

數學焦慮與學習成效分析：基於教育數據挖掘之實證研究

Math Anxiety and Learning Effectiveness Analysis: An Empirical Study Based on Educational Data Mining

孫長元，莊維仁*，張智凱
臺灣臺南大學數位學習科技學系
*10305stu@gmail.com

【摘要】本研究利用教育數據挖掘探討數學焦慮、學習行為與成績之關聯。經 K-Means、特徵排序與決策樹分析顯示，學生群體具顯著異質性，存在「高焦慮高投入」類別；「線上課程資源」為預測成績首要因子。決策樹揭示，低壓力者適合「加法策略」增加投入，高壓力者應採「減法策略」專注核心作業。資料模式顯示，策略選擇對高焦慮學生具重要調節潛力，本研究據此提出適性化學習建議，為教育實務提供參考。

【關鍵字】數學焦慮；教育數據挖掘；K-Means；決策樹；適性化學習

Abstract: This study explores the association between math anxiety, learning behaviors, and performance using educational data mining. Through K-Means, feature ranking, and decision tree analysis, results reveal significant student heterogeneity, including a "high anxiety, high engagement" group. "Online course resources" emerge as the primary performance predictor. Decision tree models indicate low-stress students benefit from an "addition strategy" to increase engagement, while high-stress students should adopt a "subtraction strategy" focusing on core tasks. Results indicate strategy selection holds crucial moderating potential, offering personalized recommendations for educational practices.

Keywords: Math Anxiety, Educational Data Mining (EDM), K-Means Clustering, Decision Tree Analysis, Adaptive Learning

1. 前言

在 STEM 教育中，數學能力為核心素養，但全球高達 30% 的學生面臨數學焦慮 (Maloney & Beilock, 2012)。Ashcraft 與 Kirk (2001) 指出焦慮會佔據工作記憶，降低運算表現；而 Eysenck 等人 (2007) 的「效能控制理論」進一步表明，焦慮主要損害認知處理的「效率」，學生常透過補償性努力維持成績。過去研究難以解釋「高投入卻低產出」的非線性現象 (Papamitsiou & Economides, 2014)，而教育數據挖掘 (EDM) 的普及使我們能從歷程中提取潛在模式 (Baker & Yacef, 2009)。本研究旨在透過 EDM 實證數據延伸「效能控制理論」，探討：(1) 考量焦慮與投入的交互作用，是否存在異質性學生群體？(2) 哪些特徵是預測成績的關鍵因子？(3) 教育科技是否具備調節學習成效的潛力？

2. 研究方法

本研究採用 Orange Data Mining 平台 (Demšar et al., 2013)，遵循 CRISP-DM 流程進行多層次挖掘。

2.1. 資料來源與變數界定

採用 Kaggle 公開之「學生學習成效與學習行為資料集」(14,003 筆)。該數據未標註特定文化背景與教育階段，為本研究之限制，但其大樣本特性仍具普遍實證價值。關鍵變數定義如下：(1) 目標變數：期末成績 (FinalGrade, 0-4 分連續變數)；(2) 心理特徵：焦慮程度 (StressLevel, 量表轉換為 0 低、1 中、2 高之次序變數)；(3) 行為特徵：週學習時數 (StudyHours, 0-20 小時連續變數)、線上課程數 (離散整數)、作業完成度 (0-100% 連續變數)；(4) 環境特徵：教育科技使用頻率 (EduTech, 類別變數)。

2.2. 預處理與演算法設定

針對缺失值採平均數與眾數填補，並排除考卷分數以防過度擬合。核心演算法為：(1)K-Means 分群：連續變數 Z-score 標準化，採歐幾里得距離，依輪廓係數確立最佳分群 $K=3$ 。(2)Rank 排序：結合 Information Gain (計算對目標變數的熵減少量) 與 ReliefF (基於最近鄰樣本距離更新權重) 雙重演算法評估特徵貢獻。(3)決策樹 (CART)：以基尼不純度為分裂準則，為防過度擬合並維持教育可解釋性，最大樹深限為 4-5 層，並限制最小葉節點樣本數。

3. 研究結果

3.1. 異質性的學生行為畫像 (K-Means)

將樣本投射至二維空間，識別出三類群體：(1) 高焦慮-高投入群：反映「補償性努力」現象，試圖以極長工時彌補效率低落；(2) 低焦慮-穩健群：情緒調節佳，學習時數適中但產出穩定；(3) 中度焦慮-表現導向群：適度焦慮轉化為學習驅力。三類學生畫像如圖 1 所示。

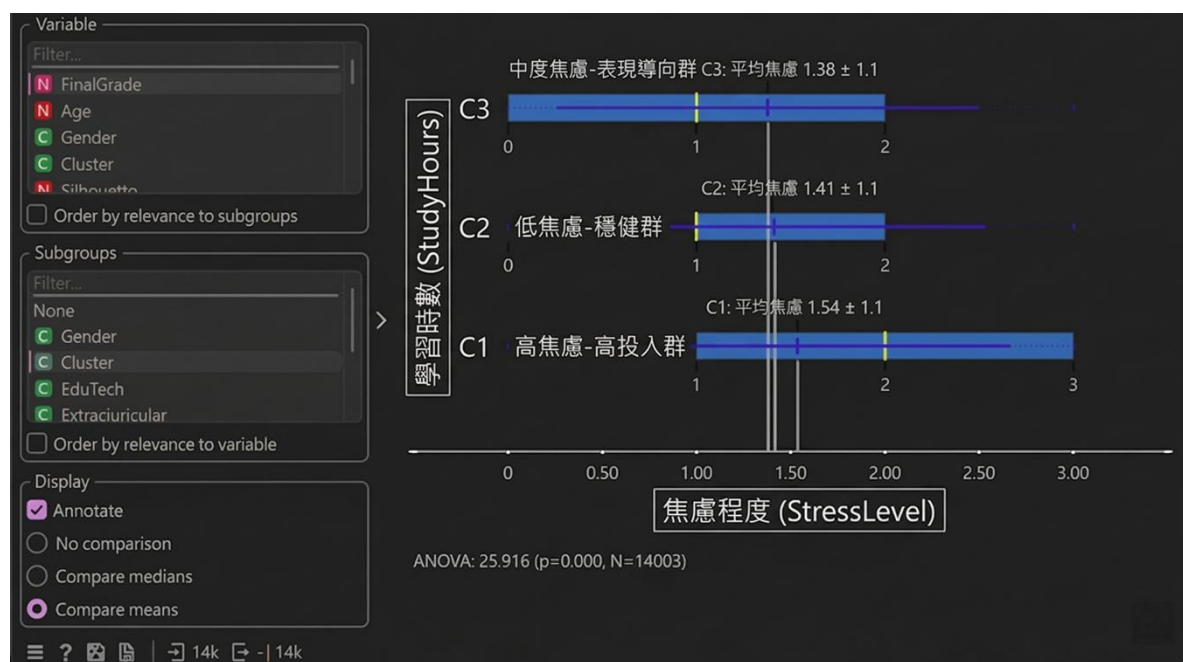


圖 1、K-Means 學生行為畫像分群圖

3.2. 關鍵預測因子 (Rank 特徵排序)

Rank 分析顯示，「線上課程修習數」為最強的正向預測因子；「焦慮程度」居次，為最關鍵的負面干擾因子。EduTech 雖未居首，但暗示其具備情境依賴的調節潛力。特徵排序結果如圖 2 所示。

		#	Univar. reg.	RReliefF
1	N OnlineCourses		NA	0.176
2	N Attendance		NA	0.170
3	N AssignmentCompletion		NA	0.167
4	N LearningStyle		NA	0.112
5	N StudyHours		NA	0.105
6	N Motivation		NA	0.079
7	N StressLevel		NA	0.074
8	N Resources		NA	0.068
9	C EduTech	2	NA	0.009
10	C Extracurricular	2	NA	0.007
11	C Internet	2	NA	0.006
12	C Discussions	2	NA	0.005

圖 2、特徵重要性
3.3. 科技介入與
(Decision Tree)

排序圖
策略路徑

決策樹以「焦慮程度」為根節點，劃分出兩條行為路徑：低壓群的「加法策略」：無焦慮干擾下，「主動投入」為關鍵。修習較多線上課程且出席率極高者，平均成績達最高點(2.2)；缺乏資源投入者則跌至最低(1.1)。高壓群的「減法策略」：對於高焦慮學生，在高作業完成度的前提下「克制」線上課程數量，成績可維持中上(1.7)；反之，若試圖攝取過多資源，成績顯著下滑(1.4)，此模式高度吻合認知負荷過載現象。結果如圖 3 所示：

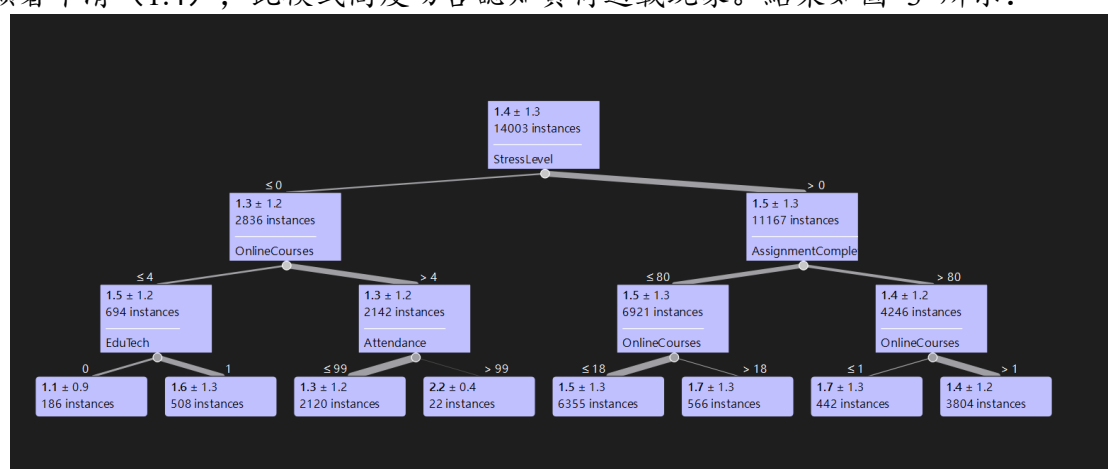


圖 3、決策樹路徑分析圖

4. 討論與結論

本研究運用 EDM 方法，為「效能控制理論」提供了大數據層面的實證支持。研究不只停留在預測成績，而是透過決策樹路徑具體描繪出認知負荷受限的現象。數據模式顯示，成績表現與其對應壓力狀態的學習策略具備高度關聯。低焦慮者適合資源最大化的「加法哲學」，展現自我調節學習中的資源管理能力 (Zimmerman, 2002)；高焦慮學生在攝取過多資源時成績下滑，更適合「減法哲學」以減輕認知負荷。此外，教育科技的「外包記憶」與適性化鷹架設計 (Mayer, 2014)，在高焦慮群體中展現出正向調節的潛力。

對教育實務之啟示與未來展望：教學績效不僅取決於學生特質，更仰賴教師的教學實踐與課堂環境營造。建議教師應針對高焦慮學生給予「精簡任務」的減法鷹架，並營造低批判

性的安全課堂環境以降低成就焦慮。此外，此「資源減法策略」亦有潛力轉移(Transferability)應用至科學或語言學習等同樣容易引發高認知負荷的學科中。

本研究揭示之數據模式尚不足以作出絕對的因果斷言，這些發現更適理解為基於數據模式的策略啟示。為進一步增強結果的穩健性與方法論說服力，未來研究將考慮引入隨機森林(Random Forest)或支持向量機(SVM)等機器學習演算法，與現有決策樹模型進行對比分析，並結合生理訊號更精準地捕捉焦慮機制的動態影響

參考文獻

- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, 130(2), 224–237.
<https://doi.org/10.1037/0096-3445.130.2.224>
- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
- Demšar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M., ... & Zupan, B. (2013). Orange: Data mining toolbox in Python. *Journal of Machine Learning Research*, 14, 2349–2353.
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2), 336–353.
<https://doi.org/10.1037/1528-3542.7.2.336>
- Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences*, 16(8), 404–406.
<https://doi.org/10.1016/j.tics.2012.06.008>
- Mayer, R. E. (Ed.). (2014). *The Cambridge Handbook of Multimedia Learning* (2nd ed.). Cambridge University Press.
- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, 17(4), 49–64.
- Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601–618. <https://doi.org/10.1109/TSMCC.2010.2053532>
- Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)*, 252–254. ACM.
<https://doi.org/10.1145/2330601.2330661>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2