Phần 2. BUYS­­-COMPUTER

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Age** | **Income** | **Student** | **Credit\_rating** | **Buys\_Computer** |
| 1 | 33 | high | no | fair | no |
| 2 | 30 | high | no | excellent | no |
| 3 | 35 | high | no | fair | yes |
| 4 | 41 | medium | no | fair | yes |
| 5 | 45 | low | yes | fair | yes |
| 6 | 43 | low | yes | excellent | no |
| 7 | 36 | low | yes | excellent | yes |
| 8 | 32 | medium | no | fair | no |
| 9 | 34 | low | yes | fair | yes |
| 10 | 42 | medium | yes | fair | yes |
| 11 | 31 | medium | yes | excellent | yes |
| 12 | 37 | medium | no | excellent | yes |
| 13 | 39 | high | yes | fair | yes |
| 14 | 44 | medium | no | excellent | no |

1. Phân nhóm

Với dữ liệu trên ta có:

* Số lượng phần tử: n=14
* Số nhóm: K= 3
* Khoảng cách h=(45-30)/3=5
* Ta có khoảng: (30, 35); (35, 40); (40,45)

Bảng dữ liệu sau khi phân nhóm:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Age** | **Income** | **Student** | **Credit\_rating** | **Buys\_Computer** |
| 1 | 30-35 | high | no | fair | no |
| 2 | 30-35 | high | no | excellent | no |
| 3 | 35-40 | high | no | fair | yes |
| 4 | 40-45 | medium | no | fair | yes |
| 5 | 40-45 | low | yes | fair | yes |
| 6 | 40-45 | low | yes | excellent | no |
| 7 | 35-40 | low | yes | excellent | yes |
| 8 | 30-35 | medium | no | fair | no |
| 9 | 30-35 | low | yes | fair | yes |
| 10 | 40-45 | medium | yes | fair | yes |
| 11 | 30-35 | medium | yes | excellent | yes |
| 12 | 35-40 | medium | no | excellent | yes |
| 13 | 35-40 | high | yes | fair | yes |
| 14 | 40-45 | medium | no | excellent | no |

1. Tính GINIsplit

* **Với Age:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | Không Buys\_computer | có Buys\_computer | Tổng |
| **30-35** | **3** | **2** | **6** |
| **35-40** | **0** | **4** | **4** |
| **40-45** | **2** | **3** | **5** |

Gini (Age = 30-35) = 0.48

Gini (Age = 35-40) = 0

Gini (Age = 40-45) = 0.48

Ginisplit (Age) = (5/14)\*0.48 + (4/14)\*0+ (5/14)\*0.48 = 0.3428571429

* **Với Income**

|  |  |  |  |
| --- | --- | --- | --- |
| **Income** | Không Buys\_computer | có Buys\_computer | Tổng |
| **hign** | **2** | **2** | **4** |
| **medium** | **1** | **3** | **4** |
| **low** | **2** | **4** | **6** |

Gini (Income = hign) = 0.5

Gini (Income = medium) = 0.444

Gini (Income = low) = 0.375

Ginisplit (Income) = (4/14)\*0.5 + (4/14)\*0.444+ (6/14)\*0.375 = 0.4304285714

* **Với Student**

|  |  |  |  |
| --- | --- | --- | --- |
| **Student** | Không Buys\_computer | có Buys\_computer | Tổng |
| **Yes** | **4** | **3** | **7** |
| **No** | **1** | **6** | **7** |

Gini (Student=No) = 0.489

Gini (Student=yes) = 0.244

Ginisplit (Student) = (7/14)\*0.490 + (7/14)\*0.244 = 0.3665

* **Với Credit\_rating**

|  |  |  |  |
| --- | --- | --- | --- |
| **Credit\_rating** | Không Buys\_computer | có Buys\_computer | Tổng |
| **Fair** | **2** | **6** | **8** |
| **excellent** | **3** | **3** | **6** |

Gini (Credit\_rating= fair) = 0.375

Gini (Credit\_rating= excellent) = 0.5

Ginisplit (Student) = (8/14)\*0.375 + (6/14)\*0.5 = 0.4285714286