

## 英國政府使用公式來預測學生被取消考試時的分數，猜猜誰做得好？

數據處校務研究中心 葉昱佐

- 改寫自 Kelsey Piper. (2020, Aug 22). The UK used a formula to predict students' scores for canceled exams. Guess who did well. *Vox Media*.
- Retrieved from <https://www.vox.com/future-perfect/2020/8/22/21374872/uk-united-kingdom-formula-predict-student-test-scores-exams>

每年春天，英國學生會參加 A 級考試(普通教育高級程度證書 General Certificate of Education: Advanced Level, 簡稱 A-Level)，作為申請大學的依據。

今年因為 Covid-19 疫情，導致春季的 A 級考試被取消。取而代之的是，政府採取了一種頗具爭議的方法來評估這些沒有考試成績的學生的入學資格：使用數學規則來預測學生在考試中的表現，然後將這些估算值作為實際成績的替代方案。

政府採取的方法非常簡單，它想猜測如果學生參加考試其表現如何，**並使用兩個輸入(input)：今年學生的成績，以及學生就讀學校的歷史記錄。**

因此，如果在校成績最好的學生通常在 A 級考試都獲得良好的分數，則甲生在校成績優異將被預測為在考試中取得良好的分數；如果該校在歷史上成績好的學生並沒有相對應獲得 A 級考試的高分，即使甲生在成績優秀，也會被預測為較低的考試分數。

整體效果如何？公式預測結果指出，家境富裕的學生比獲得相同成績但家境貧窮的學生預測表現更好，因為對學校「預期」表現的調整與學校的富裕程度密切相關。有錢的學生往往在 A 級考試表現較好，因此預測過程會偏向就讀富裕學校且成績較好的學生。

預測結果使得學生、父母及老師極為不滿，英國政府因而調整政策並宣布，如果學生在校成績優異，卻被預測考試分數較低者，將以老師估計學生可能會拿到的分數為主(以高分為主)。

英國一下子發生許多事情：Covid-19 大流行，糟糕的行政決策加劇了這種情況。儘管幾乎與 AI 無關，這也說明了通常被稱為「AI 偏誤」或「AI 倫理」這種困境。

## 預測不公平的世界

想像一下，在一個世界，有錢的學生和貧窮的學生吸毒的可能性是一樣的，但是貧窮的學生被逮捕的可能性是有錢學生的五倍。每當有人被逮捕時，系統都會嘗試預測他們是否會再次犯罪，也就是那些因為吸毒而被逮捕的人是否會在一年內再次因吸毒被逮捕。如果他們可能再次犯罪，將受到更嚴厲的懲罰；反之，則予以緩刑釋放。

由於有錢的學生不太可能被逮捕，因此系統會正確地預測他們不太可能被再次逮捕，並得出他們不太可能再次犯罪，故建議較輕的刑罰。貧窮的學生被再次逮捕可能性較高，因此系統將他們標記為可能的再次犯罪者，並建議嚴厲刑罰。

這是非常不公平的，吸毒傾向根本上沒有差異，但是因為系統在某個階段存在差異，並在下一階段以其為依據做出刑事判決而擴大了差異。

有學者認為，該算法不應該預測誰將重新入獄，而應該預測再犯的可能性。但是我們唯一可測量違反情形的代理變項(proxy variable，註 1)是入獄情形，而導致最終的偏誤。

如果該預測系統為用來預測再次犯罪情形，且在大型資料集上經過訓練的 AI，則稱為「AI 偏誤」。但我們很容易看出，AI 實際上並不是問題的關鍵。如果這個決定是由人類法官從他們多年的司法經驗中得出自己對於累犯直覺做出的，那也是不公平的。

「評估學生成績，並根據學校過去一年的表現進行調整」的方法也是「人工智慧」精神的體現。有些評論家認為英國的學校決策是 AI 偏誤的一個例子。但我們更可以將其歸類為「預測偏誤(prediction bias)」，也就是指當我們進行某些變項的預測時，最後會得出不平等的預測結果，並通常會受到種族、財富以及國籍等因素的深刻影響，而反歧視法律

的出現普遍禁止將上述因素納入考量，以避免所謂不公平的現象。

人工智慧只是我們用來預測的一種工具，但它們並不是唯一會導致上述缺失的方法或系統。例如，具有相同收入和債務的夫妻申請信用卡，卻獲得截然不同的信用額度，即為本地銀行業者做出的決定，而非依照複雜的公式結果。

特別是在刑事司法中，使用累犯資料進行訓練的算法會得出帶有種族差異的量刑建議，例如，不公正地要求將黑人囚禁比白人囚禁更長的時間。但是，當不使用此方法時，法官亦會根據量刑準則和個人直覺做出這些決定，也會產生種族差異。

### **預測結果產生的差異真的比實際結果的差異更不公平嗎？**

重要的是，不要使用不公正的系統或方法作為結果或政策的依據。執意這樣做的結果，最終將傷害過去曾受到懲罰的人，並將不平等深深刻在社會的記憶裡。值得思考的是，為什麼一個預測貧困學生考試表現較差的系統會比參加考試的結果產生更大的憤怒呢？因為也同時顯示貧困學生表現逐年惡化的情形。

對於學校考試成績等事件，不僅在預測結果方面存在差異，在實際結果亦存在不同：富裕的學生通常在考試中獲得更高的分數，原因有很多，包含就讀更好的學校、選擇更好的家庭教師、更多的學習時間等。如果學生真的如期參加考試，那麼富裕的學生與貧窮的學生在考試分數上將存在很大差距，並與預測的分數中有相似的結果。

當這樣的差距同時在預測中及現實生活中發現時，我們應正視問題的發生及解決之道，否則類似的問題將一再發生。

註 1：代理變項(proxy variable)：當一個變項無法被測量時，研究者通常會尋找一個可以被測量的相近變項，稱為「代理變項」。(https://mypaper.pchome.com.tw/readingstatistic/post/1321881499)