

Bài tập Lý thuyết 10

Thông tin sinh viên

Bài 1.

Câu a.

Câu b. Trong các hành tinh trên, chỉ 9 hành tinh đo được nhiệt độ bề mặt:

Câu 2. Một cửa hàng có dữ liệu về khách hàng như sau

Thông tin sinh viên

- MSSV: 22850034
- Họ và tên: Cao Hoài Việt
- Email: viet.ch2612@gmail.com

Bài 1.

Câu a.

Vào tháng 9 năm 2012, các nhà thiên văn học đã tìm ra được 800 hành tinh ngoài hệ trời. Bằng cách dùng kính thiên văn, các nhà khoa học đã có thống kê như sau

| Kích thước | Quỹ Đạo | Số lượng | Con người có thể ở được? |
|------------|---------|----------|--------------------------|
| Lớn | Gần | 20 | Yes |
| Lớn | Xa | 170 | Yes |
| Nhỏ | Gần | 139 | Yes |
| Nhỏ | Xa | 45 | Yes |
| Lớn | Gần | 130 | No |
| Lớn | Xa | 30 | No |
| Nhỏ | Gần | 11 | No |
| Nhỏ | Xa | 255 | No |

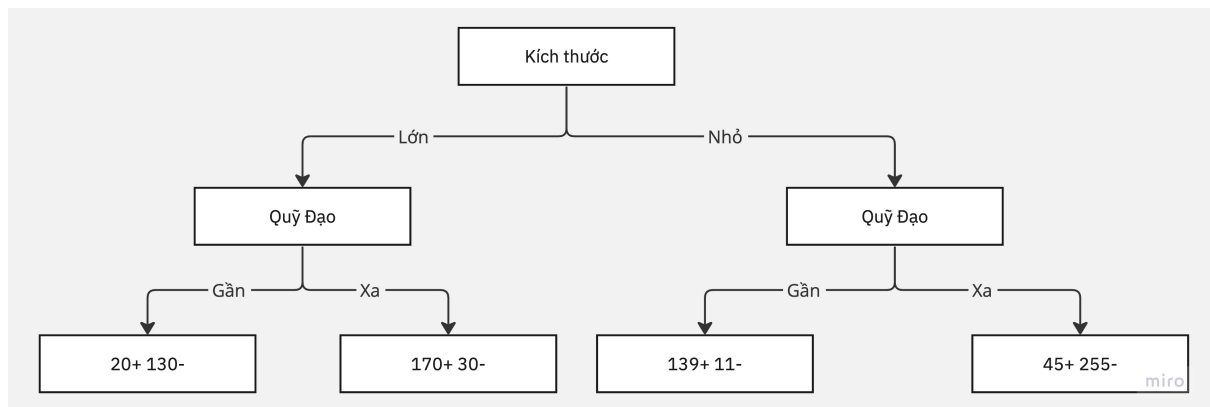
Chúng ta xác định rằng “Con người có thể ở được” hay không ở đây là output, còn “Số lượng” sẽ được sử dụng để tính Entropy

Ta tính Entropy cho hai thuộc tính Kích thước và Quỹ đạo

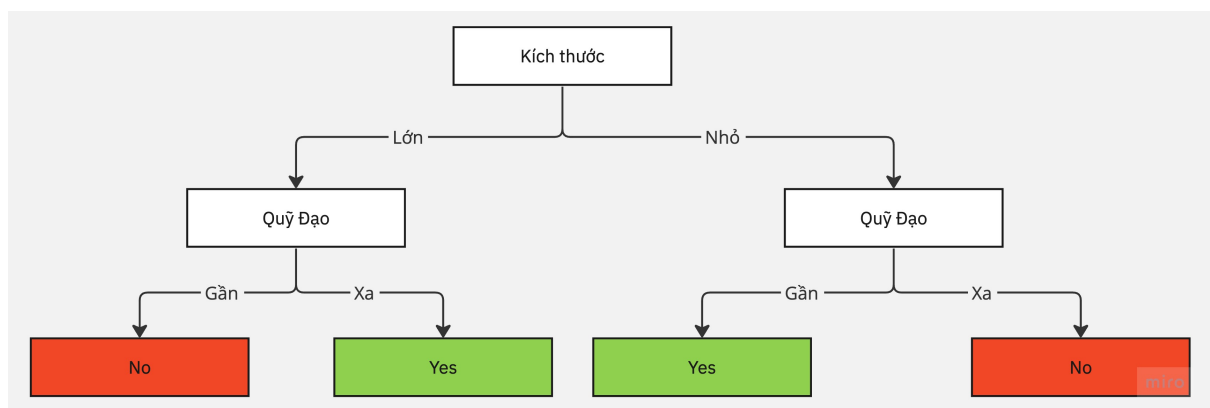
| Kích thước | AE= | 0.9841297958 | |
|------------|-----|--------------|--------------|
| 0.4375 | Lớn | 190+ 160- | 0.9946937954 |
| 0.5625 | Nhỏ | 184+ 266- | 0.9759133518 |

| Quỹ Đạo | AE= | 0.9901599939 | |
|---------|-----|--------------|--------------|
| 0.375 | Gần | 159+ 141- | 0.9974015886 |
| 0.625 | Xa | 215+ 285- | 0.9858150372 |

Như vậy ta sẽ chọn root là Kích thước, lá sẽ là Quỹ đạo: Gần hoặc Xa



Chúng ta sẽ lấy kết quả có nhiều hơn. Kết quả cây ID3 cuối cùng như sau



Câu b. Trong các hành tinh trên, chỉ 9 hành tinh đo được nhiệt độ bề mặt:

| Kích thước | Quỹ Đạo | Nhiệt độ bề mặt | Ở được |
|------------|---------|-----------------|--------|
| Lớn | Xa | 205 | No |
| Lớn | Gần | 205 | No |
| Lớn | Gần | 260 | Yes |

| Kích thước | Quỹ Đạo | Nhiệt độ bề mặt | Ở được |
|------------|---------|-----------------|--------|
| Lớn | Gần | 380 | Yes |
| Nhỏ | Xa | 205 | No |
| Nhỏ | Xa | 260 | Yes |
| Nhỏ | Gần | 260 | Yes |
| Nhỏ | Gần | 380 | No |
| Nhỏ | Gần | 380 | No |

Ta sẽ thay 3 lớp nhiệt độ vào để dễ tính toán.

- Gọi T1 sẽ là có $T < 232.5$
- Gọi T2 sẽ là $232.5 \leq T \leq 320$
- Gọi T3 sẽ là $T > 320$

Ta có bảng mới như sau

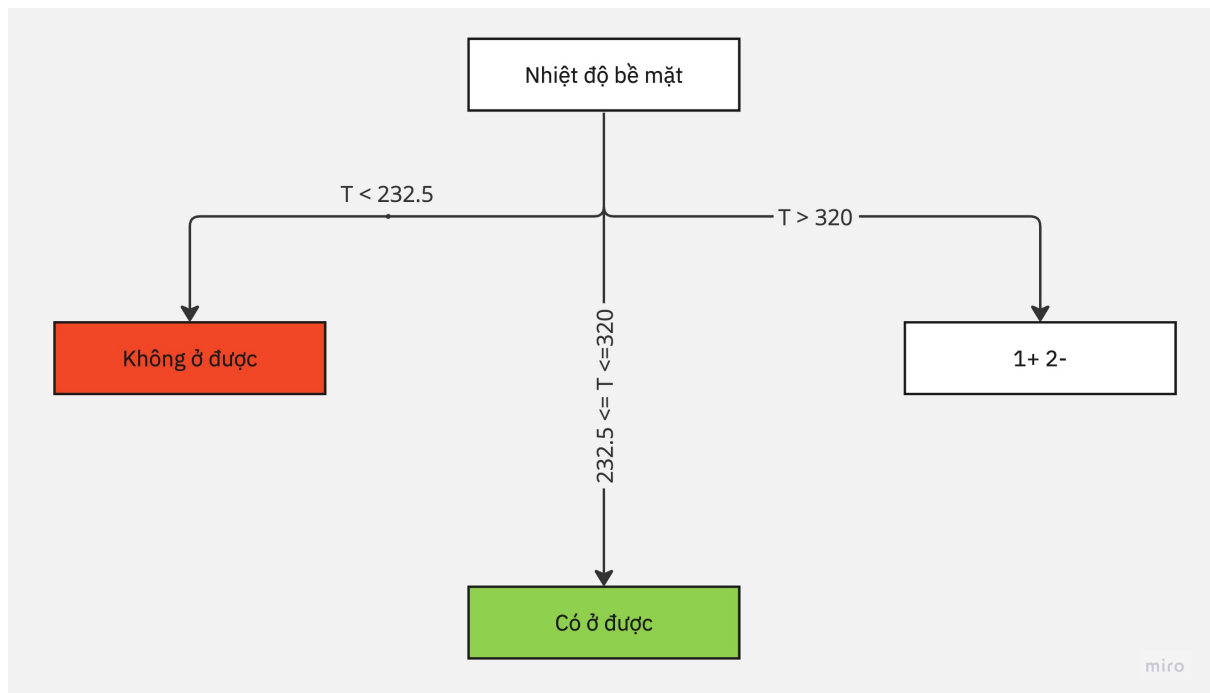
| Kích thước | Quỹ Đạo | Lớp nhiệt độ bề mặt | Ở được |
|------------|---------|---------------------|--------|
| Lớn | Xa | T1 | No |
| Lớn | Gần | T1 | No |
| Lớn | Gần | T2 | Yes |
| Lớn | Gần | T3 | Yes |
| Nhỏ | Xa | T1 | No |
| Nhỏ | Xa | T2 | Yes |
| Nhỏ | Gần | T2 | Yes |
| Nhỏ | Gần | T3 | No |
| Nhỏ | Gần | T3 | No |

Ta tính Entropy cho các thuộc tính trên

| Kích thước | AE= | 0.9838614414 | |
|--------------|-----|--------------|--------------|
| 0.4444444444 | Lớn | 2+ 2- | 1 |
| 0.5555555556 | Nhỏ | 3+ 2- | 0.9709505945 |
| | | | |
| Quỹ Đạo | AE= | 0.972765278 | |
| 0.6666666667 | Gần | 3+ 3- | 1 |
| 0.3333333333 | Xa | 1+ 2- | 0.9182958341 |

| | | | |
|----------------------------|------------|---------------------|--------------|
| | | | |
| Lớp nhiệt độ bề mặt | AE= | 0.3060986114 | |
| 0.3333333333 | T1 | 3- | 0 |
| 0.3333333333 | T2 | 3+ | 0 |
| 0.3333333333 | T3 | 1+ 2- | 0.9182958341 |

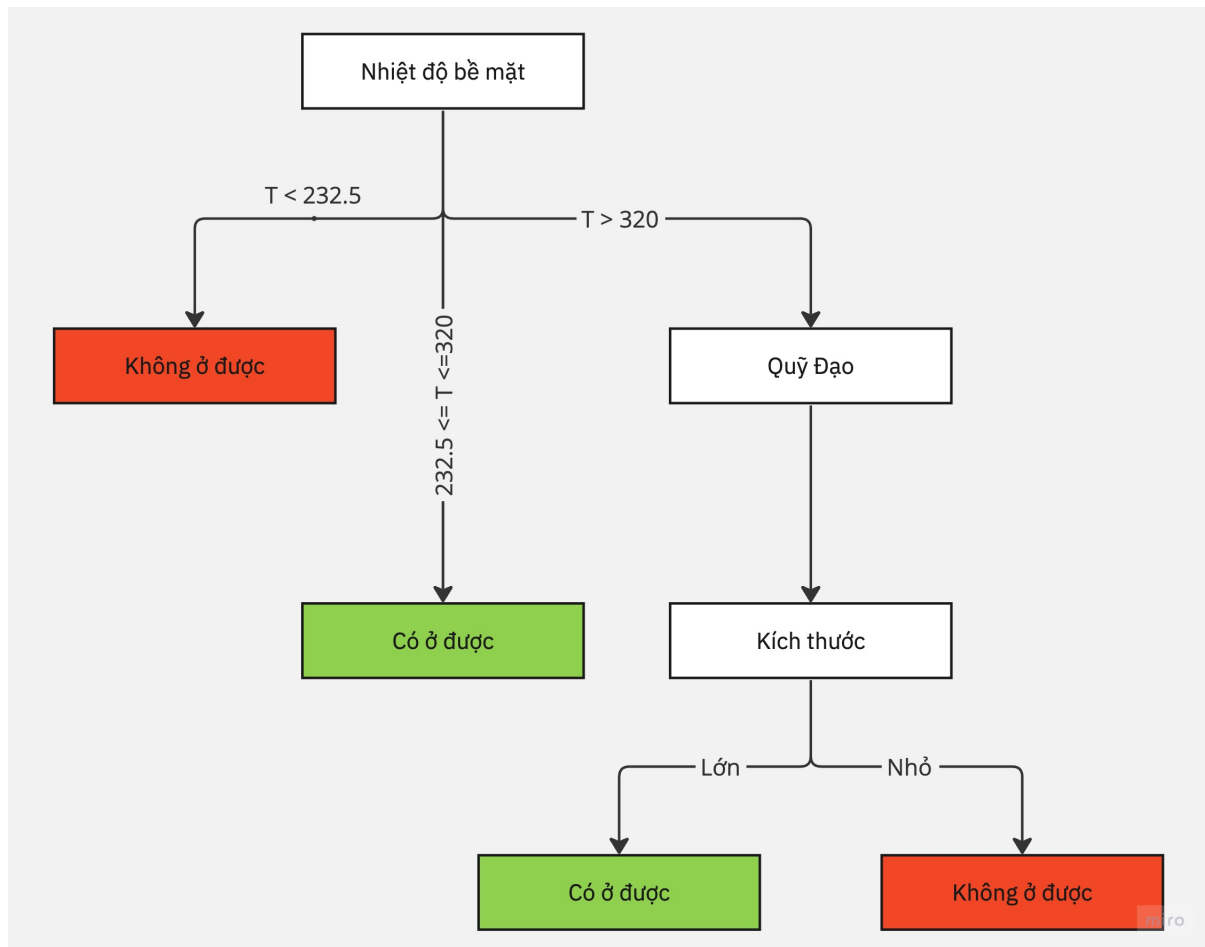
Vì “Lớp nhiệt độ bề mặt nhỏ nhất” nên ta sẽ chọn làm gốc



Vì nhánh $T > 320$ chưa rõ ràng nên chúng ta sẽ xét tiếp

| Kích thước | AE= | 1 | |
|-------------------|------------|---------------------|--------------|
| 0.3333333333 | Lớn | 1+ | 1 |
| 0.6666666667 | Nhỏ | 2- | 1 |
| | | | |
| Quỹ Đạo | AE= | 0.9182958341 | |
| 1 | Gần | 1+ 2- | 0.9182958341 |
| 0 | Xa | 0 | 0 |

Như vậy chúng ta sẽ chọn tiếp Quỹ đạo và cuối cùng là kích thước. Cây ID3 cuối cùng sẽ như sau



Từ cây ID3 trên thì với một hành tinh (Lớn, gần, 280) thì “Có ở được”

Câu 2. Một cửa hàng có dữ liệu về khách hàng như sau

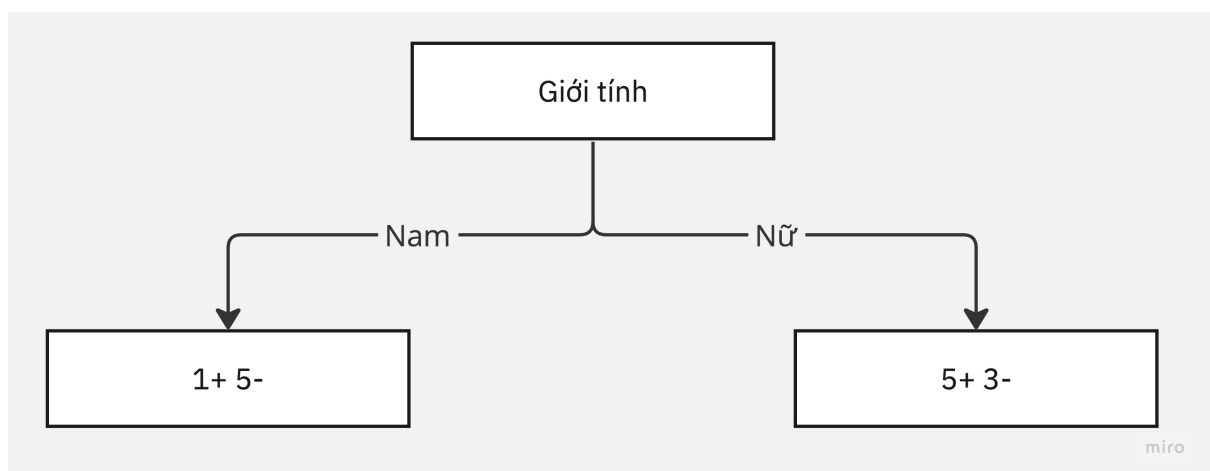
| Tuổi | Giới tính | Nghề nghiệp | Có mua hàng không? |
|------------|-----------|-------------|--------------------|
| Trẻ | Nam | Sinh viên | Có |
| Trung Niên | Nam | Công nhân | Không |
| Già | Nữ | Công nhân | Không |
| Trung Niên | Nam | Công nhân | Không |
| Già | Nữ | Kỹ sư | Có |
| Trung Niên | Nữ | Kỹ sư | Có |
| Trẻ | Nữ | Sinh viên | Không |
| Trung Niên | Nam | Kỹ sư | Không |
| Trung Niên | Nữ | Công nhân | Có |

| | | | |
|-----|-----|-----------|-------|
| Trẻ | Nam | Kỹ sư | Không |
| Trẻ | Nữ | Kỹ sư | Có |
| Già | Nữ | Công nhân | Không |
| Già | Nam | Kỹ sư | Không |
| Trẻ | Nữ | Công nhân | Có |

Ta tính Entropy như sau

| | | | |
|--------------------|------------|---------------------|--------------|
| Tuổi | AE= | 0.9521453229 | |
| 0.3571428571 | Trẻ | 3+ 2- | 0.9709505945 |
| 0.3571428571 | Trung niên | 1+ 2- | 0.9182958341 |
| 0.2857142857 | Già | 2+ 3- | 0.9709505945 |
| | | | |
| Giới tính | AE= | 0.8239718967 | |
| 0.4285714286 | Nam | 1+ 5- | 0.6500224216 |
| 0.5714285714 | Nữ | 5+ 3- | 0.9544340029 |
| | | | |
| Nghề nghiệp | AE= | 0.9649839289 | |
| 0.1428571429 | Sinh viên | 1+ 1- | 1 |
| 0.4285714286 | Công nhân | 2+ 4- | 0.9182958341 |
| 0.4285714286 | Kỹ sư | 3+ 3- | 1 |

Vì giới tính có Entropy nhỏ nhất nên ta sẽ chọn làm gốc

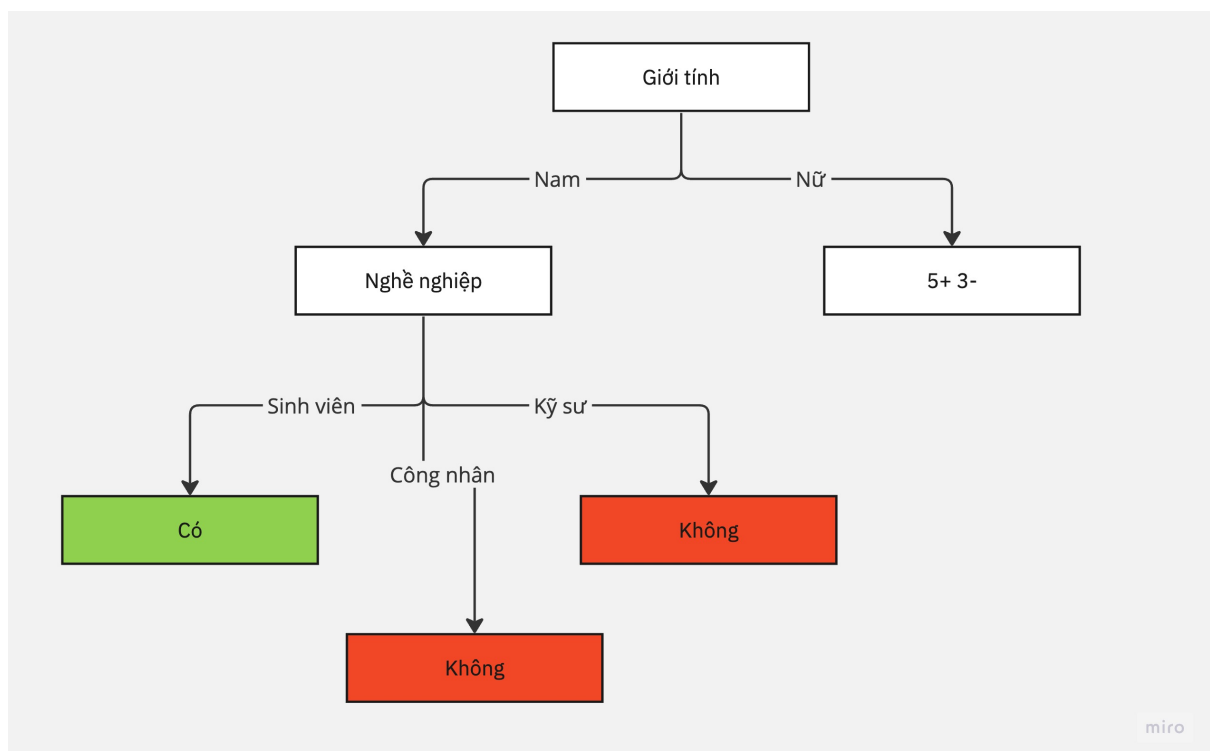


Ta tính tiếp Entropy cho các trường hợp giới tính Nam và Nữ

Đối với nhánh Nam

| Tuổi | AE= | 0.3333333333 | |
|--------------|------------|--------------|---|
| 0.3333333333 | Trẻ | 1+ 1- | 1 |
| 0.5 | Trung niên | 3- | 0 |
| 0.1666666667 | Già | 1- | 0 |
| | | | |
| Nghề nghiệp | AE= | 0 | |
| 0.1666666667 | Sinh viên | 1+ | 0 |
| 0.3333333333 | Công nhân | 2- | 0 |
| 0.5 | Kỹ sư | 3- | 0 |

Như vậy ta sẽ có kết quả liền khi chọn Nghề nghiệp mà không cần tới xét tuổi

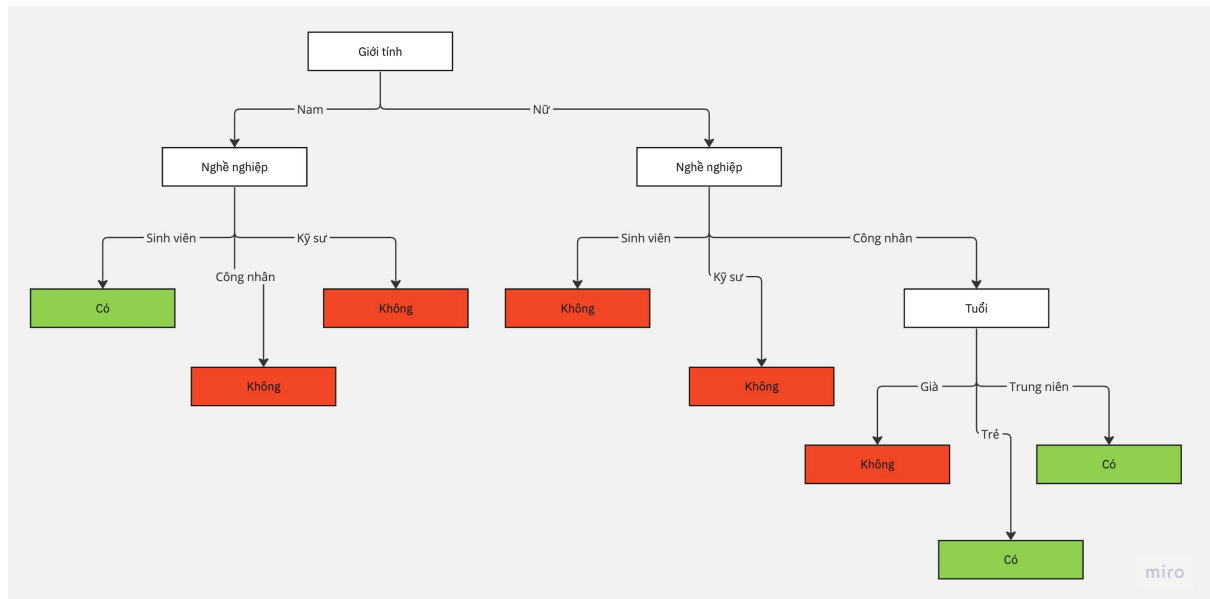


Tiếp đó ta tính tiếp tới Giới tính bằng Nữ

| Tuổi | AE= | 0.6887218755 | |
|-------|------------|--------------|--------------|
| 0.375 | Trẻ | 2+ 1- | 0.9182958341 |
| 0.25 | Trung niên | 2+ | 0 |
| 0.375 | Già | 1+ 2- | 0.9182958341 |

| Nghề nghiệp | AE= | 0.5 | |
|-------------|-----------|-------|---|
| 0.125 | Sinh viên | 1- | 0 |
| 0.5 | Công nhân | 2+ 2- | 1 |
| 0.375 | Kỹ sư | 3+ | 0 |

Ta chọn tiếp Nghề nghiệp do Entropy nhỏ nhất và sau đó là Tuổi. Cây ID3 cuối cùng của chúng ta sẽ như sau



Như vậy, với một khách hàng mới {Tuổi: Trung niên, Giới tính: Nữ, Nghề nghiệp: Sinh viên} thì dự đoán khách này sẽ **KHÔNG** mua.