作業二

壹、 請使用 weka 完成以下題目，並截圖結果附上適當說明，以 PDF 文件呈現：

1. 載入 dataset.arff，將 year、iso_code、rank 欄位刪除 (5%)
2. 使用 ReplaceMissingValues 將全部欄位的空值以各欄位平均數填入 (5%)
3. 將 freedom 與 freedom_own_foreign_currency 轉為 Nominal，並說明為何 Numeric 無法使用在 Decision tree (5%)
4. 以 70% 切割訓練資料，使用 J48 對 freedom 進行分類，並截圖分類準確率、混淆矩陣及視覺化的 Decision tree (10%)

貳、 請使用 python 完成以下題目，並在文字框附上適當註解，以 ipynb 檔繳交：

1. 以 DataFrame 格式載入 dataset.csv (5%)
2. 請檢查並列出 train.csv 中每個欄位的空值個數 (5%)
3. 由於年份久遠的數據統計不夠完整，請把在 2004 年之前的數據刪除 (5%)
4. 將 eco_freedom 欄位是空值的資料刪除 (5%)
5. 將 freedom_own_foreign_currency 欄位轉為數字型態 (例如：HIGH=2，NORMAL=1，LOW=0) (5%)
6. 將 eco_freedom 欄位重製為 freedom (eco_freedom 的分數小於六屬於不自由，non-freedom = 0，freedom = 1) (5%)

7. 將全部欄位的空值都由此欄位的平均值填入(10%)

8. 請以 year、iso_code、countries、rank、freedom 以外的欄位作為訓練資料，建立 Decision tree 來預測 freedom，將訓練資料比例設為 50%，random_state 設為 12，max_leaf 設為 5，stratify = y，並繪出 Decision tree 的樹狀圖 (10%)

9. 計算出在 8. 測試資料上的平均準確率 (5%)

10. 請用 8. 的結果評估決策樹好壞(使用 classification_report)產生類似以下結果 (5%)

	precision	recall	f1-score	support
non-freedom	0.81	0.69	0.75	101
freedom	0.93	0.96	0.95	440
accuracy			0.91	541
macro avg	0.87	0.83	0.85	541
weighted avg	0.91	0.91	0.91	541

11. 請分別以訓練資料比例 60%、70%、80%、90% 建立 Decision tree，random_state 皆設為 12，並將不同資料比例與平均準確率的比較結果以 DataFrame 呈現，如右圖所示。(10%)

	split_proportion	score
0	50/50	0.940067
1	60/40	0.937587
2	70/30	0.913124
3	80/20	0.925208
4	90/10	0.895028

12. 呈上題，將此比較結果以折線圖呈現，如下圖所示：(5%)

繳交期限：3/14　23:59

第一題請繳交.PDF 檔，檔名為 ECT_HW2_學號.pdf，請適當附文字說明。

第二題請繳交.ipynb 檔與整理後的 csv 檔，檔名為 ECT_HW2_學號.ipynb 與 ECT_HW2_學號.csv，程式中請適當附上註解

遲交一天扣該次作業 5%（最多扣 50%）