1. Github link: https://github.com/dingkk123/NYCU2023-machine-learning

2.Reference: TPS-Aug22 9th solution | Kaggle

### 3.Brief introduction:

我是用 logistic regression 來當作我的 model,還有使用 k-fold 來切割資料。

### 4. Methodology:

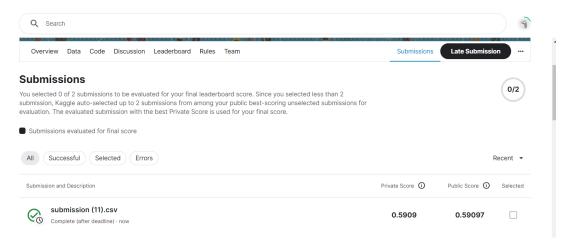
Data-preprocessing:我使用 HuberRegressor,還有 KNNImputer 來處理資料中有異常部分,因為他的 input 有一些缺失值,需要用這兩個方法來補差值。

Model architecture:在進行 logistic regression 前,會先用 standardscalar 將資料 進行預先處理,並且整個操作分成兩種,我這兩種分別 select 了兩種不同的 feature 來進行 train,並且將 feature 存起來。

# 5.Summary:

這次的作業蠻難的,需要花一定的時間才能將 model 決定出來,並且需要去處理 input 的 data 來提高 private score,然後需要去 discussion 去看其他人的一些方法來了解怎麼讓 score 提高,但是可以透過這次的作業來更了解機器學習的一些實作。

### 6.Result screenshot:



# 7.feature\_list\_link:

https://drive.google.com/drive/folders/1HaVjuY9tNaVWdhswES8mO5QCzx4QRYQo?usp=share link