**商業智慧與資料探勘**

**(Business Intelligent and Data Mining)**

**112 學年度第一學期 期中考試**

1. 請參考附件資料檔案。該檔案中之資料為：一個網路上評價各式各樣新車和中古車是否值得購買的資料檔，其中共有七個變數(buying\_price, maint\_price, persons, doors, lug\_boot size, safety, Suggestion)。各變數之定義及屬性值說明如下：

|  |  |  |
| --- | --- | --- |
| **Variable** | **Name** | **Definition** |
| X1 | buying price | 購買價格 |
| X2 | maint price | 維修保養花費 |
| X3 | Persons | 可搭乘人數 |
| X4 | doors | 車門數 |
| X5 | lug\_boot size | 行李箱的大小 |
| X6 | safety | 安全性 |
| Y | Suggestion | 是否值得購買 |

請根據資料，進行以下的分析議題(可以自行擴充)，並提供其分析機制(分析的邏輯及分析的流程跟結果)

Hint: 你可以參考下面的議題去思考!! (只是參考，請勿受限)

1. 幫忙根據(X1到X3)變數選擇適當的分群方法進行分群(3群 or 4群?)，並對各群之屬性進行說明(例如：群1為購買價格(高)、維修保養花費(高)、可搭乘人數(少))
2. 分群後，請針對各群資料，根據所有變數(X1到X6) 使用LDA**方法**進行“是否值得購買”(Y)預測，並用 confusion matrix 評估預測結果
3. 注意：記得看看是否需要使用SMOTE!

P.S. 請將答案(描述跟分析截圖等資訊)存成 word 檔案，連同整理的資料檔案、Weka檔案或python 檔案等，放在 “學號\_姓名”目錄下(例如: 112578000\_chih3c)，並一起壓縮成 “學號\_姓名.rar”(例如: **112578000\_chih3c.rar**)，**在11/21凌晨12:00 以前**寄給我(chih3c@mail.ntut.edu.tw) !! 謝謝，加油!