----------------------------------------------------------------------------------------------------------------------

期末作業提供兩個公開數據集，分別為

Dataset\_2 - a dataset for heart attack classification

Dataset\_1 - a dataset for diabetes classification

選擇一個資料集，根據這學期學習的各項機器學習知識以及範例程式，設定不同項目進行分類模型效能探討，例如：

1. 不同特徵擷取技術
2. 不同學習演算法
3. 學習演算法參數最佳化
4. 集成學習 - majority-voting classifier, bagging, boosting, …
5. …

注意:

訓練集以及測試集之分割為測試集佔全部數據集的25%，並且random\_state設定為1，即

X\_train, X\_test, y\_train, y\_test = \

train\_test\_split(X, y, test\_size=0.25, random\_state=1)

不論自訂對於哪些項目進行討論，最終一定要明確說出你所使用的是哪一個數據集

並且呈現針對上述測試集，你所找到的最佳分類模型，包含數據前處理、學習演算法及其參數設定，以及最終分類效能的指標(accuracy, precision, recal, andl confusion matrix)等結果