**巨量資料分析與應用 期末報告**

班級：商資三甲

分組編號：第 16 組

成員：學號 C109193104 姓名 鄭怡芳（組長）

學號 C109193110 姓名 張珈薰（組員）

題目：桃園市珍貴樹木數據分類

資料集名稱：桃園市珍貴樹木資料2017

資料集來源：https://data.gov.tw/dataset/26129

資料集欄位描述：欄位名稱、欄位描述

使用分析方法：多元分類（Multiclass Classification）。

描述程式碼動作：

1. 先進入Ubuntu作業系統，並啟動hadoop。
2. 下載桃園市珍貴樹木資料2017資料集，並上傳至hdfs。
3. 進入spark環境並匯入程式庫。

**多元分類-桃園市珍貴樹木資料**

⚫讀入資料檔

|  |  |
| --- | --- |
| **指令** | **說明** |
| **進入Ubuntu環境** | |
| start-all.sh | 啟動hadoop，若已經啟動，則不需要再執行此指令 |
| cd ~ | 回到家目錄 |
| wget  https://data.gov.tw/dataset/26129 | 下載鳶尾花資料集 |
| hdfs dfs -mkdir /tree | 在hdfs上建一目錄 |
| hdfs dfs -put ~/Taoyuan\_City\_Precious\_Tree\_Data\_2017.csv/tree | 將資料集上傳hdfs |
| hdfs dfs -ls /tree | 查看hdfs資料 |
| **進入spark-shell環境** | |
| spark-shell | 進入spark-shell |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |