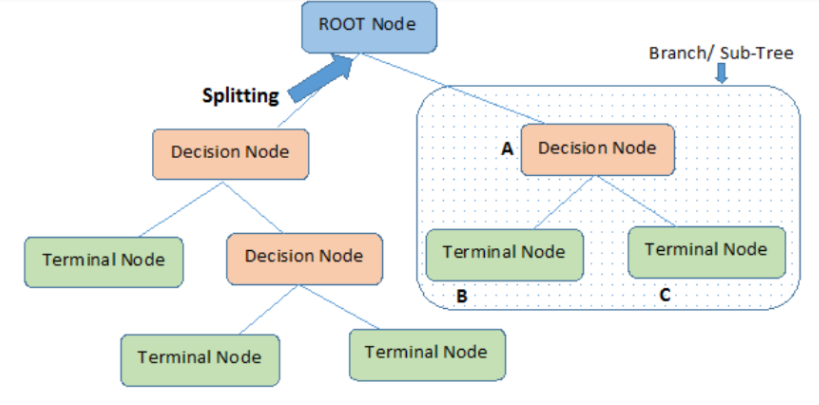
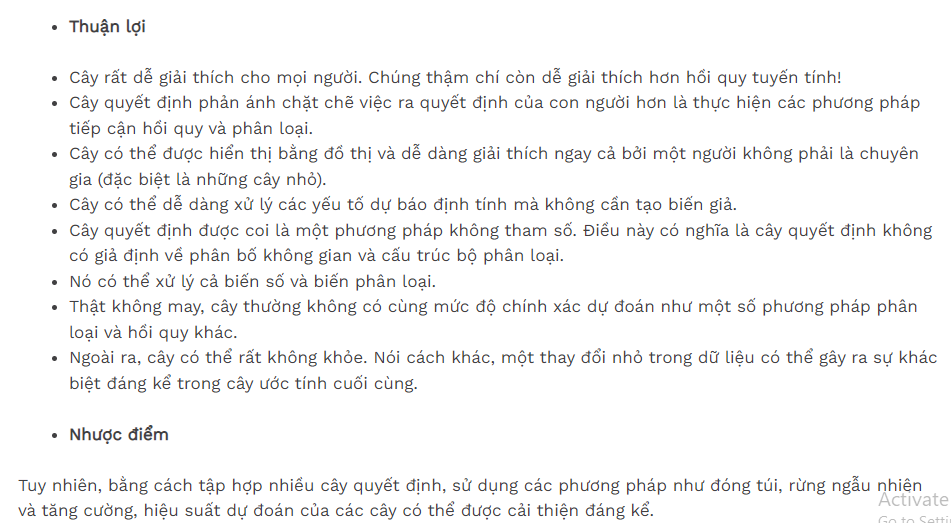
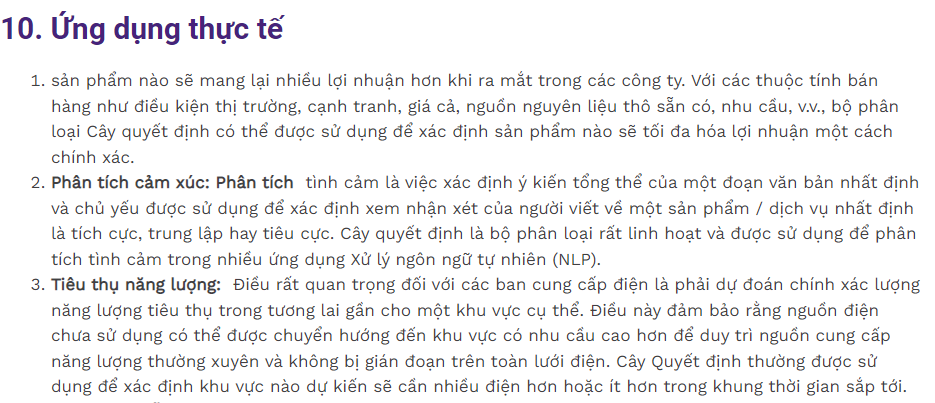


Cắt node: để giảm overfiting



1 ví dụ cây con(sub-tree) ở trên





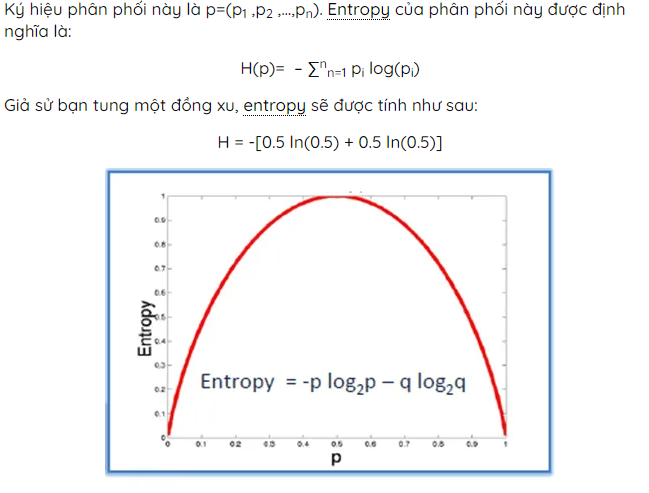
ID3 sử dụng [Entropy](https://trituenhantao.io/tu-dien-thuat-ngu/entropy/) và Information Gain.

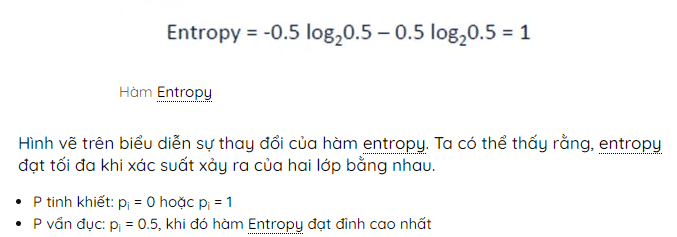
hai hệ số phổ biến là **Information Gain** và **Gain Ratio**

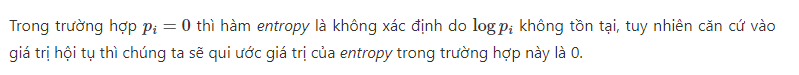
Với một phân phối xác suất của một biến rời rạc x có thể nhận n giá trị khác nhau x1,x2,…,xn.

Giả sử rằng xác suất để x nhận các giá trị này là pi=p(x=xi)

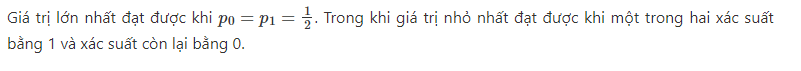
Ký hiệu phân phối này là p=(p1 ,p2 ,…,pn). [Entropy](https://trituenhantao.io/tu-dien-thuat-ngu/entropy/) của phân phối này được định nghĩa là:







[x`](https://trituenhantao.io/kien-thuc/decision-tree/)



Khảo sát hàm entropy chúng ta sẽ nhận thấy đây là một hàm không âm