109 Data Mining Project 1

I. Feature:

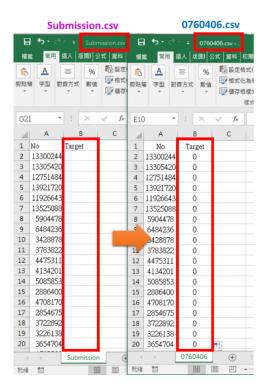
「Info」: 1 row represents a patient. 「共病症_Comorbidities」: A number indicates that the patient has a comorbid condition, and the label "0" means that there is no comorbid condition.

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.



2. <your-student ID> Report.pdf

File name: <your-student ID> Report.pdf

- EX: 0760406 Report.pdf

- Size: 12

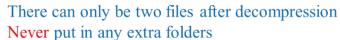
- Font: Times New Roman

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".





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 Negetive 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.