

Câu 1: Cho CSDL như sau:

	Ngày	Quang cảnh	Nhiệt độ	Độ ẩm	Gió	Chơi Tennis?
X	1	Nắng	Nóng	Cao ✓	Thấp ✓	Không đi
✗	2	Âm u	Nóng	Cao	Thấp	Đi
✗	3	Mưa	Lạnh	TB	Cao	Không đi
✗	4	Âm u	TB	Cao	Thấp	Đi
✗	5	Mưa	TB	Cao	Thấp	Đi
✗	6	Mưa	Lạnh	TB	Thấp	Đi
X	7	Nắng	TB	Cao ✓	Thấp ✓	Không đi
X	8	Nắng	Lạnh	TB ✗	Thấp ✓	Đi
✗	9	Âm u	Lạnh	TB	Cao	Đi
✗	10	Mưa	TB	TB	Thấp	Đi
X	11	Nắng	Nóng	Cao ✓	Cao ✗	Không đi
X	12	Nắng	TB	TB ✗	Cao ✗	Đi
✗	13	Âm u	TB	Cao	Cao	Đi
✗	14	Âm u	Nóng	TB	Thấp	Đi
	15	Mưa	TB	Cao	Cao	?

a. Tìm các luật phân lớp dựa trên cây quyết định với độ đo **Information Gain**:

$$Gain(S, A) = Entropy(S) - \sum_{x \in Value(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Trong đó:

- Value(A) là tập tất cả các giá trị có thể có đối với thuộc tính A và S_v là tập con của S mà A có giá trị là v

- Với S bao gồm c lớp, thì Entropy của S được tính bằng công thức sau:

$$Entropy(S) = - \sum_{i=1}^c P_i \log_2 P_i$$

Ở đây p_i là tỉ lệ của các mẫu thuộc lớp i trong tập S.

b. Cho biết lớp của mẫu #15?

a)

$$Entropy(S) = -4/14 * \log_2(4/14) - 10/14 * \log_2(10/14) = 0.86$$

Với từng thuộc tính, ta có độ đo Information Gain tương ứng là:

$$Gain(S, \text{Quang cảnh}) = 0.86 - (5/14 * (-2/5 * \log_2(2/5) - 3/5 * \log_2(3/5)) + 5/14 * 0 + 4/14 * (-3/4 * \log_2(3/4) - 1/4 * \log_2(1/4))) = 0.28$$

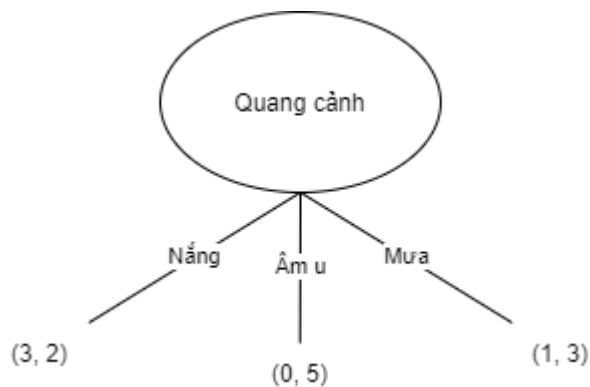
$$Gain(S, \text{Nhiệt độ}) = 0.86 - (4/14 * (-2/4 * \log_2(2/4) - 2/4 * \log_2(2/4)) + 4/14 * (-3/4 * \log_2(3/4) - 1/4 * \log_2(1/4))) = 0.06$$

$$Gain(S, \text{Độ ẩm}) = 0.86 - (7/14 * (-4/7 * \log_2(4/7) - 3/7 * \log_2(3/7)) + 7/14 * (-6/7 * \log_2(6/7) - 1/7 * \log_2(1/7))) = 0.07$$

$$Gain(S, \text{Gió}) = 0.86 - (9/14 * (-7/9 * \log_2(7/9) - 2/9 * \log_2(2/9)) + 5/14 * (-3/5 * \log_2(3/5) - 2/5 * \log_2(2/5))) = 0.02$$

Vì thuộc tính Quang cảnh có Information Gain lớn nhất nên được chọn để làm nút gốc cho cây quyết định.

Với bộ số (số lượng không đi, số lượng đi), ta có cây quyết định như sau:



Ta tiếp tục xét với hai nhánh có nút lá chưa đồng nhất:

- Nhánh nắng:

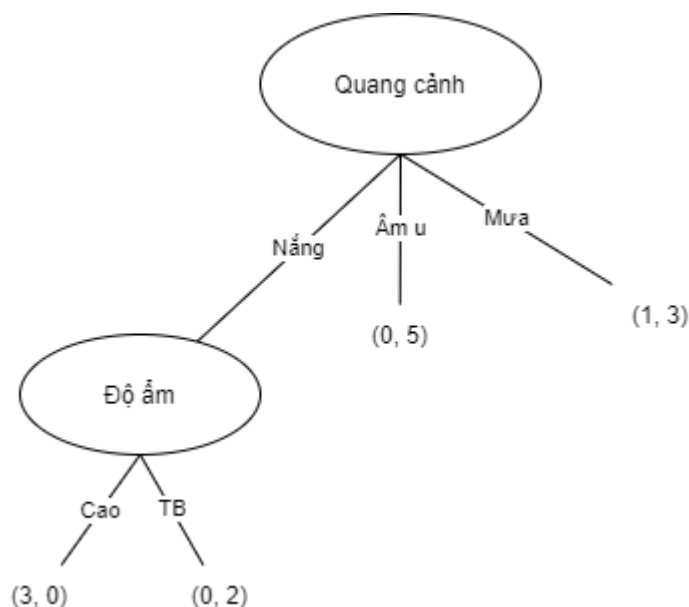
$$\text{Entropy}(S_{\text{nắng}}) = -\frac{3}{5} \cdot \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \cdot \log_2\left(\frac{2}{5}\right) = 0.97$$

$$\text{Gain}(S_{\text{nắng}}, \text{Nhiệt độ}) = 0.97 - \left(\frac{2}{5} \cdot 0 + \frac{3}{5} \cdot \left(-\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right)\right) + \frac{1}{5} \cdot 0\right) = 0.57$$

$$\text{Gain}(S_{\text{nắng}}, \text{Độ ẩm}) = 0.97 - \left(\frac{3}{5} \cdot 0 + \frac{2}{5} \cdot 0\right) = 0.97$$

$$\text{Gain}(S_{\text{nắng}}, \text{Gió}) = 0.97 - \left(\frac{3}{5} \cdot \left(-\frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right)\right) + \frac{2}{5} \cdot \left(-\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right)\right)\right) = 0.02$$

Vì thuộc tính Độ ẩm có Information Gain lớn nhất nên được chọn để làm nút tiếp theo nhánh nắng:



- Nhánh mưa:

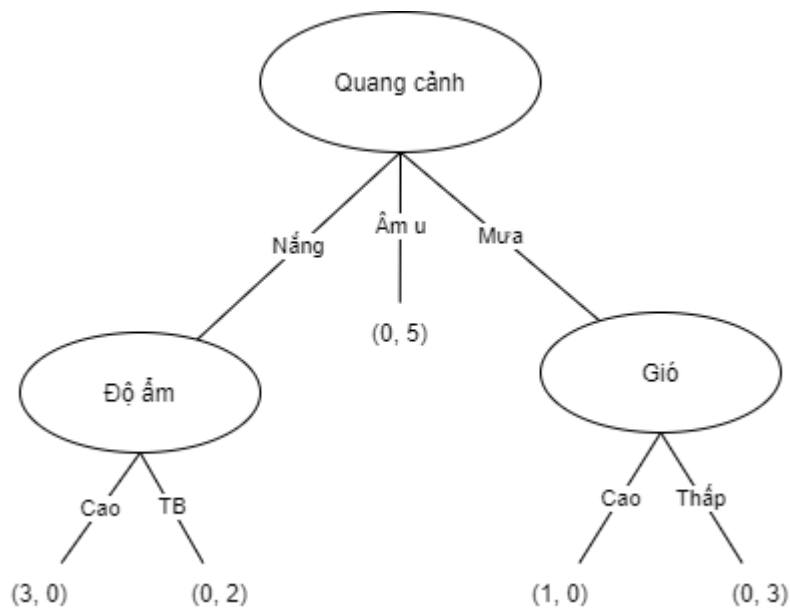
$$\text{Entropy}(S_{\text{mưa}}) = -\frac{1}{4} \cdot \log_2\left(\frac{1}{4}\right) - \frac{3}{4} \cdot \log_2\left(\frac{3}{4}\right) = 0.81$$

$$\text{Gain}(S_{\text{mưa}}, \text{Nhiệt độ}) = 0.81 - \left(\frac{2}{4} \cdot \left(-\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right)\right) + \frac{2}{4} \cdot 0\right) = 0.31$$

$$\text{Gain}(S_{\text{mưa}}, \text{Độ ẩm}) = 0.81 - \left(\frac{3}{4} \cdot \left(-\frac{2}{3} \cdot \log_2\left(\frac{2}{3}\right) - \frac{1}{3} \cdot \log_2\left(\frac{1}{3}\right)\right) + \frac{1}{4} \cdot 0\right) = 0.12$$

$$\text{Gain}(S_{\text{mưa}}, \text{Gió}) = 0.81 - \left(\frac{3}{4} \cdot 0 + \frac{1}{4} \cdot 0\right) = 0.81$$

Vì thuộc tính Gió có Information Gain lớn nhất nên được chọn để làm nút tiếp theo nhánh mưa:



Vì tất cả các nốt lá đều đồng nhất nên ta rút ra luật:

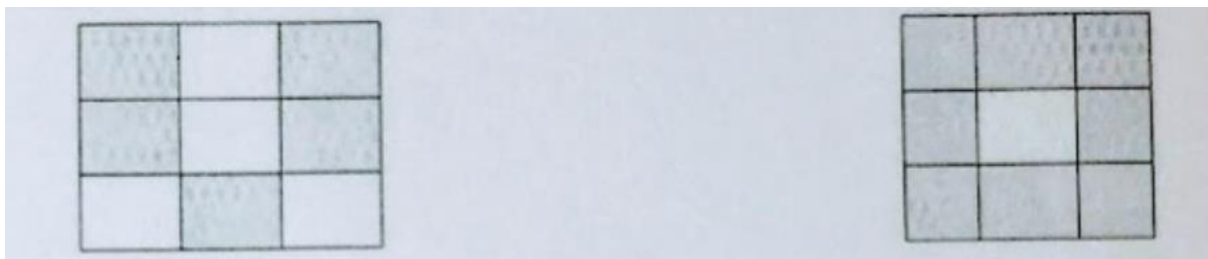
- Nếu quang cảnh nắng, độ ẩm cao thì không đi chơi tennis.
- Nếu quang cảnh nắng, độ ẩm trung bình thì đi chơi tennis.
- Nếu quang cảnh âm u thì đi chơi tennis.
- Nếu quang cảnh mưa, gió cao thì không đi chơi tennis.
- Nếu quang cảnh mưa, gió thấp thì đi chơi tennis.

b)

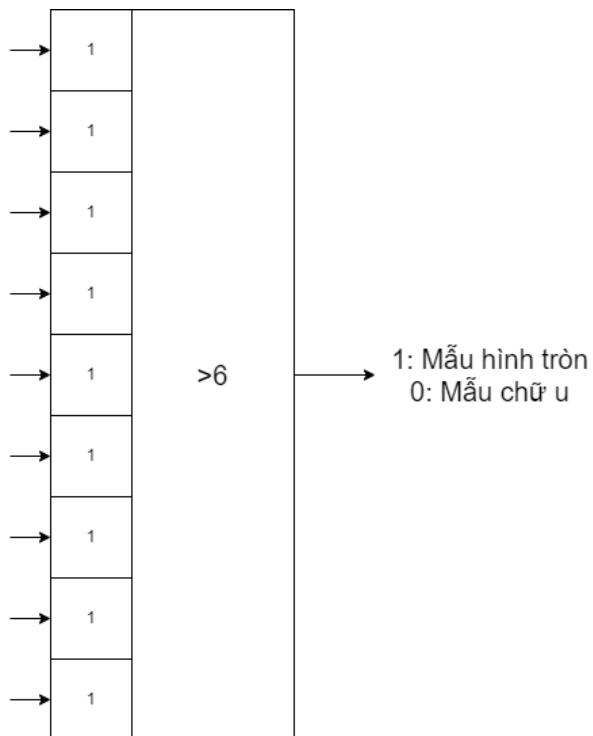
Dựa trên luật vừa rút ra, ta có:

- Mẫu 15 có quang cảnh mưa, gió cao nên sẽ thuộc lớp không đi chơi tennis.

Câu 2. Thiết kế một mạng nơ-ron nhân tạo có thể phân biệt mẫu nào trong số hai mẫu sau đây nằm trong khung nhìn của chúng.



Gọi hình thứ nhất là mẫu chữ U, hình thứ hai là mẫu chữ O, khi đó mạng nơ-ron phân biệt với 9 đầu vào tương ứng với 9 ô vuông của mẫu là:



Câu 3: Cho một cơ sở tri thức được biểu diễn dưới dạng logic mệnh đề như sau:

$$KB = \{A, B, A \vee C, K \wedge E \leftrightarrow A \wedge B, \neg C \rightarrow D, E \vee F \rightarrow \neg D\}$$

Kiểm tra các câu sau có rút ra được từ tập cơ sở trên hay không

a) $B \wedge C$?

b) $C \vee E \rightarrow F \wedge B$?

a)

Bước	Công thức	Nguồn gốc
1	A	Cho trước
2	B	Cho trước
3	$A \vee C$	Cho trước
4	$K \wedge E \leftrightarrow A \wedge B$	Cho trước
5	$\neg C \rightarrow D$	Cho trước
6	$E \vee F \rightarrow \neg D$	Cho trước
7	$A \wedge B$	1,2 And-Introduction
8	$A \wedge B \rightarrow K \wedge E$	4 Biconditional Elimination

9	$K \wedge E$	7, 8 Modus Ponens
10	E	9 And-Elimination
11	$E \vee F$	10 Or-Elimination
12	$\sim D$	11, 6 Modus Ponens
13	C	12, 5 Modus Tolens
14	$B \wedge C$	13, 2 And-Introduction

b)

Xét trường hợp $A = B = C = E = \text{True}$, $D = F = \text{False}$, khi đó ta có:

Công thức	Giá trị	Nguồn gốc
A	True	KB
B	True	KB
$A \vee C$	True	KB
$K \wedge E \leftrightarrow A \wedge B$	True	KB
$\sim C \rightarrow D$	True	KB
$E \vee F \rightarrow \sim D$	True	KB
$C \vee E \rightarrow F \wedge B$	False	Kết luận

Vì xảy ra trường hợp $KB = \text{True}$ và Kết luận là False nên $KB \rightarrow \text{Kết luận}$ là không hợp lệ, do đó không thể $C \vee E \rightarrow F \wedge B$ không thể rút ra từ tập cơ sở KB.