資料探勘 專案四

* 本作業主要目的為使用Python進行交易資料關聯規則分析。作業內容包括：
  + 資料集：交易資料集。相同INVOICE\_NO表示同一筆交易紀錄。
  + 數量為零或負值的交易代表退貨或註銷，應於前置處理中剔除。
  + 設定不同的支持度及信心度，記錄規則數量及執行時間之變化。應剔除冗餘規則，以減少規則數量。，則為冗餘規則。
  + 請撰寫程式可以將探勘出的關聯規則輸出存檔，也可以讀入已存檔之關聯規則。
  + 撰寫功能允許輸入數項產品，利用關聯規則推薦出現於規則右手邊的產品。同一次推薦中，相同的產品只能出現一次。
  + 請分析不同參數值設定與推薦產品數量多寡之關係。
  + 請分別使用Apriori演算法及FP-Growth演算法，並比較兩者所花費之時間。
  + 撰寫研習報告，請遵循學術論文寫作格式，內容至少必須包括下列要項：

**摘要** (學術論文摘要通常不分段，內容簡要地涵蓋動機、目的、方法與實驗結果)

**一、緒論**

**1.1動機** (探勘所選用的資料集之動機)

**1.2目的** (分析該資料集之目的)

**二、資料集**

**2.1 真實資料集** (簡介包括名稱、筆數、欄位、型態…、並以資料表格呈現部份資料內容)

**三、方法**

**3.1實作說明** (簡要說明實作過程)

**3.2操作說明** (說明如何操作繳交的程式)

**四、實驗**

**4.1前置處理** (說明對資料集做了甚麼前置處理 )

**4.2實驗設計** (實驗如何進行、參數如何設定以及使用的評估指標等）

**4.3實驗結果** (以圖表呈現結果、分析結果及解讀)

(應分析績效好壞之原因、比較不同分群方法之執行時間、實驗結果以表格或圖形統整)

**五、結論**

**六、參考文獻**

* **評分標準**

內容(難度、豐富性、方法正確性)：60%，論文格式：20%，寫作技巧(文筆通順)：20%

* **任何複製剪貼網站資料或學長姐以前作業，屬於抄襲，該作業分數將不會超過50分，嚴重者零分。**
* 此份報告最多20頁，請加註頁碼。請遵循投稿論文寫作格式，使用一欄式，無需製作論文目錄。可以觀摩 ACM Transactions on Knowledge Discovery from Data期刊論文的寫作技巧與格式。
* 紙本報告於上課前繳交或直接放入許老師的系辦信箱。專案程式直接上傳GitHub，專案名稱：hw4+組長學號，並新增助教為專案成員，以利下載評分。缺紙本或程式皆視為未完成繳交作業。
* 繳交紙本時，在左上角訂一訂書針即可，無須其他任何裝訂。
* 每遲交一天，扣該次作業成績10分。
* 繳交日期：112/01/05（四）