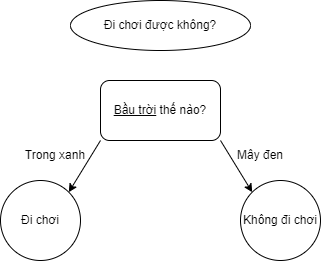
Decision Tree: ID3

Thuật toán ID3 (Iterative Dichotomiser 3) là một kỹ thuật xây dựng cây quyết định dựa trên cơ chế chọn lọc các đặc trưng để tối ưu hóa thông tin nhận vào (Information Gain). Thuật toán này thường được dùng trong phân lớp (classification).

Ý tưởng của thuật toán là mình cần tìm ra đặc trưng tốt nhất làm nút gốc dựa trên một tiêu chuẩn nhất định. Với một đặc trưng được chọn, mình sẽ chia dữ liệu vào các nút con tương ứng với các giá trị của đặc trưng ấy, mà rồi tiếp tục áp dụng phương pháp này cho mỗi nút con. Việc chọn đặc trưng tốt nhất ở mỗi bước được gọi là một cách chọn tham lam. Sau tất cả thì mình muốn tìm đặc trưng có lượng thông tin nhận vào tốt nhất làm nút gốc.

Trong cây quyết định, mỗi đặc trưng trong tập dữ liệu là một *câu hỏi*, và các dữ liệu được phân chia vào các nút con tượng trưng cho *các câu trả lời* cho *câu hỏi* ấy.



*Đặc trưng “Bầu trời” và 2 lớp của nó dưới dạng cây quyết định.*

Trong cây quyết định, một phép phân chia tốt nhất là khi dữ liệu trong mỗi nút con hoàn toàn phụ thuộc vào một lớp