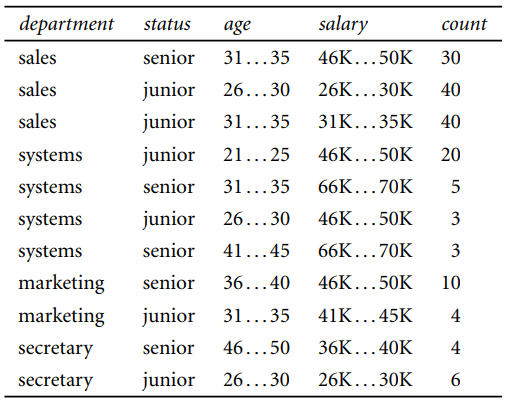
Buổi 5:

1. Trình bày ngắn gọn các bước chính của phân loại cây quyết định.

* Tính toán độ lợi thông tin của mỗi thuộc tính
* Chọn thuộc tính có độ lợi thông tin cao nhất làm nút gốc
* Phân chia tập mẫu: Chia dữ liệu học thành các tập con phân việt sao cho trong một tập có một giá trị cố định
* Xây dựng một cây có các nhanh ứng với các giá trị của thuộc tính
* Lặp lại tính trình: với mỗi nhánh, lặp lại tiến trình này từ bước 1
* Dừng tiến trình: Tại mỗi lần lặp, một thuộc tính bị loại trừ. Tiến trình dừng khi không còn thuộc tính để xem xét, hoặc khi tất cả dữ liệu đang được xem xét trong nhánh con có cùng giá trị kết luận

1. Điều quan trọng là phải tính toán độ phức tạp trong trường hợp xấu nhất của thuật toán cây quyết định. Cho tập dữ liệu D, số thuộc tính n, và số bộ giá trị huấn luyện | D |, cho thấy rằng chi phí tính toán để xây dựng cây quyết định tối đa là n × | D | × log (| D |).
2. Bảng bên dưới bao gồm dữ liệu huấn luyện từ cơ sở dữ liệu nhân viên. Dữ liệu đã được tổng quát hóa. Ví dụ: “31 ... 35” cho độ tuổi đại diện cho độ tuổi từ 31 đến 35. Đối với một hàng nhất định, số lượng biểu thị số lượng bộ dữ liệu có các giá trị cho phòng ban, trạng thái, tuổi và mức lương được cung cấp trong hàng đó . Xây dựng cây quyết định



1. Hoàn thiện cây quyết định (ID-3) cho bài toán chơi bóng chày

Table

Description automatically generated

* Excel