1. Lý thuyết
   1. ID3 (**Iterative Dichotomiser 3**): Là một thuật toán phân loại theo cách tiếp cận tham lam bằng cách chọn thuộc tính tốt nhất nhằm mang lại Information Gain (IG - lợi ích của thông tin) tối đa hoặc Entropy tối thiểu (entropy dùng để chỉ trạng thái ngẫu nhiên hoặc không có trật tự).
   2. Input: Là 1 tập DL đã được gán nhãn
   3. Output: Là 1 cây quyết định
   4. Method: Sử dụng hàm “entropy“ để đo độ tinh khiết (purity) / độ vẩn đục (impurity) của 1 phép phân chia
      1. B1: Tính entropy tại root node:

A black and white math symbols

Description automatically generated

* + 1. B2: Tính entropy tại child node:

A number of mathematical symbols

Description automatically generated with medium confidence

* + 1. B3: Định nghĩa “information gain“ dựa trên thuộc tính x:



* + 1. Tại mỗi node, thuộc tính được chọn được xác định dựa vào:

Text

Description automatically generated

* 1. Điều kiện dừng
     1. Tất cả các mẫu của nút xem xét thuộc cùng một lớp
     2. Không có thuộc tính nào để phân chia
     3. Không còn mẫu nào