### ****Tổng quan về hệ thống gợi ý****

Hệ thống gợi ý (Recommender systems hoặc Recommendation systems) là một dạng của hệ hỗ trợ ra quyết định, cung cấp giải pháp mang tính cá nhân hóa mà không phải trải qua quá trình tìm kiếm phức tạp. Hệ gợi ý học từ người dùng và gợi ý các sản phẩm tốt nhất trong số các sản phẩm phù hợp.

Hệ thống gợi ý sử dụng các tri thức về sản phẩm, các tri thức của chuyên gia hay tri thức khai phá học được từ hành vi con người dùng để đưa ra các gợi ý về sản phẩm mà họ thích trong hàng ngàn hàng vạn sản phẩm có trong hệ thống. Các website thương mại điện tử, ví dụ như sách, phim, nhạc, báo...sử dụng hệ thống gợi ý để cung cấp các thông tin giúp cho người sử dụng quyết định sẽ lựa chọn sản phẩm nào. Các sản phẩm được gợi ý dựa trên số lượng sản phẩm đó đã được bán, dựa trên các thông tin cá nhân của người sử dụng, dựa trên sự phân tích hành vi mua hàng trước đó của người sử dụng để đưa ra các dự đoán về hành vi mua hàng trong tương lai của chính khách hàng đó. Các dạng gợi ý bao gồm: gợi ý các sản phẩm tới người tiêu dùng, các thông tin sản phẩm mang tính cá nhân hóa, tổng kết các ý kiến cộng đồng, và cung cấp các chia sẻ, các phê bình, đánh giá mang tính cộng đồng liên quan tới yêu cầu, mục đích của người sử dụng đó.

### **Thu thập thông tin – Collecting Information**

Hệ thống gợi ý hoạt động bằng cách thu thập thông tin – ghi nhận những gì bạn yêu cầu – chẳng hạn như những bộ phim bạn xem trên ứng dụng xem phim, xếp hạng và đánh giá mà bạn đã gửi, những sản phẩm bạn đã cho vào giỏ hàng, những từ khóa mà bạn đã tìm kiếm và các hành động khác bạn đã thực hiện trong quá khứ.

Có lẽ quan trọng hơn, họ có thể theo dõi các lựa chọn bạn đã thực hiện: Những gì bạn click vào và cách bạn duyệt qua các trang web. Thời gian bạn xem một bộ phim nào đó mất bao lâu, chẳng hạn. Hoặc quảng cáo nào bạn click vào hoặc bạn bè của bạn tương tác.

Tất cả thông tin này được truyền vào các trung tâm dữ liệu khổng lồ và được tổng hợp thành các bảng dữ liệu (table) đa chiều, phức tạp và nhanh chóng phình to kích thước.

Chúng có thể lớn hàng trăm terabyte và cứ phát triển liên tục.

Chúng quá lớn như vậy không phải vì một lượng lớn dữ liệu được thu thập từ từng cá nhân, mà bởi vì chỉ một số ít dữ liệu được thu thập từ rất rất nhiều cá nhân.

Nói cách khác, các bảng dữ liệu này rất thưa tớt – hầu hết thông tin, hầu hết các dịch vụ này có trên hầu hết chúng ta đều gần như trống rỗng.

Nhưng tổng hợp lại, các bảng này chứa rất nhiều thông tin về sở thích của một số lượng rất lớn các cá nhân.

Và điều đó giúp các công ty đưa ra quyết định thông minh về những loại người dùng nhất định có thể sẽ thích.

### **Content Filtering, Collaborative Filtering**

Mặc dù có một số lượng lớn các thuật toán và kỹ thuật gợi ý, hầu hết đều thuộc một trong hai loại chính: Content Filtering và Collaborative Filtering.

Collaborative Filtering giúp bạn tìm thấy những gì bạn thích bằng cách tìm kiếm những người dùng tương tự như bạn.

Vì vậy, trong khi hệ thống gợi ý có thể không biết gì về sở thích âm nhạc của bạn, nếu nó biết bạn và một người dùng khác cùng thích một cuốn sách nào đó, nó có thể gợi ý cho bạn một bài hát mà người dùng kia đã thích.

Ngược lại, Content Filtering hoạt động bằng cách hiểu các tính năng, thuộc tính của từng sản phẩm.

Vì vậy, nếu một hệ thống gợi ý thấy bạn thích những bộ phim “You’ve Got Mail” và “Sleepless in Seattle”, thì nó có thể gợi ý một bộ phim khác cho bạn với sự tham gia của Tom Hanks và Meg Ryan, chẳng hạn như “Joe Versus the Volcano”.

Đó chắc chắn chỉ là những ví dụ rất đơn giản để bạn hình dung.