金融大數據 2023/4/7 筆記

* 關聯分析**不是找因果關係**，只是找兩者之間**同時發生**的機率。
  + 不是因果關係，只是同時發生。
* 服務品質的構面：

1. 可靠性
2. 回應性
3. 確實性
4. 關懷性
5. 有形性

→ 每個題目需要3-5題，因為要刪除1-2題。

→ 取平均，以gamez0study.csv範本為例，就是取這17題的平均分數。

〔Apriori 演算法〕

<https://ithelp.ithome.com.tw/articles/10218530>

* min\_support 最小支持度→要超過多少數值才會放進來分析，數值介於0~1之間，因為不知道實際多少數值，所以用比例%。
* 因為掃描每項商品、每筆交易，所以效率極差。
* Confidence→要超過多少數值才願意推薦
  + 先解決min\_support再解決Confidence

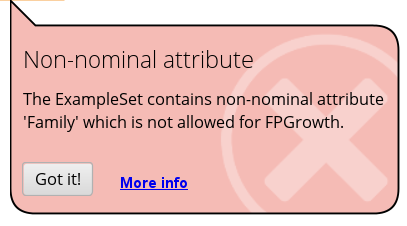
當Support值提高，則分析項目變少 → 頻繁項目集（Frequence）

Ex. 有多少客人進來會買這商品，比方說0.2，代表100人近來有2人會買。

**影響：**

1. 當Support越低→項目越多→規則越多；當Support越高→項目越少→規則越少。
2. 當Confidence越低→規則越多；當Confidence越高→規則越少。

◎在RapidMinder中，mini support=0.95是列出0~0.95的項目。

錯誤：  
解析：只能處理文字類的資料，不能處理數字類的資料。必須要先轉成Binominal。

* Binominal→二元  
  Numerical to Binominal→數字轉二元

一張含有 圖表 的圖片

自動產生的描述

Ex.若要分析Education，則Support要定為0.01

資料劃分 方式：

1. Get\_Dummy
2. 數字→級距 Ex. Age → 0~10歲、10~20歲…

Case→類別型資料



當前者發生，後者發生的機率。

EX. 402筆住民宿，其中374筆開車。

一張含有 圖表 的圖片

自動產生的描述

* 可以將Education、City、Language、Skiing合併為Other其它
* Select Attribute，將Bathing、Active、Recreation、Wander刪除，提高Support及Confidence的數值。  
  → 分子、分母概念