

## 比較性分析：關聯規則

學生姓名與學號

### 1. 資料集描述：

#### 1-1. 請描述資料集的資訊：

請整理成類似「[鐵達尼號逃生資料集](#)」的形式。需有資料來源、簡介、資料筆數、屬性 (含名稱、資料類型、定義、類別或數值的單位、缺失值筆數)、目標屬性、資料預處理流程

### 2. 關聯規則分析：熱點分析

2-1. 請對所有目標值(Target value)進行Hot Spot熱點分析，並在下面貼上 **Associator output** 的探勘結果細節。多個目標值探勘結果之間以分隔線隔開。這樣子老師可以瞭解您執行的參數、資料與結果，看看分析操作是否正確。

#### 參考答案

```
=== Run information ===
```

```
Scheme: weka.associations.HotSpot -c last -V first -S 0.33 -M 30 -length 1 -I 0.01 -R
Relation: WekaODF
Instances: 585
Attributes: 30
    Gender
    Age
```

...(略)...

```
=== Associator model (full training set) ===
```

```
Hot Spot
```

```
=====
```

```
Total population: 585 instances
Target attribute: School
Target value: GP [value count in total population: 381 instances (65.13%)]
Minimum value count for segments: 126 instances (33% of target value total population)
Maximum branching factor: 30
Maximum rule length: 1
Minimum improvement in target: 1%
```

```
[FatEdu > 2]: 234 ==> [School=GP]: 180 <conf:(0.77)> lift:(1.18) lev:(0.05) conv:(1.48)
[TravelTime <= 1]: 324 ==> [School=GP]: 245 <conf:(0.76)> lift:(1.16) lev:(0.06) conv:(1.41)
[Address=urban]: 403 ==> [School=GP]: 303 <conf:(0.75)> lift:(1.15) lev:(0.07) conv:(1.39)

...(略)...

-----

=== Run information ===

Scheme:    weka.associations.HotSpot -c last -V 2 -S 0.33 -M 30 -length 1 -I 0.01
-R
Relation:  WekaODF
Instances: 585

...(略)...
```

**2-2.** 請列出每個目標值(**Target value**)探勘結果中前**5**條關聯規則。不同目標值之間探勘出的關聯規則之間會有所差異，請寫出造成這個差異的可能理由、以及您覺得最合理和最不合理的規則及理由。推測理由的時候，請注意因果的合理性。

**2-3.** 您覺得這個熱點分析的結果符合您的預期嗎？那些結果符合您的預期，那些沒有呢？請寫下您的理由。

**2-4.** 如果您覺得上述分析結果不合理，請試著調整資料與**HotSpot**的參數再來分析看看，並在下方寫下您覺得合理的分析結果，並在後面貼上**Associator output**的探勘結果細節。

---

### 3. 綜合比較

**3-1.** 請比較熱點分析所找到的規則，以及上次授課中用決策樹找出來的結果，兩

者有何異同。您認為在什麼場合適合使用那種方式來分析呢？

[\(跳到頁首\)](#)