

# Thuật toán Quinlan

- Thuật toán Quinlan: là thuật toán học theo quy nạp dùng luật, đa mục tiêu. Nó được phát triển bởi John Ross Quinlan – nhà khoa học máy tính – đưa ra năm 1979 với ý tưởng: chọn thuộc tính quan trọng nhất để tạo cây quyết định.

# Ý tưởng thuật toán Quinlan

- Cho một bảng quan sát (cơ sở dữ liệu) là một tập hợp các mẫu với các thuộc tính nhất định của các đối tượng nào đó.
- Sử dụng một độ đo để định lượng và đề ra một tiêu chuẩn nhằm chọn lựa một thuộc tính mang tính chất phân loại để phân bảng này thành các bảng con nhỏ hơn. Từ các bảng con này dễ dàng phân tích tìm ra qui luật chung.
- Từ đó thiết lập cây quyết định cho thấy thứ tự của thuộc tính đang xét.
- Tìm cây quyết định, xây dựng bộ luật, và đưa ra quyết định một số mẫu cụ thể.

VD1

Cho cơ sở dữ liệu gồm các mẫu sau:

Mẫu	Thời tiết	Lá cây	Nhiệt độ	Quyết định (mùa)
1	Mưa	Rụng	Thấp	Đông
2	Nắng	Xanh	Trung bình	Xuân
3	Nắng	Vàng	Trung bình	Thu
4	Nắng	Xanh	Cao	Hè
5	Nắng	Rụng	Thấp	Đông
6	Tuyết	Vàng	Thấp	Đông
7	Mưa	Rụng	Trung bình	Thu
8	Mưa	Xanh	Cao	Hè
9	Tuyết	Xanh	Thấp	Đông
10	Tuyết	Rụng	Thấp	Đông
11	Mưa	Vàng	Trung bình	Thu
12	Mưa	Xanh	Trung bình	Xuân
x	Mưa	Vàng	Cao	?
y	Tuyết	Rụng	Trung bình	?
z	Tuyết	Vàng	Trung bình	?

- Từ mẫu 1 đến mẫu 12 hãy rút ra bộ luật cho sự quyết định Mùa theo thuật toán Quinlan.
- Áp dụng cho biết kết quả các mẫu x,y,z.

# Hướng dẫn giải

Bước 0: Gọi vectơ độ đo  $v=(\text{Xuân}, \text{Hè}, \text{Thu}, \text{Đông})$

Bước 1: Tính vectơ độ đo của các thuộc tính ở CSDL ban đầu

+Thuộc tính **thời tiết**

$$V_{(\text{thời tiết}=\text{nắng})} = (1/4, 1/4, 1/4, 1/4);$$

$$V_{(\text{thời tiết}=\text{tuyết})} = (0, 0, 0, 1);$$

$$V_{(\text{thời tiết}=\text{mưa})} = (1/5, 1/5, 2/5, 1/5);$$

+Thuộc tính **lá cây**

$$V_{(\text{lá cây}=\text{Vàng})} = (0, 0, 2/3, 1/3);$$

$$V_{(\text{lá cây}=\text{Xanh})} = (2/5, 2/5, 0, 1/5);$$

$$V_{(\text{lá cây}=\text{Rụng})} = (0, 0, 1/4, 3/4);$$

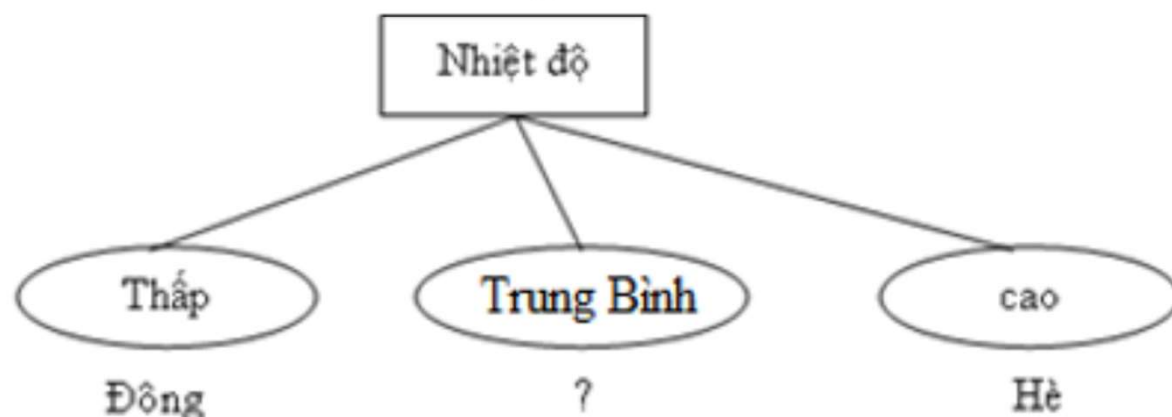
+Thuộc tính Nhiệt độ

$$V_{(\text{nhiệt độ}=\text{Trung bình})} = (2/5, 0, 3/5, 0);$$

$$V_{(\text{nhiệt độ}=\text{thấp})} = (0, 0, 0, 1);$$

$$V_{(\text{nhiệt độ}=\text{cao})} = (0, 1, 0, 0);$$

Chọn thuộc tính Nhiệt độ làm thuộc tính phân loại:



Bước 2: CSDL ứng với Nhiệt độ bằng Trung bình

#	Thời tiết	Lá cây	Quyết định (Mùa)
2	Nắng	Xanh	Xuân
3	Nắng	Vàng	Thu
7	Mưa	Rụng	Thu
11	Mưa	Vàng	Thu
12	Mưa	Xanh	Xuân

Tính vectơ độ đo của các thuộc tính

+Thuộc tính **thời tiết**

$$V_{(\text{thời tiết}=\text{nắng})} = (1/2, 0, 1/2, 0);$$

$$V_{(\text{thời tiết}=\text{mưa})} = (1/3, 0, 2/3, 0);$$

+Thuộc tính lá cây

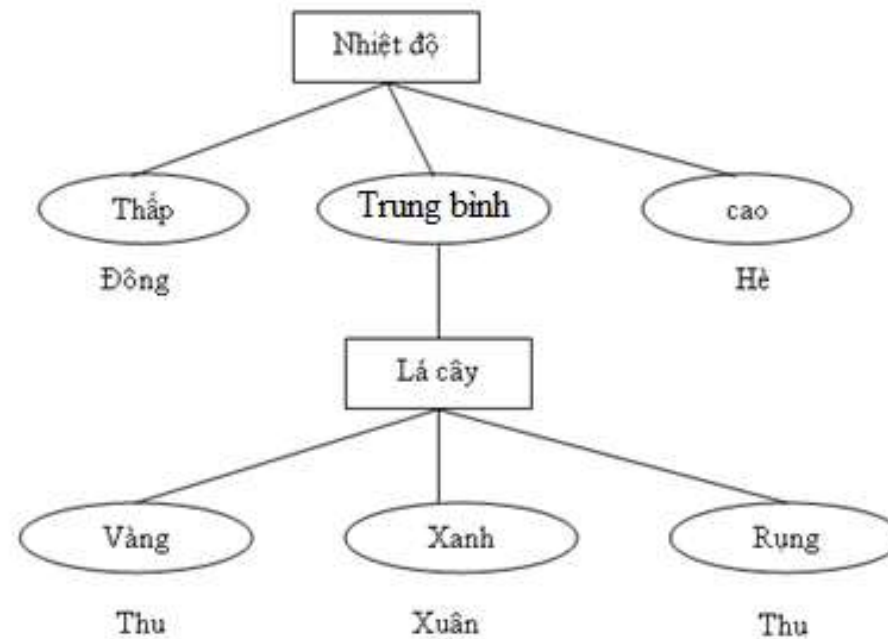
$$V_{(\text{lá cây}=\text{Vàng})} = (0, \quad 0, \quad 1, \quad 0);$$

$$V_{(\text{lá cây}=\text{Xanh})} = (1, \quad 0, \quad 0, \quad 0);$$

$$V_{(\text{lá cây}=\text{Rụng})} = (0, \quad 0, \quad 1, \quad 0);$$

Chọn thuộc tính Lá cây làm thuộc tính phân loại

Từ các bước trên, ta có cây quyết định sau:



**Bước 3: Tập luật** Từ cây quyết định trên, ta có tập luật sau:

Luật 1: Nếu nhiệt độ thấp thì mùa Đông

Luật 2: Nếu nhiệt độ cao thì mùa Hè

Luật 3: Nếu nhiệt độ trung bình và lá cây Xanh thì mùa Xuân

Luật 4: Nếu nhiệt độ trung bình và lá cây màu Vàng hoặc lá rụng thì mùa Thu



#	Thời tiết	Lá cây	Nhiệt độ	Quyết định (Mùa)	Luật
x	Mưa	Vàng	Cao	Hè	Luật 2
y	Tuyết	Rụng	Trung bình	Thu	Luật 4
z	Tuyết	Vàng	Trung bình	Thu	Luật 4