**DATA MINING** Q56074027\_吳偲華

* 特徵 : sex, weight, height, visceral fat, body fat, obesity, status
* 規則 :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 營養不良 | | 健康 | | 營養過剩 | |
| 內臟脂肪<4 | 體脂肪<14 | 4<=內臟脂肪<=6 | 14<=體脂肪<=20 | 內臟脂肪>6 | 體脂肪>20 |
| BMI<18.5 | 肥胖度<-10 | 18.5<=BMI<=24 | -10<=肥胖度<=10 | BMI>24 | 肥胖度>10 |

男生:

女生:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 營養不良 | | 健康 | | 營養過剩 | |
| 內臟脂肪<2 | 體脂肪<17 | 2<=內臟脂肪<=4 | 17<=體脂肪<=24 | 內臟脂肪>4 | 體脂肪>24 |
| BMI<18.5 | 肥胖度<-10 | 18.5<=BMI<=24 | -10<=肥胖度<=10 | BMI>24 | 肥胖度>10 |

* 方法 :

Step 1 : 建立特徵和資料

Step 2 : 建立規則和分類答案

Step 3 : 將資料分成訓練和測試資料

Step 4 : 建立一個分類器(entropy)

Step 5 : 進行預測

Step 6 : 計算績效

* 條件比較 :
  + 資料量 : 500 VS 10000，21條規則

|  |  |  |
| --- | --- | --- |
| 資料量 | 500(筆) | 1000(筆) |
| 績效分數 | 0.67 | 0.93 |

* + 資料量 : 500 VS 10000，9條規則

|  |  |  |
| --- | --- | --- |
| 資料量 | 500(筆) | 1000(筆) |
| 績效分數 | 0.92 | 0.98 |

* 結論 :

利用決策樹進行分類，在同樣規則數量下，建立的資料量愈大，分類出來的效果就會愈好；反之則愈差。而訂定的規則愈少，分類出來的效果相對較好；反之則愈差。資料量少時，可思考是否刪減一些不那麼重要的規則，以提升效果；反之，則可以多加一些規則，精細的分類。