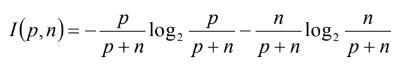
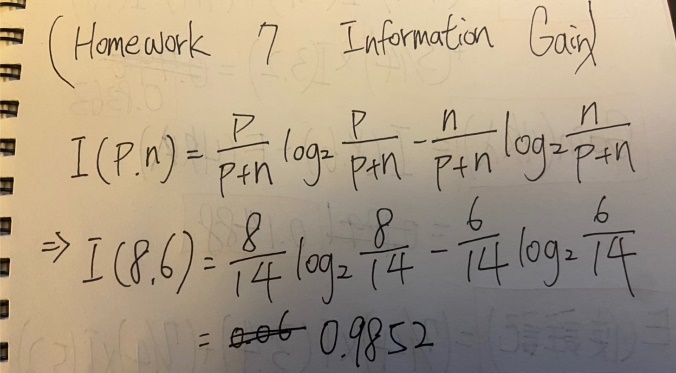
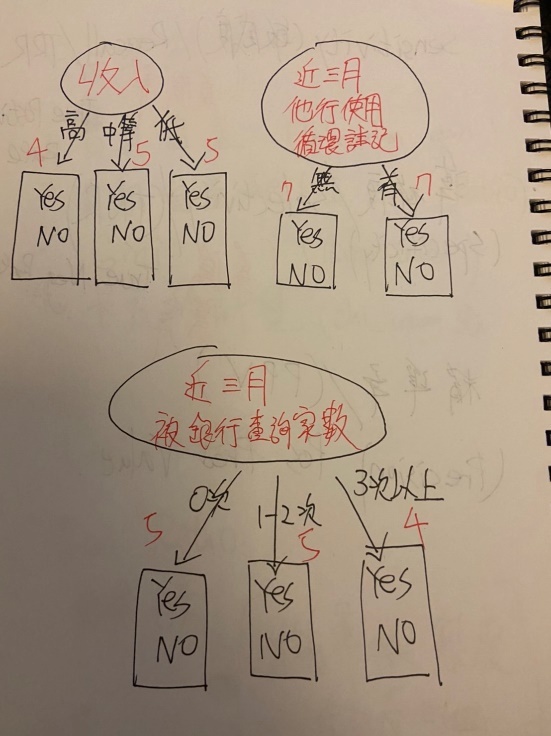
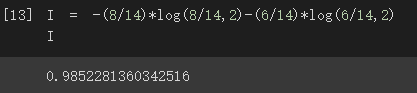
**1. 請以「是否違約」為 target，利用 info gain 以及 gain ratio 計算應該使用哪一個特性來做第一層決測樹之分割**

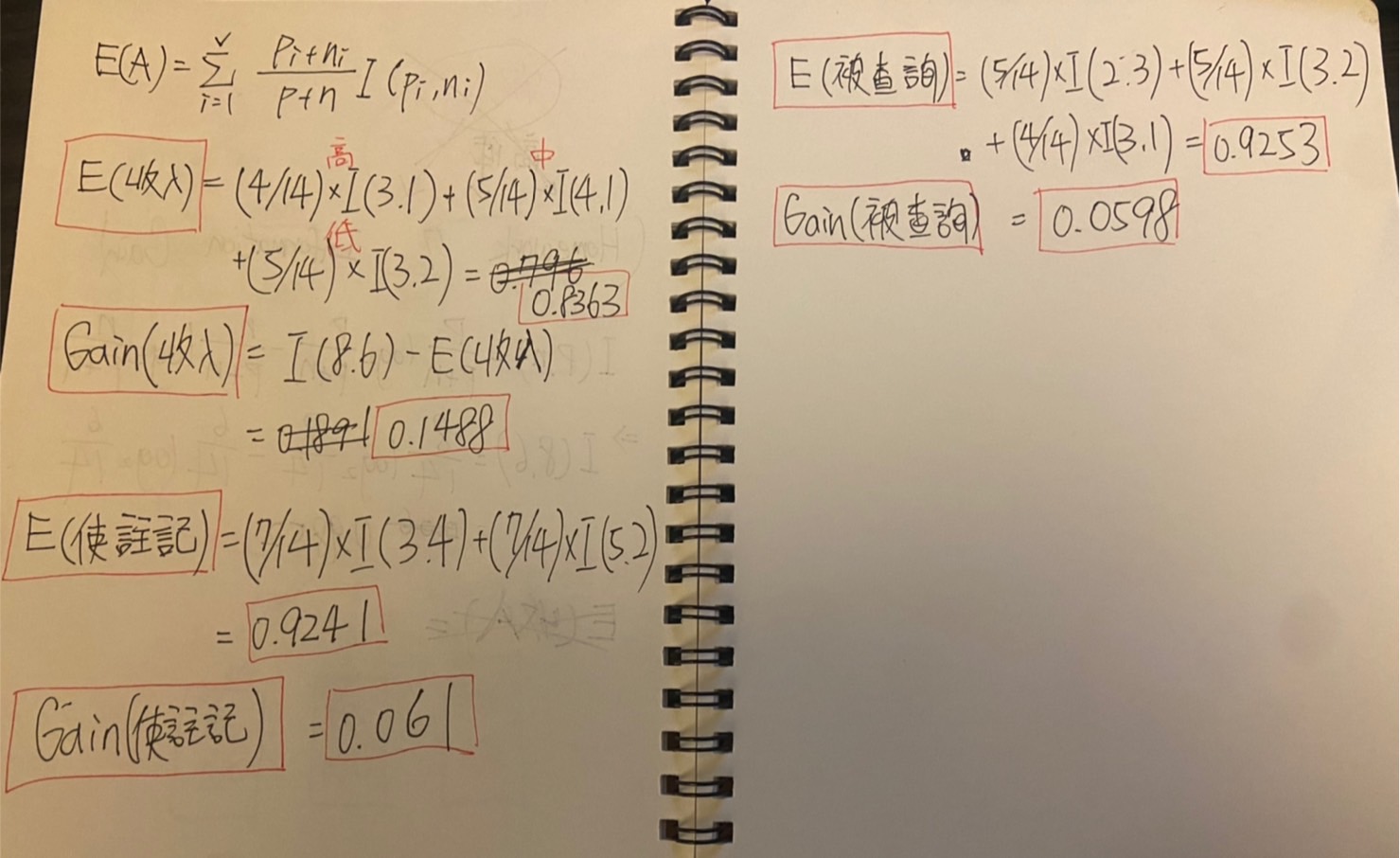
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 客戶編號 | 收入 | 最近三個月他行使用循環註記 | 最近三個月被銀行查詢家數 | 是否違約 |
| 1 | 高 | 無 | 0次 | No |
| 2 | 高 | 無 | 0次 | No |
| 3 | 高 | 無 | 1~2次 | Yes |
| 4 | 中等 | 無 | 3次以上 | Yes |
| 5 | 低 | 有 | 3次以上 | Yes |
| 6 | 低 | 有 | 1~2次 | Yes |
| 7 | 低 | 有 | 1~2次 | Yes |
| 8 | 中等 | 無 | 0次 | Yes |
| 9 | 低 | 有 | 0次 | No |
| 10 | 中等 | 有 | 3次以上 | Yes |
| 11 | 中等 | 有 | 0次 | Yes |
| 12 | 中等 | 無 | 1~2次 | No |
| 13 | 高 | 有 | 1~2次 | No |
| 14 | 低 | 無 | 3次以上 | No |

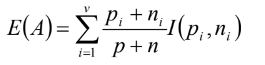
**Ans(Info gain):**

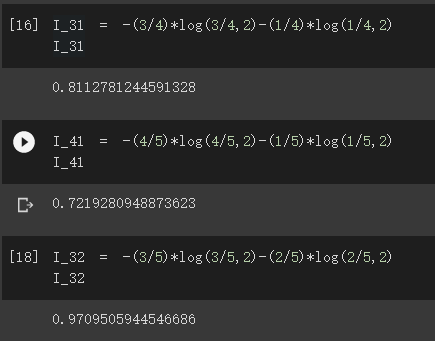




**目標總資訊量: I(8,6)**



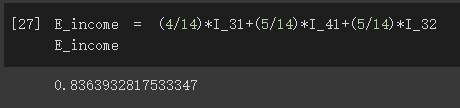




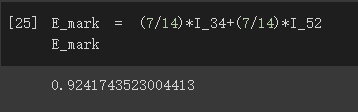
**I(3,2)**

**I(4,1)**

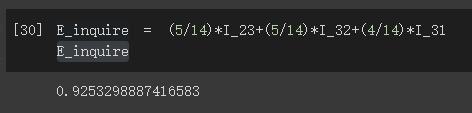
**I(3,1)**



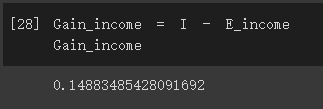
**E(收入)**



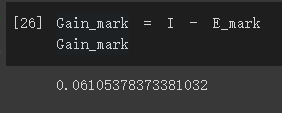
**E(使標記)**



**E(被查詢)**

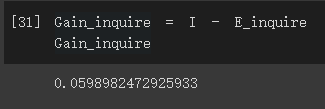


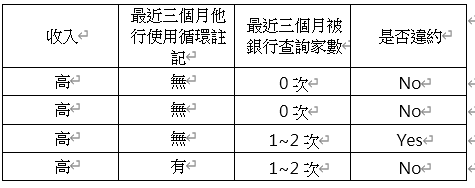
**Gain(收入)最高**



**使用 ”收入”**

**作為第一層決測樹之分割**

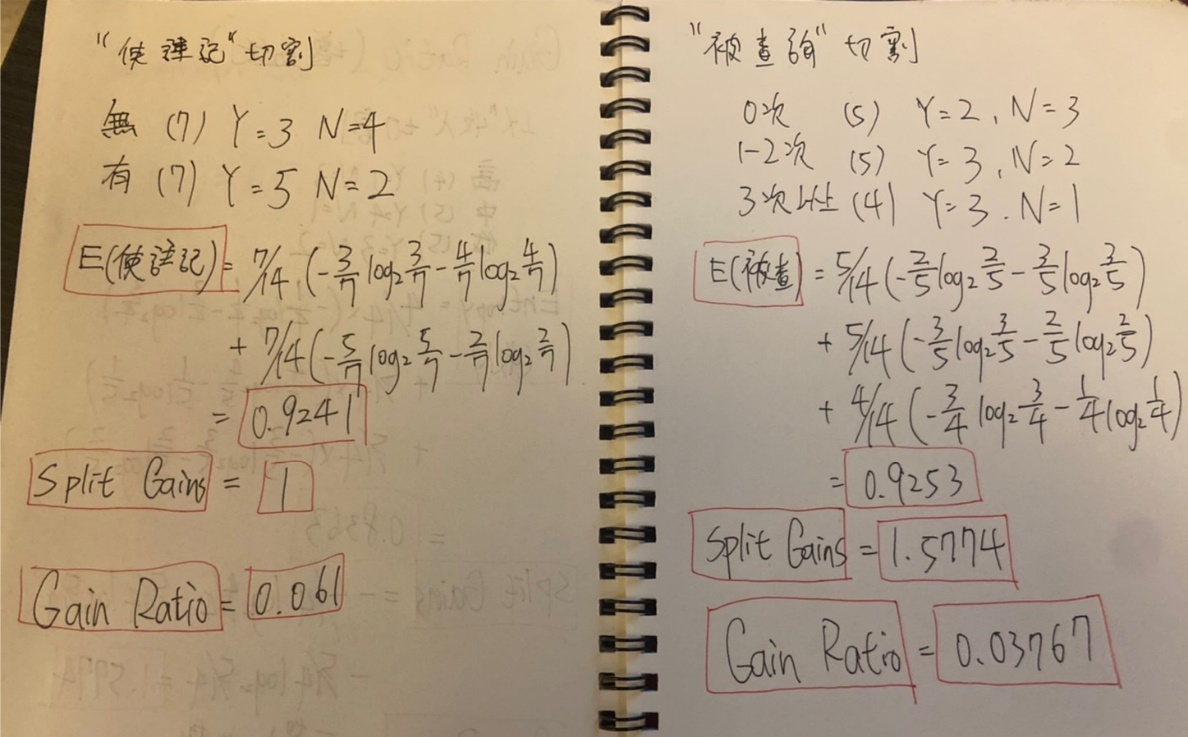
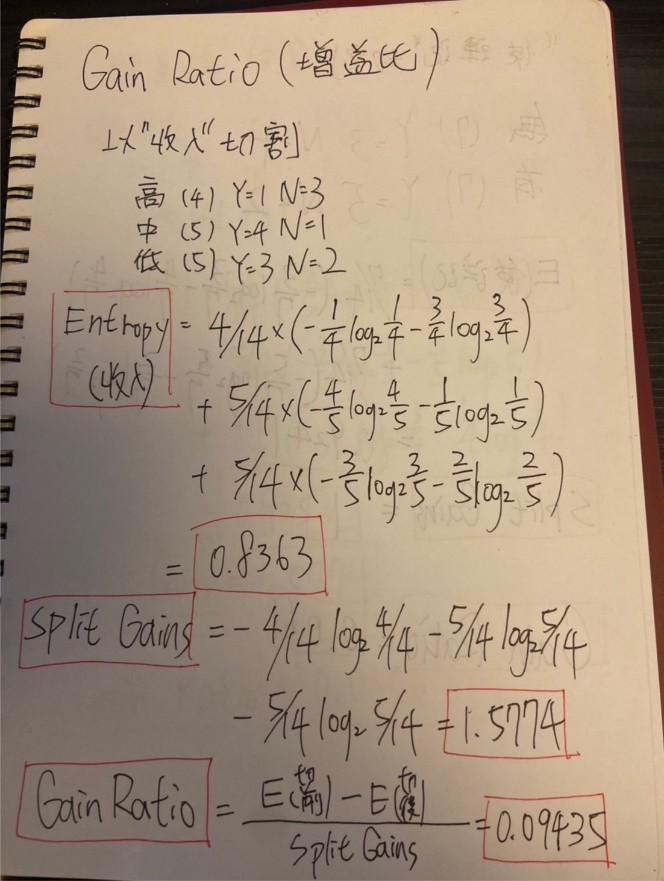




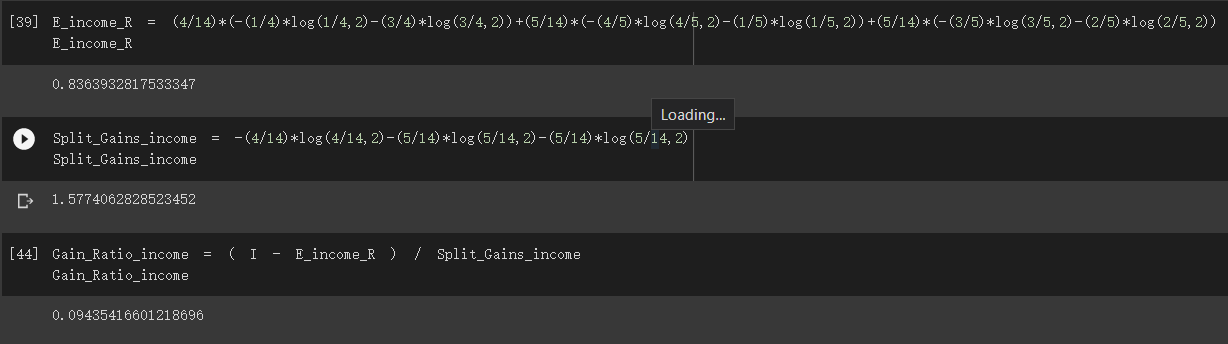




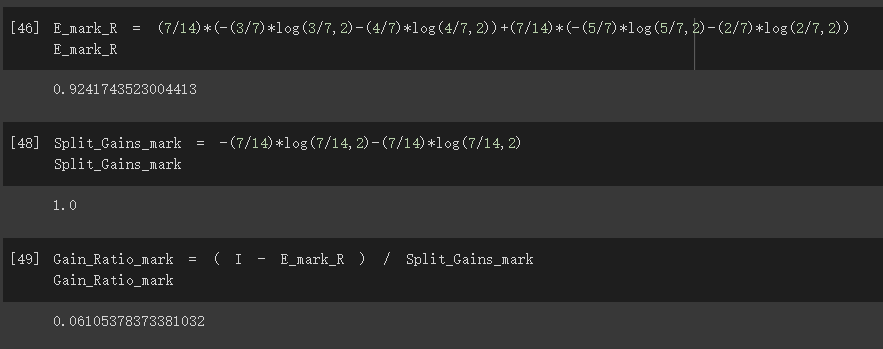
**Ans(Gain Ratio):**

****

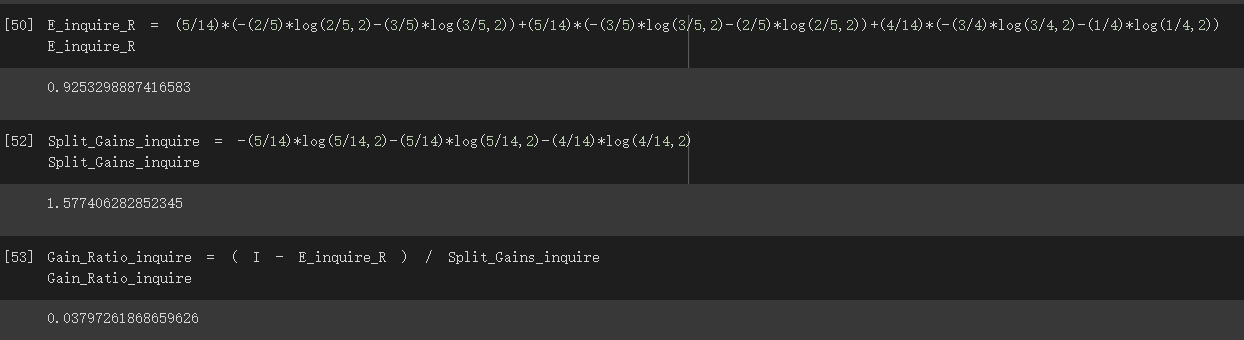
**Gain R (收入)最高**



**Gain R (收入)最高**



**Gain R (使註記)**



**Gain R (被查詢)**

2. 請用 Gini Index 計算下表各分支屬性之純度，並決定能獲得較多資訊之分支屬性。