Data Mining 作業一 Association Rules & Frequent Itemset

此次作業主要目的在讓同學學習運用 Weka 由 Foodmart Database 超級市場的顧客及其交易資料中，以Apriori, FP-Tree Algorithm探勘Frequent Itemsets, Association Rules, Multi-level Association Rules, Quantitative Association Rules。作業中有些小題直接操作 Weka 即可完成、有些小題則必須撰寫小程式做資料的轉換或必須執行 SQL 查詢。

作業給定的 Foodmart Database（下載連結https://reurl.cc/N6nOyn）可利用Microsoft Access開檔並執行SQL query或匯出成其他格式。Weka input的檔案格式為ARFF。ARFF可以是Relational Table形式也可以是Transaction形式(Sparse ARFF)，請參考Weka的格式說明http://www.cs.waikato.ac.nz/ml/weka/arff.html

作業每人繳交一份。上傳檔名格式為HW1\_學號，EX: HW1\_XXX.doc / .pdf

此次作業交易資料只針對**1998**年的資料 (sales\_fact\_1998+sales\_dec\_1998)

1. 請利用Weka 中的Apriori 演算法，從Foodmart資料庫的交易資料中，探勘符合Minimum Support = 0.0001且Minimum Confidence = 0.9的Association Rules，並列出Confidence最高的前10條Rules。若無法跑出結果，請簡述其原因。
2. 請利用Weka 中的FP-Growth演算法，從Foodmart資料庫的交易資料中，探勘符合Minimum Support = 0.0001 且Minimum Confidence = 0.9的Association Rules，並列出Confidence最高的前10條Rules。若無法跑出結果，請簡述其原因。
3. 有時候我們有興趣的資料不只有產品間的資訊，也會想要由User Profile探勘顧客的基本資料。請運用Weka，給定Minimum Support = 0.05且Minimum Confidence= 0.9的條件下，探勘Foodmart顧客基本資料的屬性{State\_Province, Yearly\_Income , Gender , Total\_Children , Num\_Children\_at\_Home , Education , Occupation, Houseowner , Num\_cars,owned } 間的association rule。(列出10條)
4. 請運用Weka探勘Foodmart資料庫中，顧客背景資料與其交易資料之間的關係(Quantitative Association Rules)。例如80%女性顧客常買保養品。請自行嘗試設定Minimum Support Minimum Confidence，找出10條你覺得有意義的Rules。請說明你的作法及相關參數設定。
5. 在美國由於聖誕節，12月是購物的旺季。請探勘分析比較12月與1~11月的顧客購物行為。有哪些相似的地方，有哪些差異的地方？