Data Science HW1

B10815057 廖聖郝

1. 執行程式:
2. pip install pandas
3. pip install sklearn
4. python judge\_rain\_tomorrow.py
5. 程式架構流程:
6. 讀取檔案，包括訓練與測試資料



1. 資料預處理(preprocess函式內)
   * 1. 把不是數字的資料建立虛擬數值



* + 1. 填補空缺資料



* + 1. 日期欄位只取月份(因年份或幾號影響不大)



* + 1. 把有NAN的row都丟掉



* + 1. 將 Yes No 轉為 0 1，因測試資料無Attribute23，所以要先判斷有沒有(hasattr)再做修改



* + 1. 風向欄位做 one hot encoding



* + 1. 將訓練與測試資料都執行預處理



1. 平衡資料(resample\_data函式內)

由於下雨跟不下雨的次數相差過大，用resample將資料平衡，然後把資料打亂，得到更好的效果



1. 建立模型

分析模型採用神經網路模組: MLPClassifier(多層感知分類器)



1. 預測



1. 輸出結果到csv檔

