

109 Data Mining Project 1

I. Feature:

「Info」: 1 row represents a patient. 「共病症_Comorbidities」: A number indicates that the patient has a comorbid condition, and the label "0" means that there is no comorbid condition.

「TPR」: Time series data of "temperature, pulse, breathing rate, systolic blood pressure, diastolic blood pressure". The same "No" means the data of the same patient.

If you want to merge the two sheets "Info" and "TPR", you can use "No" as the merge index. Please note that "No" is the patient ID and you should **not** train it as a Feature otherwise your Model will learn the wrong information.

II. Submitted file name and format:

1. <your-student ID>.csv

- This file is your predict result for test data.
- Please rename your "Submission.csv" to "<your-student ID>.csv"
EX: 0760406.csv
- 「Target」 column only can be 0 or 1.
- The entire CSV file can only have two columns.

	A	B	C
1	No	Target	
2	13300244		
3	13305420		
4	12751484		
5	13921720		
6	11926643		
7	13525088		
8	5904478		
9	6484236		
10	3428878		
11	3783822		
12	4475311		
13	4134201		
14	5085853		
15	2886400		
16	4708170		
17	2854675		
18	3722892		
19	3226138		
20	3654704		

	A	B	C
1	No	Target	
2	13300244	0	
3	13305420	0	
4	12751484	0	
5	13921720	0	
6	11926643	0	
7	13525088	0	
8	5904478	0	
9	6484236	0	
10	3428878	0	
11	3783822	0	
12	4475311	0	
13	4134201	0	
14	5085853	0	
15	2886400	0	
16	4708170	0	
17	2854675	0	
18	3722892	0	
19	3226138	0	
20	3654704	0	

2. <your-student ID>_Report.pdf

- File name: <your-student ID>_Report.pdf
- EX: 0760406_Report.pdf
- Size: 12
- Font : Times New Roman
- Include:

The content must include "Data preprocessing", "Formula", "Validation method".
Can be increased if necessary.

III. Submission:

- Please upload the zip to e-Campus system "Project_01".

There can only be two files after decompression
Never put in any extra folders



IV. Rule

Items that will lose points	Project Score
Any wrong file name	-20%
Any format error	-20%
Late project per day	-20%

Item	Total score for the entire semester
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. Deadline

- 2020/11/11 (Wed.) AM 12:00 TA will upload Test file
- 2020/11/13 (Fri.) PM 11:50 upload deadline.