Data Mining mini project

說明:

經過兩次的實作作業，相信同學對於分類器、評估方法有一定程度的認識了，因此這個mini project要讓同學練習實際運用在資料集上，並進行預測。
本次要使用的資料集為: **Pima Indians Diabetes Database**

- Kaggle Competition: https://www.kaggle.com/competitions/diabetes-classification (https://www.kaggle.com/competitions/diabetes-classification)(需註冊 kaggle，並將預測結果上傳至此competition，一天上傳次數上限為 10 次，請同學盡早開始做)

繳交內容:

1. **一份報告(.pdf)，包含:**
   (1) 資料集特徵資料說明、屬性特性說明
   (2) 對特徵做甚麼樣的分析? 哪些前處理? 採用哪些特徵? 原因?
   (3) 基於什麼理由選擇哪個分類器?
   (4) 採用的評估指標結果與觀察
   (5) 將預測結果上傳至 kaggle 並截圖測試的分數
   (在 leaderboard 上的整列截圖，如下圖)

2. **程式碼(.py)、上傳到 kaggle 的預測檔案(.csv)**
   一同壓縮至一個資料夾(組別_學號姓名.zip)上傳至 moodle

注意事項:

1. 可參考網路上相關資源，但說明所採用前處理、分類器的原因，請勿直接抄網路上的程式
2. 評分以報告內容是否說明清楚、合理、以及發現為依據，並非完全以Kaggle 的測試分數為主。

繳交期限 12/09(六) 23:59 PM