

2019 ECT 作業二

1. 用 Weka 軟體對 contact-lenses.arff 建立 J48 決策樹，選擇 “Use training set” ，設定 Attribute: contact-lenses 為 Output，在過程中對重要步驟截圖並加以說明，並回答以下問題：
 - (a) 在前處理部分，右下角選擇不同屬性作為 Class，請解釋長條圖的數量、上方的數字以及不同顏色意義為何?(15%)
 - (b) 使用 Visualize Tree 或 Classifier Output 列出三個 Classification Rule 並解釋。(20%)
2. 請利用 weka 和 python 對 glass.csv 進行 Supervised learning 中的 DecisionTree 分析 ,並回答以下問題:
 - (a) 請運用 python 的 train_test_split 對 glass.csv 資料集，預測目標屬性為 Type 進行訓練集(66%)、測試集(34%)切分，請將重要程式碼截圖並說明(10%)
 - (b) 請利用參數(criterion = ‘entropy’, max_depth=3, max_leaf_nodes = 4)對切分出的訓練集進行訓練，並用 metrics.accuracy_score()分別計算出模型對於訓練集和測試集的精準度，並與 WEKA 設定演算法 J48 Percentage spilt 66%跑出的結果截圖說明並一起呈現比較。(20%)
 - (c) 請利用 graphviz 套件跑出決策樹圖形截圖並加以說明當中 X[?] 、 samples、及 value 各代表的訊息，另外運用 WEKA visualize tree 觀察決

策樹圖形並說明 WEKA 中葉節點的標籤及括號內的數字代表的意義。

(20%)

(d) 請試著調整 DecisionTreeClassifier 的參數，提升模型準確率，請截圖並

附上每次測試的結果(3~5 次)，觀察並說明準確率上升的原因，另外說明

模型在訓練集的準確度通常較測試集的準確度高的原因為何(15%)

- 繳交期限： 3/13 (三) 中午 12:00
- 請轉檔為 PDF 格式，檔名為：ECT_HW2_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 HW2 補交區，僅開放到 3/20(三)中午 12:00