

109 Data Mining Project 01

I. 使用的 Feature:

「Info」：一個 row 代表一位患者。特徵「共病症_Comorbidities」意義:一個數字代表擁有一種共病症，標示「0」則表示沒有任何共病症。

「TPR」：「溫度、脈搏、呼吸速率、收縮壓、舒張壓」的時間序列資料，同樣的「No」代表是同一位患者的資料。

若你想將「Info」、「TPR」兩個表單整合，可利用「No」作為合併依據，但請特別注意，「No」是患者編號識別碼，切勿將其當作 Feature 一併訓練。

II. 繳交檔案名稱與格式:

1. <學號>.csv

- 此份檔案是用來存放 Test 檔預測的結果
- 「Submission.csv」於上傳前，需重新命名為「學號.csv」
EX: 0760406.csv
- 「Target」欄位只能填入 0 或是 1，沒有其他選項
- 整份 CSV 檔只能有這兩欄，不能有其他多餘的欄位

No	Target
13300244	
13305420	
12751484	
13921720	
11926643	
13525088	
5904478	
6484236	
3428878	
3783822	
4475311	
4134201	
5085853	
2886400	
4708170	
2854675	
3722892	
3226138	
3654704	

No	Target
13300244	0
13305420	0
12751484	0
13921720	0
11926643	0
13525088	0
5904478	0
6484236	0
3428878	0
3783822	0
4475311	0
4134201	0
5085853	0
2886400	0
4708170	0
2854675	0
3722892	0
3226138	0
3654704	0

2. <學號>_Report.pdf

- 檔名: <學號>_Report.pdf
EX: 0760406_Report.pdf
- 字體大小: 12
- 字型: 英文(Times New Roman), 中文(標楷體)
- 內文: 必須要有「資料前處理」、「模型建製」、「驗證方法」
其餘不限，若有必要可以自行增加。

III. 上傳內容:

- 請上傳 zip 至 e3 作業繳交區

解壓縮後只能有兩個檔
千萬不要放資料夾包起來



IV. 扣分項目

扣分項目	Project 總分
檔案名稱錯誤	-20%
答案繳交格式錯誤	-20%
遲交 1 天	-20%

Item	學期總分
F1-score (base on Target=1)	10%
Project Report	10%

$$*F1 - score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} = \frac{2TP}{(2TP + FP + FN)}$$

TP: True Postive

TN: True Negative

FP: False Postive

V. 截止日期

- 2020/11/11(三) 零晨 12:00 整，助教會上傳 Test 檔
- 2020/11/13(五) 晚上 11:50 作業上傳截止