

# 2020 ECT 作業五

1. 請用 python 依照步驟對 BreastCancer.csv 進行 KMeans 分析，過程中對所有重要程式步驟進行截圖並加以說明。(60%)
    - (a) 將 radius\_mean 及 area\_mean 切為 feature，diagnosis 切為 target
    - (b) 用 cluster.Mean 設 n\_clusters=2
    - (c) 用 fit\_predict 對 feature 進行分類
    - (d) 運用 matplotlib 中的 scatter 圖，x 軸設為 radius\_mean，y 軸設為 area\_mean，c 設為分群結果，印出分類圖形
    - (e) 移除 area\_mean 中大於 2000 的資料
    - (f) 重複上述動作印出分類圖形
  2. 請用 weka 對 BreastCancer.csv，進行 simplekMeans 分析，Cluster mode 設為 Use training set，numClusters 設為 2。(40%)
    - (a) 印出 KMeans 分類結果(兩類的數目)
    - (b) 運用 Visualize cluster 得到與上述 python 印出之圖形(未進行刪減前)類似的結果
- 繳交期限: 4/22(三) 中午 12:00
  - 請轉檔為 PDF 格式, 檔名為:ECT\_HW5\_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 **HW5 補交區**,僅開放到 **4/29(三) 中午 12:00**