

# 2020 ECT 作業七

1. 利用 Weka 對 DeerHunter.arff 進行前處理及 RandomForest，依序完成

以下步驟：

- (a) ReplaceMissingValue，列出補上的值為何？(10%)
- (b) 以 LinearRegression 對 instance 進行 Outlier Detection&Remove (10%)
- (c) 進行 RandomForest Classification，Test options 設為 10-Folds Cross-validation，目標為 “yes” (10%)
- (d) Attribute Selection，以 CfsSubsetEval 及 BestFirst 篩選出適合的屬性 (10%)
- (e) 重複步驟 c 比較篩選屬性前後的差異 (10%)

2. 利用 Python 對 IRIS 及 BreastCancer 資料集進行 PCA 降維，依序完成

以下步驟：

- (a) 利用 sklearn 載入 IRIS 資料集，切分 Attribute 及 Target，挑選三個屬性印出其 3D 分布圖 (10%)
- (b) 用 PCA 對 Attribute 進行降維 n\_components 設為 2 (10%)
- (c) 印出 IRIS 降維後的分布圖 (10%)
- (d) 對 BreastCancer 資料集進行 PCA 降維 n\_components 設為 2 (10%)

(e) 印出 BreastCancer 降維後的分布圖 (10%)

- 繳交期限: 5/27(三) 中午 12:00
- 請轉檔為 PDF 格式, 檔名為: ECT\_HW7\_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 HW7 補交區, 僅開放到 6/3(三) 中午 12:00