

# Big Data Analytics HW1

Kaggle 資料及說明 (佔 50%) :

目標：利用每日天氣觀測樣本做訓練，給入當天的觀測數據，希望能預測隔天會不會降雨

Attribute1 當天日期

Attribute2 氣象站的地區

Attribute3 最低溫度(攝氏)

Attribute4 最高溫度(攝氏)

Attribute5 降雨量 (單位: 毫米)

Attribute6 蒸發量

Attribute7 當天陽光出現的時數

Attribute8 最強陣風方向

Attribute9 最強陣風速度 (單位: 公里/小時)

Attribute10 早上九點的風向

Attribute11 下午三點的風向

Attribute12 早上九點前的平均風速 (單位: 公里/小時)

Attribute13 下午三點前的平均風速 (單位: 公里/小時)

Attribute14 早上九點的相對濕度

Attribute15 下午三點的相對濕度

Attribute16 早上九點前的平均大氣壓(hpa)

Attribute17 下午三點前的平均大氣壓(hpa)

Attribute18 早上九點，雲層遮蓋天空的比例 (0 完全晴朗無雲 - 8 完全多雲)

Attribute19 下午三點，雲層遮蓋天空的比例 (0 完全晴朗無雲 - 8 完全多雲)

Attribute20 早上九點的溫度

Attribute21 下午三點的溫度

Attribute22 今天有沒有下雨

Attribute23 明天會不會下雨(Label)

分為 Training Data(要訓練的資料) 及 Test Data(欲預測資料)

作法：

將 Training Data 透過 Model 訓練後，將 Testing Data 放入訓練好的 Model 中，去預測出 Output Attribute 為 0 or 1 (No or Yes)，Submission 的文件會長得像下面這樣：

id	ans
0	1
1	0
2	0
3	1
4	1
5	1
6	1

上傳檔案格式：共有 790 列(id 為 0 到 789)，共有 2 行，第一欄為 id (表示資料順序，從 0 開始)，第二欄為 ans (表示明天是否會下雨，1 代表 Yes，0 代表 No)

書面報告的部分 (佔 50%)：格式不限但至少要有以下兩點

- 請說明如何執行程式，並附上程式碼檔案
- 簡介你所用的程式架構及演算法流程，說明怎麼實作的