

2024 年教育大數據分析競賽初賽用開放資料的說明

一、開放資料宣告

- (一) 教育大數據分析計畫辦公室（以下簡稱計畫辦公室）提供之教育大數據分析競賽初賽用開放資料(以下簡稱初賽開放資料)，為教育大數據競賽的使用而設計。該數據依照國際評量資料庫資料處理原則產生，並經匿名處理，公開資料之分析結果可視為統計與測量上相等，但無法用於實務上的現象詮釋，故不適合運用於學術論文發表或其他任何形式之公開發表。若有教育大數據分析競賽以外之使用，請使用者自行評估風險，計畫辦公室不作任何保證，然使用時仍應註明出處。
- (二) 計畫辦公室已盡可能確保開放資料之準確性和完整性，但由於資料來源為原始資料庫的部分抽取，可能含有少數錯誤、遺漏或不確定性。使用開放資料，除法令另有規定外，由使用者自行承擔，計畫辦公室不對使用開放資料而引起的任何直接、間接、特殊或後果性之損害結果，承擔賠償或補償責任。
- (三) 計畫辦公室保留隨時刪除、修改及更正開放資料的權利，而不需提前通知。
- (四) 使用開放資料，因使用者故意或過失，導致資料提供機關遭受損害，或第三人因此向資料提供機關請求損害賠償，使用者應對各機關負擔賠償責任。
- (五) 使用初賽開放資料即表示您理解並同意遵守上述開放資料宣告內容，前揭開放資料宣告內容，如有未盡事宜，計畫辦公室保留解釋權利，並以中華民國法律為準據法。

二、資料背景簡介

本次教育大數據分析競賽初賽開放資料是使用者跨平臺學習時各類操作的紀錄檔，每一筆資料記載使用者於學習平臺上之學習紀錄。由於此資料檔經匿名處理，並依照國際評量資料庫資料處理原則產生，基於此資料檔的分析結果與原始資料檔結果可視為統計與測量上相等。唯此資料檔的結果無法用於實務上的現象詮釋，僅供教育大數據分析競賽使用。

本次教育大數據分析競賽初賽用開放資料包含使用者基本資料與其在 3 個不同平臺（dp001~dp003）的學習資料，分別以 8 個檔案作呈現。詳細欄位說明請參照下表。

本次教育大數據分析競賽初賽用開放資料為跨平臺資料，請參賽者在使用這份資料進行數據分析並撰寫提案構想書時。本次提供的開放資料為 UTF-8 編碼的 CSV 格式檔案，此資料經由 Python、R、SPSS、Excel 等軟體測試，確認能正確讀取。

三、資料格式與欄位說明

1.使用者基本資料(user_data)

欄位名稱	欄位資料類型	欄位說明
user_sn	number	使用者流水號
organization_id	number	學校代碼
grade	number	使用者年級
class	number	使用者班級
seat	number	使用者座號
chinese_score	number	使用者在國文測驗的分數
math_score	number	使用者在數學測驗的分數
english_score	number	使用者在英文測驗的分數

※ 使用者基本資料表中的學校代碼、年級、班級與座號皆經過去識別化處理，已非使用者原始資料。

2.使用者在學習平臺 dp001 的學習資料

2-1.使用者在學習平臺 dp001 的練習題作答資料(dp001_prac)

欄位名稱	欄位資料類型	欄位說明
prac_sn	number	練習題作答流水號
user_sn	number	使用者流水號
date	datetime	進行練習題的時間戳記 yyyy-mm-dd hh:mm:ss
during_time	number	進行練習題的作答時間(單位是秒)
score_rate	number	練習題正確率
binary_res	string	練習題中每一小題的作答結果(1：答對，0：答錯)，以@XX@作為分隔符號
items_ans_time	string	練習題中每一小題的作答時間，以@XX@作為分隔符號(單位是毫秒)
indicator_name	string	能力指標名稱
subject_name	string	科目名稱

2-2.使用者在學習平臺 dp001 影片檢核點作答資料(dp001_exam)

欄位名稱	欄位資料類型	欄位說明
exam_sn	number	影片檢核點測驗記錄流水號
review_sn	number	影片瀏覽流水號
question_sn	number	檢核點試題流水號
question_timestamp	decimal	影片中檢核點出現的時間點(秒)，該欄位會隨著影片教材內容的修改而有所異動，時間點會以最後教材更新時間為主
binary_res	string	檢核點作答結果(1：答對，0：答錯)
finish_time	string	檢核點測驗記錄完成時間的時間戳記 yyyy-mm-dd hh:mm:ss
ans_time	number	檢核點出現後，使用者回答問題花費的時間(單位是秒)

2-3.使用者在學習平臺 dp001 影片瀏覽的資料(dp001_review)

欄位名稱	欄位資料類型	欄位說明
review_sn	number	影片瀏覽流水號
user_sn	number	使用者流水號
indicator_name	string	能力指標名稱
subject_name	string	科目名稱
video_name	string	影片名稱
video_len	decimal	影片長度(單位是秒)
start_timestamp	decimal	影片瀏覽開始的秒數
start_time	datetime	影片瀏覽開始的時間戳記 yyyy-mm-dd hh:mm:ss
end_timestamp	decimal	影片瀏覽結束的秒數
end_time	datetime	影片瀏覽結束的時間戳記 yyyy-mm-dd hh:mm:ss
finish_rate	number	影片瀏覽的進度(百分比)

2-4.使用者在學習平臺 dp001 影片瀏覽時的行為資料紀錄(dp001_review_plus)

欄位名稱	欄位資料類型	欄位說明
review_plus_sn	number	影片操作的行為流水號
review_sn	number	影片瀏覽流水號
view_time	datetime	執行影片操作的時間戳記 yyyy-mm-dd hh:mm:ss
view_action	string	影片操作的行為名稱。包含：play-播放、paused-暫停、end-結束、normal-恢復正常播放速度、slowdown-降低播放速度、speedup-提升播放速度、chkptstart-檢核點作答開始、chkptend-檢核點作答結束、continue-接續上次未看完的進度、note-停下做筆記、dragleft-影片倒轉、dragright-影片快轉、dragstart-影片開始拖曳、 <u>browse-進入影片瀏覽頁面</u> 、replay-重看一次、 <u>fuscreenoff-結束影片全螢幕放大</u> 、 <u>fuscreenon-開始影片全螢幕放大</u> 、review-檢核點作答完畢後，回到上一個檢核點開始再看一次，若是在第一個檢核點則從從頭開始播放、question-提問
timestamp	decimal	動作結束在影片的第幾秒
turbo	decimal	影片調整播放速度，只記錄調整播放速度的當下，並非紀錄當前的播放速度 (0.5：慢速、1x：正常速度、1.5x：加速)

3.使用者在學習平臺 dp002 上的學習資料(dp002_exam)

欄位名稱	欄位資料類型	欄位說明
sn	number	學習資料行為歷程資料操作紀錄流水號
user_sn	number	使用者流水號
action_time	datetime	使用者開始進行測驗的時間戳記 yyyy-mm-ddThh:mm:ss+08:00
object_type	string	測驗類型
result_duration	string	使用者進行該測驗所花費的時間(欄位使用 <u>ISO 8601</u> 的時間格式，以 PTO OMO OS 的格式呈現時間，例如 PT1M35S 代表 1 分 35 秒)
result_success	string	測驗題目是否答對(true：答對，false：答錯)
mission_id	number	測驗流水號
question_id	number	測驗題目的流水號

4.使用者在學習平臺 dp003 上的學習資料

4-1 使用者在 dp003 上的英文單字學習資料(dp003 word)

欄位名稱	欄位資料類型	欄位說明
sn	number	資料流水號
user_sn	number	使用者流水號
section_id	number	單元流水號
start_timestamp	datetime	開始執行操作的時間戳記 yyyy-mm-dd hh:mm
finish_timestamp	datetime	結束執行操作的時間戳記 yyyy-mm-dd hh:mm
total_millisecond	number	使用時間(單位是毫秒)
last_modified	number	資料最後更新時間(Unix Timestamp 格式，單位是毫秒)，例如當 last_modified 是 1717159989812 時，轉換後為 2024-05-31 20:53:09.812，表示距離從 1970-01-01 00:00:00 開始的時間差
game_sn	string	遊戲闖關編號
progress_seq	number	當前遊戲闖關題號
is_correct	number	該次作答時，對目標單字的作答結果是否答對(0:答錯，1:答對)
target_vocabulary	string	學習目標單字
wrong_times	number	對該目標單字(target_vocabulary 欄位)重複作答時，系統紀錄之累積答錯次數
correct_times	number	對該目標單字(target_vocabulary 欄位)重複作答時，系統紀錄之累積答對次數

4-2 使用者在 dp003 上的數學學習資料(dp003_math)

欄位名稱	欄位資料類型	欄位說明
sn	number	資料流水號
user_sn	number	使用者流水號
game_grade	number	系統建議適用年級，表示此遊戲適合該年級的使用者
game_semester	number	系統建議適用學年度，表示此遊戲適合該學年度的使用者
unit_id	number	單元流水號
answer_problem_num	string	回答問題的流水號
is_correct	number	回答問題是否正確(0:答錯，1:答對)
<u>game_time</u>	number	遊戲闖關時間(單位是毫秒)
last_modified	number	資料最後更新時間(Unix Timestamp 格式，單位是毫秒)，例如當 last_modified 是 1717159989812 時，轉換後為 2024-05-31 20:53:09.812，表示距離從 1970-01-01 00:00:00 開始的時間差
unit_name	string	遊戲所屬的單元名稱

ER-Model 圖

