 建立關鍵查詢欄位的索引  
  CREATE INDEX idx\_pressure\_category ON student\_depression(Academic\_Pressure\_Category);  
  CREATE INDEX idx\_degree ON student\_depression(Degree4);  
  CREATE INDEX idx\_depression ON student\_depression(Depression);
* **預計算統計視圖**：
* -- 創建常用統計視圖，減少重複計算  
  CREATE OR REPLACE VIEW v\_depression\_by\_pressure AS  
  SELECT  
   Academic\_Pressure\_Category AS pressure\_level,  
   COUNT(\*) AS total\_count,  
   SUM(Depression) AS depression\_count,  
   ROUND(SUM(Depression) / COUNT(\*), 4) AS depression\_rate  
  FROM student\_depression  
  GROUP BY pressure\_level  
  ORDER BY CASE pressure\_level  
   WHEN '低壓力' THEN 1  
   WHEN '中壓力' THEN 2  
   WHEN '高壓力' THEN 3  
   ELSE 4 END;

### (5) 實際貢獻與價值

此整合系統為校園心理健康監測提供多重價值：

1. **及早預警**：透過視覺化儀表板快速識別高風險學生群體，促進及早干預
2. **資源分配**：依據實證數據最佳化心理健康資源配置
3. **長期追蹤**：支援縱向資料分析，觀察干預措施的長期效果
4. **跨領域協作**：心理輔導人員與資料分析師的橋樑
5. **標準化工具**：可套用至不同學校或機構的心理健康監測

## 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 |
| --- | --- |
| Work/Study Hours | 0.208 |
| CGPA | 0.022 |
| Age | -0.225 |
| Study Satisfaction | -0.168 |

* 顯示學業壓力與工作/學習時數呈正相關，與年齡和學習滿意度呈負相關，與 CGPA 的相關性較弱。

## 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 69.3％、AUC 0.750
* 5.3 特徵重要性
* Logistic Regression 前五大特徵重要性：
  + Academic Pressure\_Value: 1.137
  + Age: 0.523
  + Study Satisfaction: 0.334
  + CGPA: 0.077
  + Gender\_Male: 0.022
* Random Forest 前五大特徵重要性：
  + Academic Pressure\_Value: 0.138
  + Age: 0.030
  + Study Satisfaction: 0.009
  + CGPA: 0.003
  + Gender\_Male: 0.001
* 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

## 口頭報告時間分配與重點建議（共 15 分鐘）

| 章節 | 時間 | 報告重點 | 簡報建議 |
| --- | --- | --- | --- |
| **1. 緒論與動機** | 2 分鐘 | • 大學生心理健康重要性• 學業壓力與憂鬱關聯缺乏實證 | • 引用相關統計數據• 使用簡單視覺化圖表 |
| **2. 資料與方法** | 2 分鐘 | • 問卷資料來源與特點• 變數定義與資料清洗流程 | • 呈現資料探索過程• 強調樣本大小與代表性 |
| **3. 探索分析** | 3 分鐘 | • 三組壓力層級憂鬱比例差異• **資料視覺化整合系統展示**• 互動式儀表板功能 | • **實機展示 Grafana 儀表板**• 高亮差異最顯著的數據點 |
| **4. 統計檢定** | 3 分鐘 | • 卡方檢定結果與解釋• PCA 分析關鍵發現• 相關性矩陣解讀 | • 使用熱力圖呈現相關性• 標示統計顯著性 |
| **5. 預測模型** | 3 分鐘 | • 兩種模型的準確率與 AUC• 特徵重要性排序• 模型優劣勢比較 | • 使用 ROC 曲線• 特徵重要性長條圖 |
| **6. 結論建議** | 2 分鐘 | • 學業壓力是關鍵預測因子• 校園心理健康資源配置建議• 未來研究方向 | • 總結關鍵數據• 視覺化實務建議 |

**答問時間準備建議**：針對資料視覺化整合系統，準備額外的技術細節說明，包括系統架構、資料庫設計、以及如何將此系統擴展至其他校園。特別準備回應關於系統效能與擴展性的問題。