Decision Trees

1. Giới thiệu về Decision Trees

* Cây quyết định được xây dựng bằng cách chia tập train thành các node riêng biệt, trong đó một nút chứa tất cả hoặc hầu hết một loại dữ liệu.
* Mỗi node là một thuộc tính và mỗi nhánh tương ứng với giá trị của thuộc tính đó.
* Cách xây dựng Decision Tree : một cây quyết định có thể thể được xây dựng bằng cách xem xét từng thuộc tính một

+ Chọn một thuộc tính từ tập dữ liệu.

+ Tính toán độ quan trọng của thuộc tính trong việc phân lớp dữ liệu ( tính entropy)

+ Phân chia dữ liệu dựa trên giá trị của thuộc tính tốt nhất, sau đó đi đến từng nhánh và lặp lại

1. Xây dựng Decision Tree

* DS được xây dựng bằng cách sử dụng phân vùng đệ quy để phân loại dữ liệu
* Phương thức sử dụng phân vùng đệ quy để phân loại thành các các phân khúc và giảm thiểu tạp chất ở mỗi bước.
* Tạp chất của các nút được tính bằng entropy của dữ liệu trong nút
* Entropy là lượng rối loạn thông tin hoặc lượng ngẫu nhiên trong dữ liệu.
* Entropy trong nút phụ thuộc vào có bao nhiêu dữ liệu ngẫu nhiên trong nút đó và tính cho mỗi nút.
* Trong các DS, chúng ta tìm kiếm những cây có entropy nhỏ nhất trong các nút của chúng.
* Entropy được sử dụng để tính độ đồng nhất của các mẫu trong nút đó

+ Nếu các mẫu hoàn toàn đồng nhất , entropy = 0

+ Nếu các mẫu được chia đều thì entropy = 1