

2022 全國智慧製造大數據分析競賽

初賽說明會 Q&A

(一) 初賽進行相關問題

Q1： 請問初賽有規定不能使用的套件嗎？

A1： 初賽數據開放參賽者下載至自己的設備進行分析，未限制任何分析方法、演算法、軟體或硬體等，唯需於初賽繳交文件內說明團隊解題過程。

Q2： RMSE 值會影響最後是否能進決賽，那其他的說明文件會影響進決賽的資格嗎？

A2： RMSE 值與初賽繳交的相關文件(包含「測驗報告」、「測試結果」與「完整程式碼與安裝、執行方式說明文件」)，都將作為競賽評審團進行評分與驗證之依據。

Q3： 請問可以提供去年決賽得獎的 RMSE 值作為參考嗎？

A3： 因每年初賽與決賽的題目、數據與評分方式皆不相同，去年決賽的評分標準對今年初賽無影響，故主辦單位不提供去年決賽的 RMSE 值。

Q4： 請問想看這次說明會的錄影檔，可以到哪裡看？

A4： 本場說明會的簡報、錄影檔及相關資料，皆會公告在競賽網站的最新消息(<https://imbd2022.thu.edu.tw/#news>)。

(二) 國網中心初賽平台相關問題

Q1： 我已經註冊 iService 註冊帳號，平台上顯示的金額只有 200 元額度可以使用嗎？

A1： 註冊 iService 帳號後，國網中心會贈送 200 元的額度作為體驗各項計算資源使用，但本競賽會將已註冊的 E-mail 加入到指定計畫內，不會占用贈送的 200 元額度。

此次競賽提供每隊 1 張 GPU、1 個開發型容器，使用期限至初賽結束(9 月 15 日 23:59)止，請各團隊在初賽結束前自行備份需留存的檔案，並關閉容器，9 月 15 日截止後國網中心將自動刪除所有資料。

(三) 初賽題目與數據相關問題

Q1： 初賽訓練數據中有出現 0 或'0 或 nan 或空格，代表什麼意思呢？

A1： 假設總共有 15 個工序要進行加工，但這項工件只需要經過其中 3 個工序，則只會有其中 3 個工序有數值，另外 12 個沒有動作的工序則會是 0 或'0 或 nan 或空格的型態。

Q2： 請問初賽訓練數據中 nan 或空格適合補值嗎？

A2： 初賽訓練數據中 nan 或空格表示感測器沒有動作，是否要補值或如何處理，請參賽團隊自行判斷。

Q3： 請問會提供初賽訓練數據的單位嗎？例如公分、毫公分等等

A3： 因初賽訓練數據的單位都相同，不影響解題，故主辦單位不提供初賽訓練數據的單位。

Q4： 請問初賽訓練數據有時序嗎？

A4： 初賽訓練數據沒有時序、前後順序關係，每一筆資料都是獨立的。

Q5： 請問初賽訓練數據都是數值型態嗎？

A5： 是，所有初賽訓練數據都是數值型態。

Q6： 有些初賽訓練數據是紅字或黃底，請問這些標示有特別的意義嗎？

A6： 初賽訓練數據標示紅字或黃底的數據沒有特別的意義。

Q7： 請問初賽訓練數據(train.xlsx)的「工作表 1」有特別的意義嗎？

A7： 初賽訓練數據(train.xlsx)的「工作表 1」沒有特別的意義。

Q8： 請問初賽訓練數據(train.xlsx)的欄位 A~F 的膜厚是加工前的膜厚度嗎？

A8： 不是，初賽訓練數據(train.xlsx)的欄位 A~F 的膜厚是經過欄位 G~EA 加工後的膜厚。

Q9： 請問會提供各個欄位名稱的中英文對照嗎？

A9： 初賽訓練數據的欄位名稱為數據提供廠商的商業機密，主辦單位無法再提供欄位名稱的中英文對照。

Q10： 題目中提到的四個製程是什麼意思？

A10： 四個製程表示四種不同加工的方法。

Q11：請問訓練數據與測驗數據的四個製程有順序嗎？

A11：訓練數據與測驗數據的製程都有順序，主辦單位提供的訓練數據與測驗數據都已經照製程順序排序。

Q12：請問訓練數據來源是哪些製程或加工方法？

A12：此為數據提供廠商的商業機密，主辦單位無法再提供數據來源的製程或加工方法。

Q13：請問訓練數據 6 個 output 是同時測量的嗎？

A13：訓練數據的 6 個 output 為同時量測的膜厚值。

Q14：請問訓練數據每一筆資料都是一個工件的感測資料嗎？

A14：訓練數據的每一筆資料都是同一個工件的感測資料，每一個工件原則上製程都是同一天完成，但是資料的前後筆可能不是同一天。

Q15：請問訓練數據在同一個感測器都沒有資料，代表生產同一個產品嗎？

A15：不一定，有可能某一個設備故障，啟用另一個設備代替生產。

Q16：請問同一個工件都使用同一個設備加工嗎？

A16：不一定，不同天加工的工件，因為設備會輪休，所以每次加工的設備可能會不一樣。

Q17：請問訓練數據是紀錄幾種類型的工件？例如紀錄 3 種或 5 種。

A17：訓練數據沒有特別統計總共有幾種類型的工件。

Q18：不同工件有指定加工膜厚嗎？工件與膜厚有什麼關聯？

A18：膜厚跟工件沒有直接相關性，膜厚需要從中間的感測器資料推斷，經過四個製程後能不能達到指定的膜厚。

Q19：我要如何確認膜厚的不同是因為加工的瑕疵，還是客戶要求的膜厚不同？

A19：產品會依要求不同而有不同的膜厚，須從感測器上的數據推測出膜厚的結果。

Q20：請問訓練數據和測驗數據切出的方法是從同一個資料庫 random sample 嗎？

A20：是。