

# 2019 ECT 作業四

一、 請用 python 依照步驟對 Titanic.csv 進行前處理整理出 Weka 可執行

logistic 的資料，順便準備 python 的前處理(40%)：

請將各步驟程式碼截圖並簡單說明

- (a). 利用 np.nanmedian 函數計算資料集中 Age 的中位數
- (b). 利用 np.where 將 dataframe 中空值的 Age 替換成上述的中位數
- (c). 運用 drop 函數刪除 Cabin 欄位
- (d). 運用 drop 函數刪除 PassengerId 欄位
- (e). 運用 dropna() 函數去掉資料的空值
- (f). 將 name 轉為 encoded 的資料
- (g). 運用 replace 將原資料中 Survived 欄位中的[0,1]轉為  
[No,Yes](Python 部分不用)
- (h). 運用 to\_csv 另存為 TitanicClean.csv

Python 前處理

- (i). 除上述步驟外將 Name、Sex、Ticket、Embarked 進行 encoded
- (j). 將  
  
'Pclass','Name','Sex','Age','SibSp','Parch','Ticket','Fare','Embarked'  
  
切為 feature
- (k). 將 Survived 切為 Target

二、 請用 python 對 Titanic.csv, Weka 對第一大題整理的

TitanicClean.csv 進行 logistic regression 的分類。

1. 請用 python 對 Titanic.csv 進行資料集 66%及訓練集 34%切分, 並運用 linear\_model.LogisticRegression (solver='liblinear')進行訓練, 並運用 LogisticRegression 函式庫中的 score 印出模型準確度。請將程式碼及準確度一同截圖。(30%)
  2. 請用 WEKA 對 TitanicClean.csv 同上等分切分資料集進行 LogisticRegression 訓練, 同時在 Classifier output 中找出兩個與分類結果為 No 正相關的因素。截圖並附上過程、答案及準確率。(30%)
- 繳交期限： 3/27 (三) 中午 12:00
  - 請轉檔為 PDF 格式, 檔名為：ECT\_HW4\_學號\_版本.pdf, 並同時附上 python 的 ipynb.檔,命名格式同上
  - 上傳至 LMS 作業區,遲交一天扣該次作業成績 5%
  - 補交請上傳至 HW4 補交區, 僅開放到 4/3(三)中午 12:00