

# 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

Câu 3. Sử dụng Naive Bayes để dự đoán.

Input 1.

Input 2

Câu 4. Vẽ cây ID3

Câu 5. Dùng thuật toán ILA để rút ra tập luật cho bảng dữ liệu 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

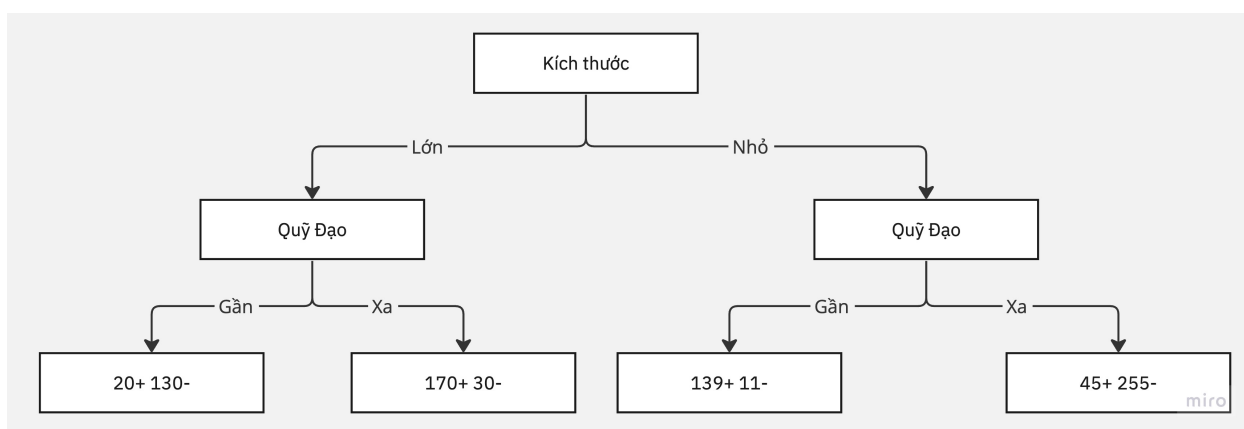
Chúng ta xác định rằng “Con người có thể ở được” hay không ở đây là output, còn “Số lượng” sẽ được sử dụng để tính Entropy

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

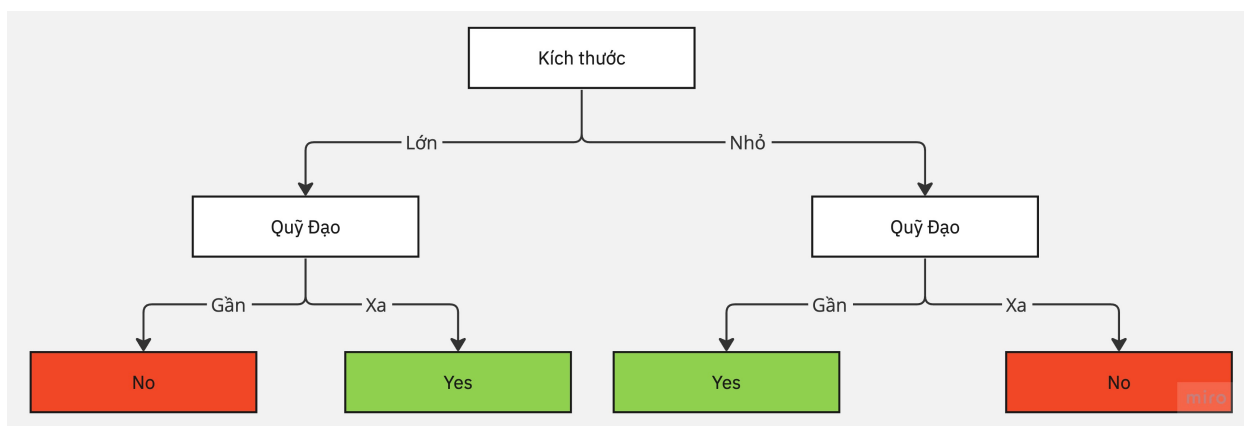
Kích thước	AE=	0.9841297958	
0.4375	Lớn	190+ 160-	0.9946937954
0.5625	Nhỏ	184+ 266-	0.9759133518

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

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

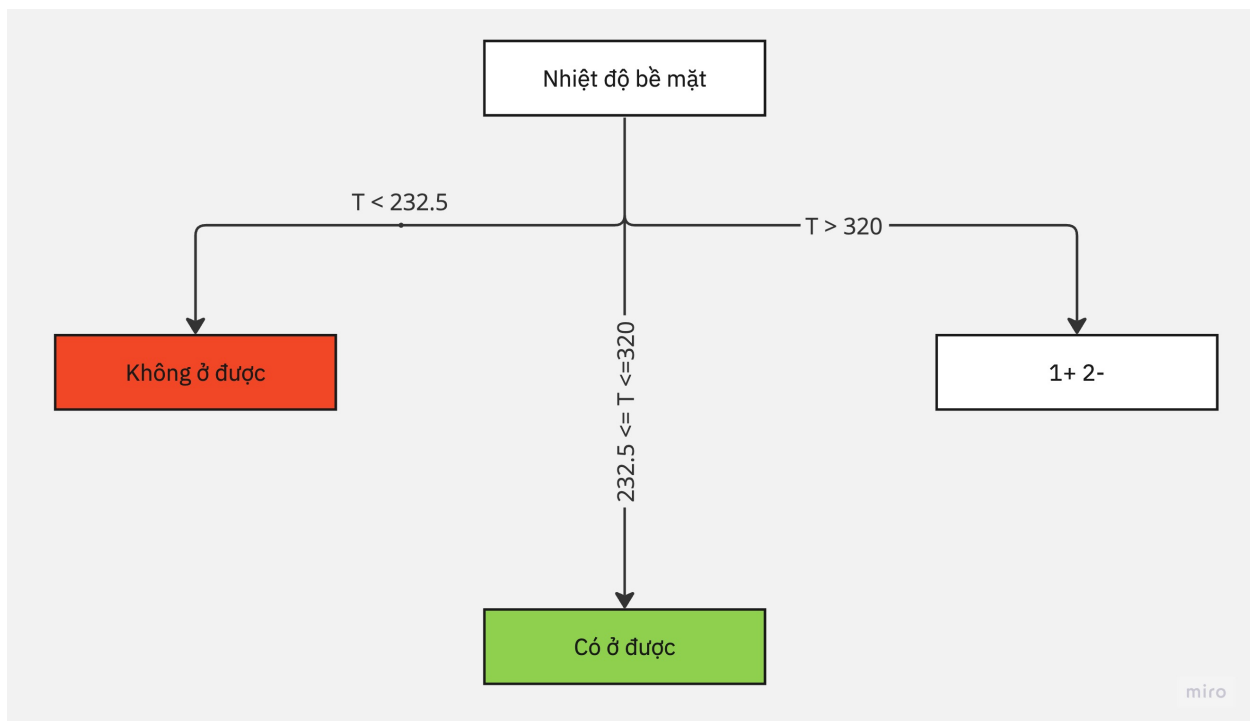
Kích thước	Quỹ Đạo	Lớp nhiệt độ bề mặt	Ở được
Lớn	Xa	T1	No
Lớn	Gần	T1	No
Lớn	Gần	T2	Yes
Lớn	Gần	T3	Yes
Nhỏ	Xa	T1	No
Nhỏ	Xa	T2	Yes
Nhỏ	Gần	T2	Yes
Nhỏ	Gần	T3	No
Nhỏ	Gần	T3	No

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

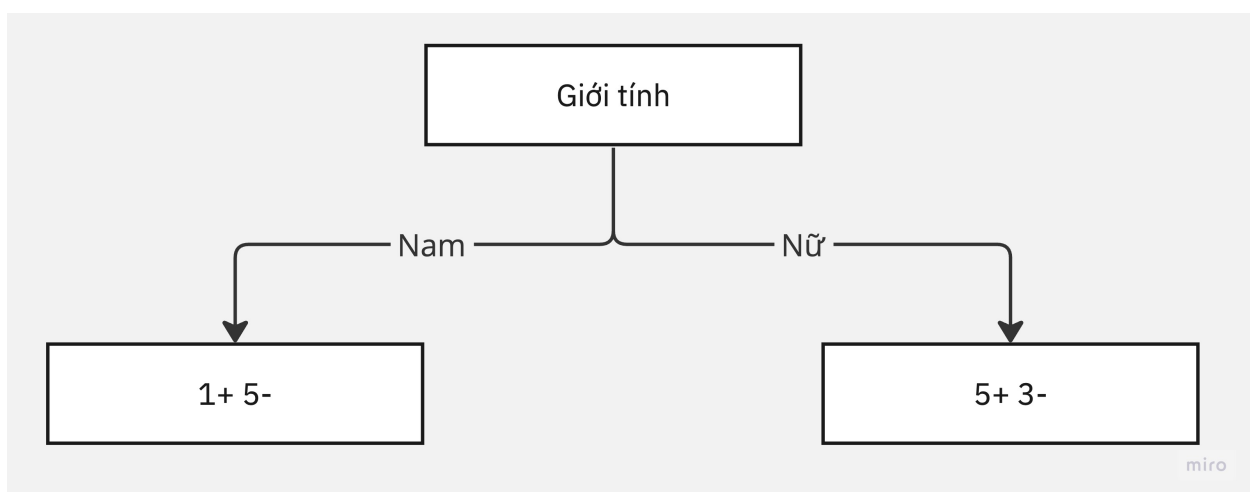
Tuổi	Giới tính	Nghề nghiệp	Có mua hàng không?
Trẻ	Nam	Sinh viên	Có
Trung Niên	Nam	Công nhân	Không
Già	Nữ	Công nhân	Không
Trung Niên	Nam	Công nhân	Không
Già	Nữ	Kỹ sư	Có
Trung Niên	Nữ	Kỹ sư	Có
Trẻ	Nữ	Sinh viên	Không
Trung Niên	Nam	Kỹ sư	Không

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

<b>Tuổi</b>	<b>AE=</b>	<b>0.9521453229</b>	
0.3571428571	Trẻ	3+ 2-	0.9709505945
0.3571428571	Trung niên	1+ 2-	0.9182958341
0.2857142857	Già	2+ 3-	0.9709505945
<b>Giới tính</b>	<b>AE=</b>	<b>0.8239718967</b>	
0.4285714286	Nam	1+ 5-	0.6500224216
0.5714285714	Nữ	5+ 3-	0.9544340029
<b>Nghề nghiệp</b>	<b>AE=</b>	<b>0.9649839289</b>	
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

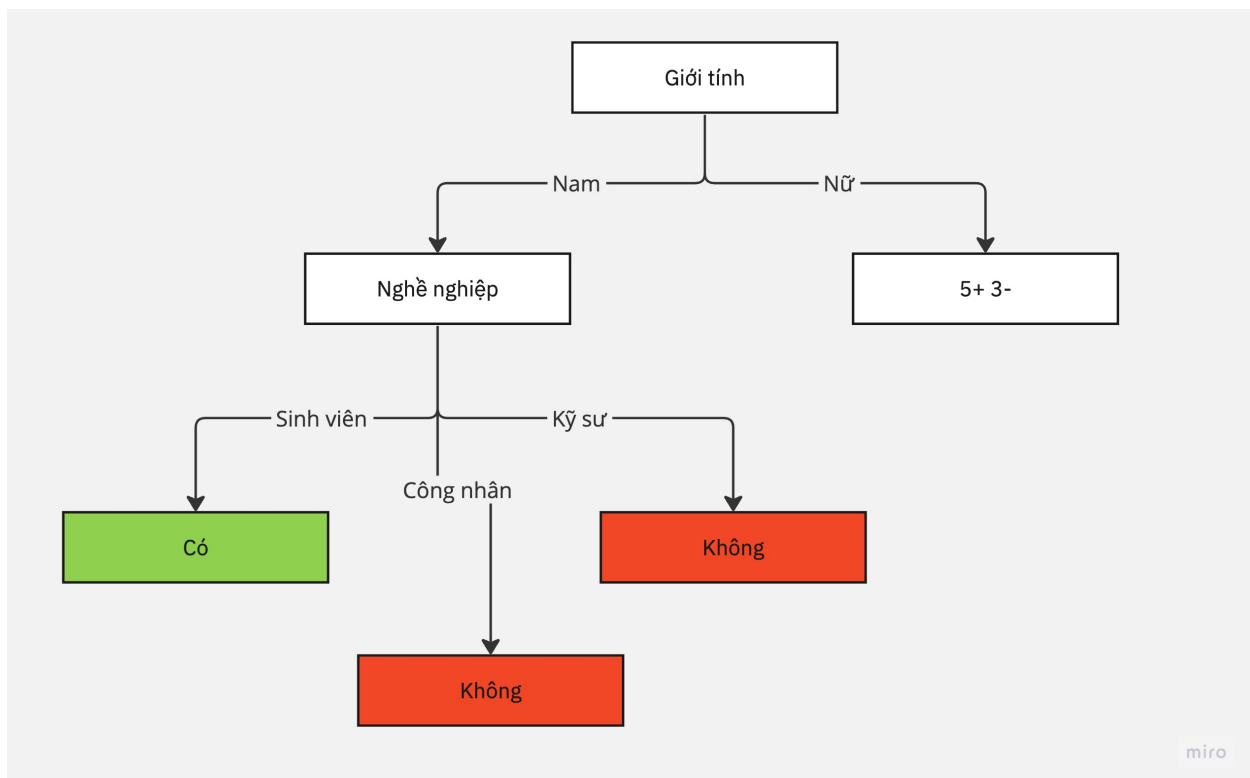


Ta tính tiếp Entropy cho các trường hợp giới tính Nam và Nữ

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

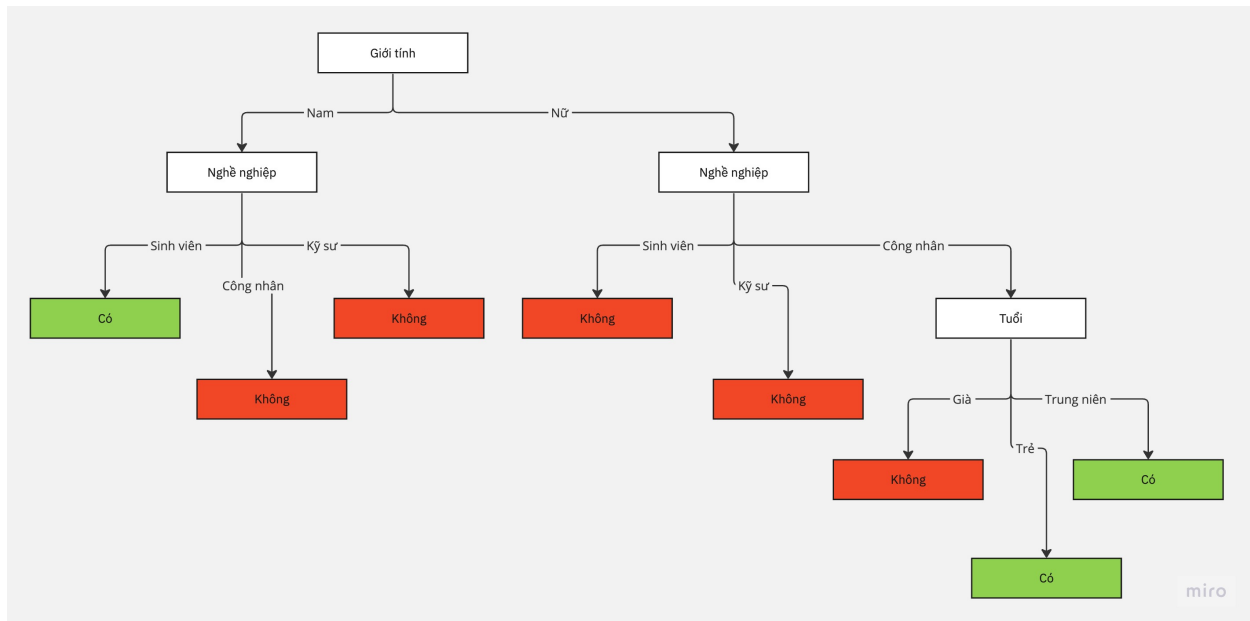


Tiếp đó ta tính tiếp tới Giới tính bằng Nữ

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.

### Câu 3. Sử dụng Naive Bayes để dự đoán.

Tuổi	Thu nhập	Là sinh viên	Xếp hạng tín dụng	Mua PC
<= 30	Cao	Sai	Khá	Không
<= 30	Cao	Sai	Tốt	Không
31..40	Cao	Sai	Khá	Có
>40	Trung bình	Sai	Khá	Có
>40	Thấp	Đúng	Khá	Có
>40	Thấp	Đúng	Tốt	Không
31..40	Thấp	Đúng	Tốt	Có
<= 30	Trung bình	Sai	Khá	Không



Tuổi	Thu nhập	Là sinh viên	Xếp hạng tín dụng	Mua PC
<= 30	Thấp	Đúng	Khá	Có
>40	Trung bình	Đúng	Khá	Có
<= 30	Trung bình	Đúng	Tốt	Có
31..40	Trung bình	Sai	Tốt	Có
31..40	Cao	Đúng	Khá	Có
>40	Trung bình	Sai	Tốt	Không

## Input 1.

Tuổi	Thu nhập	Là sinh viên	Xếp hạng tín dụng	Mua PC
31..40	Trung bình	Đúng	Khá	?

$$P(Y = \text{Có}|X) \sim P(X|Y = \text{Có}) * P(Y = \text{Có})$$

$$P(Y = \text{Không}|X) \sim P(X|Y = \text{Không}) * P(Y = \text{Không})$$

$$P(Y = \text{Có}) = 9/14$$

$$P(Y = \text{Không}) = 5/14$$

	Tuổi	Thu nhập	Là sinh viên	Xếp hạng tín dụng	Mua PC
	31..40	Trung bình	Đúng	Khá	P(Y X)
Y = Có	4/9	4/9	6/9	6/9	
Y = Không	0/5	2/5	1/5	2/5	

Vì có biến bằng 0 nên chúng ta sẽ làm tròn Laplace

$$P(Y = \text{Có}) = 10/16$$

$$P(Y = \text{Không}) = 6/16$$

	Tuổi	Thu nhập	Là sinh viên	Xếp hạng tín dụng	Mua PC
	31..40	Trung bình	Đúng	Khá	P(Y X)
Y = Có	5/12	5/12	7/12	7/12	~ 0.0369
Y = Không	1/8	3/8	2/8	3/8	~ 0.0016

Vì Y = 'Có' cao hơn nên dự đoán của input Tuổi = 31..40, Thu nhập Trung bình, Có là sinh viên và Xếp hạng tín dụng: Khá thì **CÓ** mua PC.

## Input 2

$$P(Y = \text{Có}|X) \sim P(X|Y = \text{Có}) * P(Y = \text{Có})$$

$$P(Y = \text{Không}|X) \sim P(X|Y = \text{Không}) * P(Y = \text{Không})$$

$$P(Y = \text{Có}) = 9/14$$

$$P(Y = \text{Không}) = 5/14$$

	Tuổi	Thu nhập	Là sinh viên	Xếp hạng tín dụng	Mua PC
	<=30	Trung bình	Đúng	Khá	$P(Y X)$
Y = Có	2/9	4/9	6/9	6/9	0.0282
Y = Không	1/5	2/5	1/5	2/5	0.0023

Như vậy, với input Tuổi ≤ 30, thu nhập Trung bình, Là sinh viên và xếp hạng tín dụng khá thì khả năng vẫn sẽ là **CÓ**

## Câu 4. Vẽ cây ID3

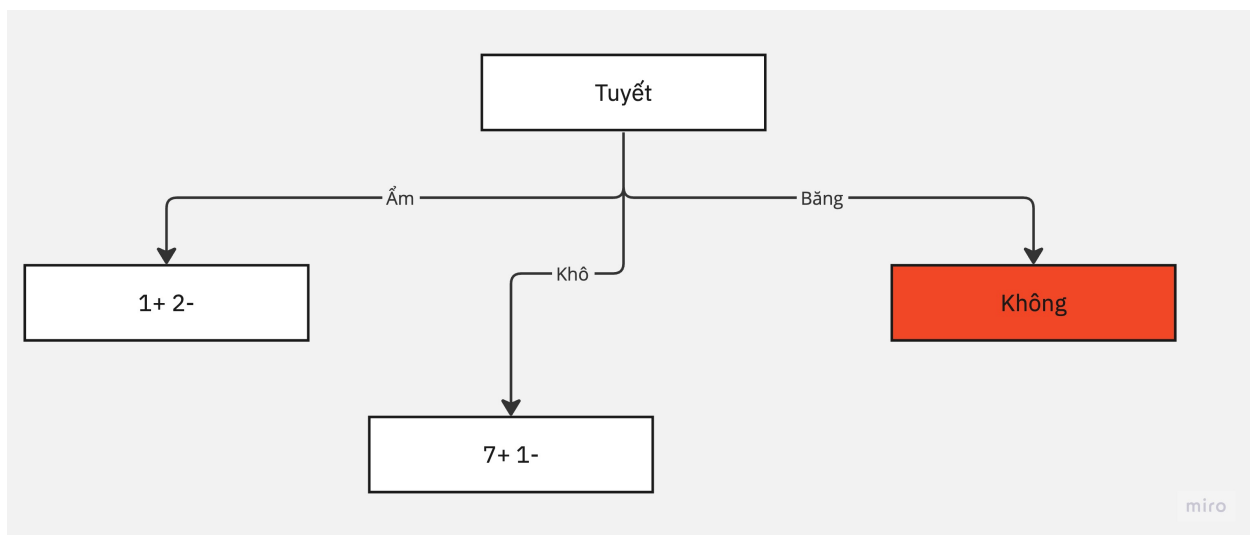
Mẫu	Tuyết	Thời tiết	Mùa	Sức khỏe	Trượt tuyết
1	Ấm	Sương mù	Vắng	Tốt	Không
2	Khô	Nắng	Vắng	Bị thương	Không
3	Khô	Nắng	Vắng	Tốt	Có
4	Khô	Nắng	Cao điểm	Tốt	Có
5	Khô	Nắng	Vừa phải	Tốt	Có
6	Băng	Gió	Cao điểm	Mệt mỏi	Không
7	Ấm	Nắng	Vắng	Tốt	Có
8	Băng	Sương mù	Vừa phải	Tốt	Không
9	Khô	Gió	Vắng	Tốt	Có
10	Khô	Gió	Vắng	Tốt	Có
11	Khô	Sương mù	Vắng	Tốt	Có
12	Khô	Sương mù	Vắng	Tốt	Có
13	Ấm	Nắng	Vừa phải	Tốt	Có
14	Băng	Sương mù	Vắng	Bị thương	Không

Chúng ta tính Entropy cho các thuộc tính

Tuyết	AE=	0.507385932	
0.2142857143	Ấm	1+ 2-	0.9182958341

0.5714285714	Khô	7+ 1-	0.5435644432
0.2142857143	Băng	3-	0
<b>Thời tiết</b>	<b>AE=</b>	<b>0.822126786</b>	
0.4285714286	Nắng	5+ 1-	0.6500224216
0.2142857143	Gió	2+ 1-	0.9182958341
0.3571428571	Sương mù	2+ 3-	0.9709505945
<b>Mùa</b>	<b>AE=</b>	<b>0.7682062502</b>	
0.6428571429	Vắng	6+ 3-	0.6666666667
0.2142857143	Vừa phải	2+ 1-	0.9182958341
0.1428571429	Cao điểm	1+ 1-	1
<b>Sức khỏe</b>	<b>AE=</b>	<b>0.5374587709</b>	
0.7857142857	Tốt	9+ 2-	0.6840384356
0.1428571429	Bị thương	2-	0
0.07142857143	Mệt mỏi	1-	0

Vì Tuyết có Entropy thấp nhất nên ta sẽ chọn làm gốc

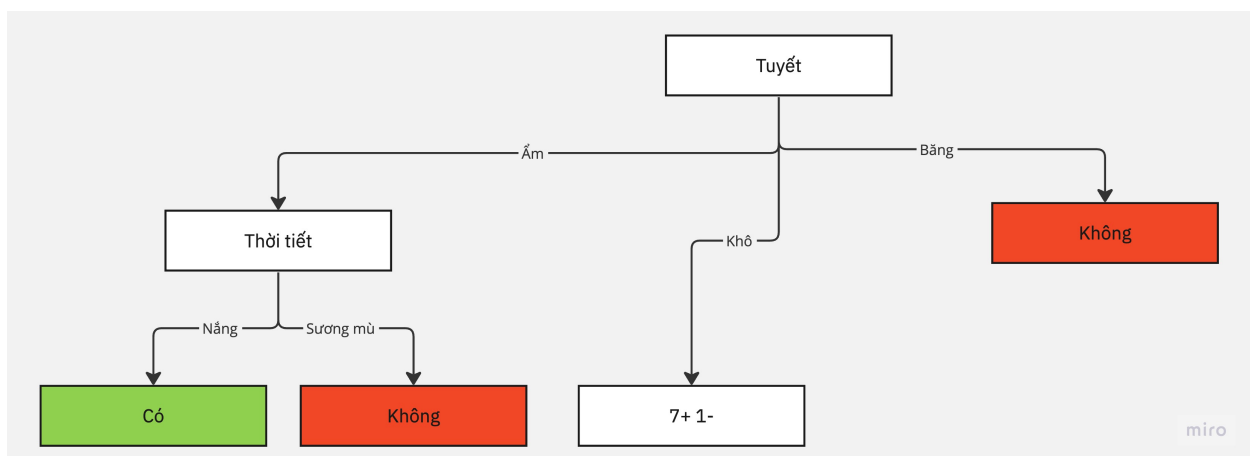


Ta tính tiếp với nhánh Tuyết = Ấm

<b>Thời tiết</b>	<b>AE=</b>	<b>0</b>	
0.6666666667	Nắng	2+	0
0	Gió	0	0

0.3333333333	Sương mù	1-	0
<b>Mùa</b>	<b>AE=</b>	<b>0.6666666667</b>	
0.6666666667	Vắng	1+ 1-	1
0.3333333333	Vừa phải	1+	0
0	Cao điểm	0	0
<b>Sức khỏe</b>	<b>AE=</b>	<b>0.9182958341</b>	
1	Tốt	2+ 1-	0.9182958341
0	Bị thương	2-	0
0	Mệt mỏi	1-	0

Như vậy ta chọn nhánh tiếp theo là “Thời tiết” cho Tuyết = Ấm

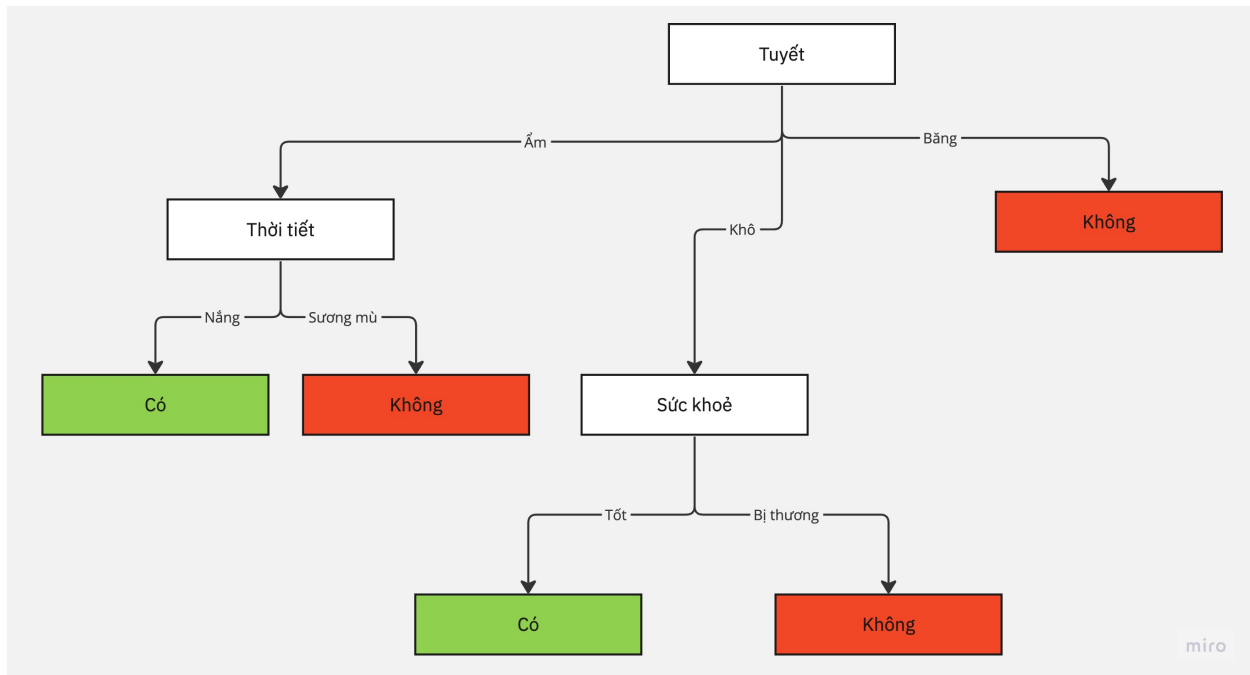


Ta tính tiếp cho nhánh Tuyết = Khô

<b>Thời tiết</b>	<b>AE=</b>	<b>0.9056390622</b>	
0.5	Nắng	3+ 1-	0.8112781245
0.25	Gió	2+	1
0.25	Sương mù	2+	1
<b>Mùa</b>	<b>AE=</b>	<b>0.4875168162</b>	
0.75	Vắng	5+ 1-	0.6500224216
0.125	Vừa phải	1+	0
0.125	Cao điểm	1+	0

Sức khỏe	AE=	0	
0.875	Tốt	7+	0
0.125	Bị thương	1-	0
0	Mệt mỏi	0	0

Như vậy ta sẽ chọn sức khỏe. Cây hoàn chỉnh của chúng ta sẽ như sau



Từ đó, kết quả dự đoán như sau

A. (Ấm, Gió, Cao điểm, Mệt mỏi) → Vì thời tiết = Gió không có nên chúng ta sẽ chọn đại là Có.

B. (Khô, Sương Mù, Vừa phải, Bị Thương) → Kết quả sẽ là KHÔNG.

## Câu 5. Dùng thuật toán ILA để rút ra tập luật cho bảng dữ liệu sau.

Weather	Trees	Temperature	Season (Class)
Rainy	Yellow	Average	Autumn
Rainy	Leafless	Low	Winter
Snowy	Leafless	Low	Winter
Sunny	Leafless	Low	Winter
Rainy	Leafless	Average	Autumn
Rainy	Green	High	Summer

Rainy	Green	Average	Spring
Sunny	Green	Average	Spring
Sunny	Green	High	Summer
Sunny	Yellow	Average	Autumn
Snowy	Green	Low	Winter

Đầu tiên chúng ta chia thành những bảng con

Weather	Trees	Temperature	Season (Class)
Rainy	Yellow	Average	Autumn
Rainy	Leafless	Average	Autumn
Sunny	Yellow	Average	Autumn

Weather	Trees	Temperature	Season (Class)
Rainy	Green	Average	Spring
Sunny	Green	Average	Spring

Weather	Trees	Temperature	Season (Class)
Rainy	Green	High	Summer
Sunny	Green	High	Summer

Weather	Trees	Temperature	Season (Class)
Rainy	Leafless	Low	Winter
Snowy	Leafless	Low	Winter
Sunny	Leafless	Low	Winter
Snowy	Green	Low	Winter

Ta thấy Yellow chỉ xuất hiện ở Autumn nên ta sẽ có rule đầu tiên

```
IF 'Trees' == 'Yellow'
THEN 'Season' = 'Autumn';
```

Nhưng rule trên mới chỉ cover được 2 dòng nên chúng ta cần thêm 1 rule nữa. Ta thấy cặp Trees = Leafless và Temp = Average chỉ xuất hiện ở Autumn nên chúng ta sẽ lấy điều kiện này tiếp

```
IF 'Trees' == 'Leafless'
THEN 'Season' = 'Autumn';
```

Với Spring, ta thấy với Trees = Green và Temp = Average thì chỉ xuất hiện ở Spring nên ta sẽ sử dụng điều kiện này.

```
IF 'Trees' == 'Green' AND 'Temperature' == 'Average'  
THEN 'Season' = 'Spring';
```

Tương tự ta thấy khi Trees = Green và Temp = High thì sẽ là Summer

```
IF 'Trees' == 'Green' AND 'Temperature' == 'High'  
THEN 'Season' = 'Summer';
```

Và với Temperature = Low thì sẽ là Winter

```
IF 'Temperature' == 'Low'  
THEN 'Season' = 'Summer';
```