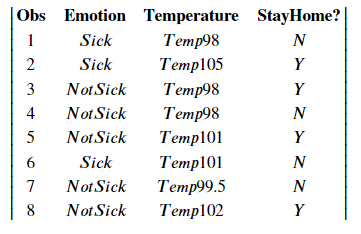
là một trong những phương pháp được sử dụng trong các thuật toán cây quyết định để quyết định sự phân tách tối ưu từ nút gốc và các lần phân tách tiếp theo.

(Trước khi tiếp tục, bạn có thể muốn xem lại [Việc đưa ra Quyết định với Cây](https://medium.com/@StevenLoaiza/making-decisions-with-trees-559c8db5af59) )

Để đặt nó vào ngữ cảnh, cây quyết định đang cố gắng tạo ra các câu hỏi tuần tự để nó phân vùng dữ liệu thành các nhóm nhỏ hơn. Sau khi phân vùng hoàn tất, một quyết định dự đoán sẽ được thực hiện tại nút đầu cuối này (dựa trên tần số).





Giả sử chúng ta có một lis t quan sát, cho biết nếu một người quyết định ở nhà đi làm. Chúng tôi cũng có hai tính năng, đó là nếu chúng bị ốm và nhiệt độ của chúng.

Chúng ta cần chọn tính năng, cảm xúc hoặc nhiệt độ, để chia nhỏ dữ liệu. Một biện pháp tạp chất Gini sẽ giúp chúng tôi đưa ra quyết định này.

**Def:** Gini Impurity cho chúng ta biết xác suất phân loại sai một quan sát là bao nhiêu.

Lưu ý rằng Gini càng thấp thì sự phân chia càng tốt. Nói cách khác, khả năng phân loại sai càng thấp.