

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH TRƯ**ỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



DS300 HỆ KHUYẾN NGHỊ

Giới thiệu về hệ khuyến nghị

Giảng viên: ThS. Nguyễn Văn Kiệt CN. Huỳnh Văn Tín

Bộ môn Khoa học Dữ liệu Khoa Khoa học và Kỹ thuật Thông tin

Nội dung

- Bối cảnh xuất hiện các nghiên cứu
- Khái niệm hệ khuyến nghị
- Vai trò hệ khuyến nghị
- Phương pháp phổ biến và xu hướng

Những khó khắn khi đưa ra quyết định

- Cái máy ảnh nào tôi nên mua?
- □ Tôi nên đi đâu vào ngày nghỉ?
- □ Cuối tuần này tôi nên đọc cuốn sách nào?
- □ Tôi nên theo dõi ai trên mạng xã hội?
- □ Tôi nên tìm các bài báo có nội dung hấp dẫn ở đâu?
- □ Bộ phim nào hay nhất **cho gia đình** tôi tối nay?
- □ Nếu quan tâm, xem thêm hai conference tutorials
 - SIGKDD2014, Recommendation in Social Media
 - RecSys2014, Personalized Location Recommendation

Khi nào vấn đề xảy ra?

- □ Có nhiều sự lựa chọn
- Không có ưu điểm rõ ràng trong số đó.
- Chúng ta không có đủ nguồn lực để kiểm tra tất cả các tùy chọn (quá tải thông tin).
- Chúng ta không có đủ kiến thức và kinh nghiệm để lựa chọn, hoặc
 - Lười, nhưng không muốn bỏ lỡ những thứ tốt.
 - Defensive decision making

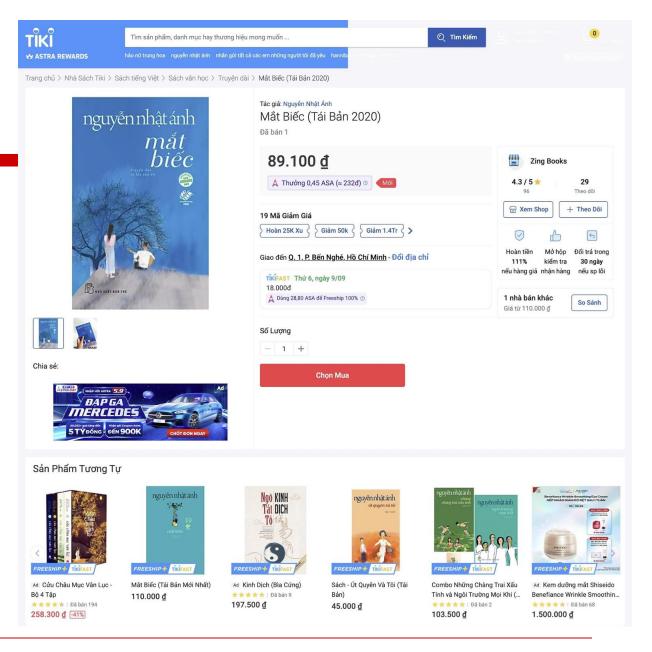
Mục tiêu của khuyến nghị:

Để đưa ra danh sách ngắn các mặt hàng phù hợp với sở thích của người dùng

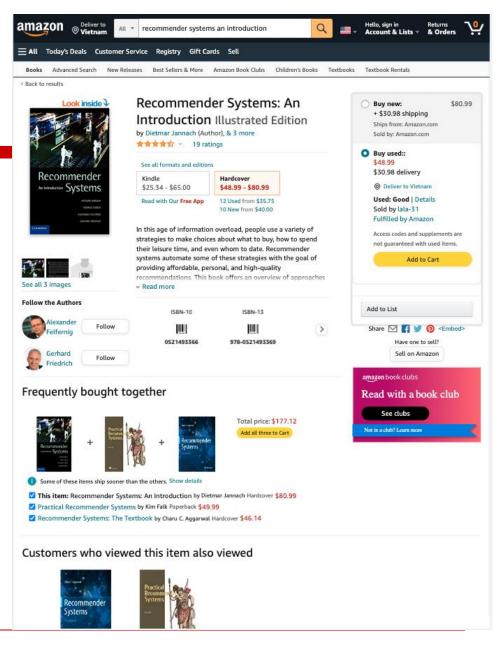
Các giải pháp thông thường cho vấn đề

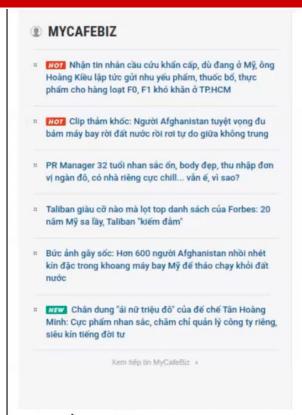
- □ Hỏi ý kiến bạn bè
- Có được thông tin từ bên thứ 3
- □ Thuê đội ngũ chuyên gia
- Làm theo số đông
 - Chọn 1 sản phẩm từ danh sách top-n
 - Sản phẩm bán chạy nhất trên Amazon
- □ Chúng ta có thể làm tự động những thế ở trên không?
 - Sử dụng một giải thuật khuyến nghị
 - Như được biết là các hệ khuyến nghị

Minh họa khuyến nghị sản phẩm trên TIKI.VN



Minh họa khuyến nghị sản phẩm trên Amazon.com





khuyến nghị tin tức Nguồn hình ảnh: cafebiz.vn

Lịch đi học trở lại của học sinh cả nước: Có nhiều thay đổi I ĐOC THÊM

18-08-2021 - 09:31 AM | Thời sư



Lịch đi học trở lạ



18-08-2021 | Bộ GD8 17-08-2021 | MỚI NH



16-08-2021 | Chính t



NÓNG: TP.HCM công bố thời gian dự kiến bắt đầu năm học mới

Thời sự - 2 giờ trước

Mới đây, Sở GD-ĐT TP.HCM đã gửi báo cáo lên UBND thành phố về phương án bắt đầu năm học mới 2021-2022.



Chính thức: Hà Nội công bố thời gian đi học trở lại của học sinh

Tin hay nên đọc - 1 ngày trước

Mới đây, Sở GD-ĐT Hà Nội đã công bố thời gian đi học trở lại của học sinh trên khắp địa bán.



MỚI NHẤT: Lịch đi học trở lại của học sinh cả nước

Tin hay nên đọc - 1 ngày trước

Tính đến ngày 17/8, đã có gần 50 địa phương công bố thời gian tựu trường, khai giảng của học sinh trên khắp địa bàn.



Bộ GD&ĐT đề nghị địa phương tiếp nhận học sinh từ vùng dịch về quê học tập

Thời sự - 10 giờ trước

Bộ GD&ĐT vừa có văn bản gửi Chú tịch UBND các tính, TP trực thuộc T.Ư về việc tiếp nhận, tạo điều kiện cho học sinh học tập tại nơi cư trú do dịch COVID-19.

Ý tưởng chính đằng sau hệ khuyến nghị

Sử dụng dữ liệu lịch sử như sở thích trước đây của người dùng hoặc sở thích trước đây của người dùng tương tự để dự đoán các lượt thích trong tương lai

- Sở thích của người dùng có thể vẫn ổn định và thay đổi nhẹ theo thời gian:
 - Bước 1: Bằng cách xem sở thích của người dùng hoặc nhóm trong quá khứ, chúng ta cố gắng dự đoán sở thích của họ trong tương lai.
 - Bước 2: Sau đó, chúng ta giới thiệu các mặt hàng mà họ quan tâm.
- Tổng quát hóa: Hệ thống khuyến nghị lấy một tập hợp người dùng U và một tập hợp các sản phẩm I và tìm ra một hàm *f* sao cho:

$$f: U \times I \to R$$

Khuyến nghị và Tìm kiếm

- □ Một cách để nhận được câu trả lời là sử dụng công cụ tìm kiếm
- Công cụ tìm kiếm tìm những thấy kết quả phù hợp với truy vấn do người dùng cung cấp
- Các kết quả thường được cung cấp dưới dạng danh sách được sắp xếp theo thứ tự về mức độ liên quan của mục đó với truy vấn đã cho
- ☐ Xem xét truy vấn "best 2014 movie to watch"
 - Kết quả tương tự cho một đứa trẻ 8 tuổi và một người lớn

Kết quả của công cụ tìm kiếm không được tùy chỉnh

Những thách thức của hệ khuyến nghị

□ Vấn đề Cold Start

- Hệ thống đề xuất sử dụng dữ liệu lịch sử hoặc thông tin do người dùng cung cấp để đề xuất các mặt hàng, sản phẩm, v.v.
- Khi người dùng tham gia các trang web, họ vẫn chưa mua bất kỳ sản phẩm nào hoặc không có lịch sử.
- Thật khó để suy luận những gì họ sẽ thích khi họ bắt đầu trên một trang web.

□ Dữ liệu thưa (Data Sparsity)

- Khi thông tin lịch sử hoặc thông tin trước đó không đủ.
- Không giống như vấn đề Cold Start, vấn đề này nằm trong toàn bộ hệ thống và không chỉ riêng cho một cá nhân.

Những thách thức của hệ khuyến nghị

□ Attacks

- Push Attack: Đẩy xếp hạng lên bằng cách tạo người dùng giả mạo
- Nuke attack: Tấn công DDoS, dừng toàn bộ hệ thống khuyến nghị

Privacy

Sử dụng thông tin riêng tư của một người để giới thiệu cho người khác.

Explanation

Hệ thống đề xuất thường đề xuất các mặt hàng mà không giải thích lý do tại sao các mặt hàng này được đề xuất

Hệ khuyến nghị là gì?

- Recommender Systems/ Recommendation Systems viết tắt là RecSys.
- Là những hệ thống được thiết kế để hướng người dùng đến những đối tượng quan tâm, yêu thích, khi lượng thông tin quá lớn vượt quá khả năng xử lý của con người. [P. Resnick and H. R. Varian, 1997; Burke, 2002]
- Là những công cụ, phần mềm, kỹ thuật cung cấp những đề xuất các đối tượng có thể hữu ích với người dùng. Những đề xuất liên quan đến quyết định của người dùng như: Sản phẩm nào nên mua, bài hát nào nên nghe hay tin tức nào nên đọc. [Francesco Ricci et al. 2011]

Hệ khuyến nghị là gì?

Có thể hiểu

•Là những hệ thống giúp thông tin liên quan chủ động tìm đến người dùng thay vì người dùng phải đi tìm kiếm thông tin liên quan, khi lượng thông tin quá lớn vượt quá khả năng xử lý của con người.



Vai trò hệ khuyến nghị

- Đối với người dùng
- Đối với nhà cung cấp



Đối với người dùng

- Giảm chi phí tìm kiếm.
- Cung cấp thông tin, giúp người dùng hiểu biết về lĩnh vực đối tượng/sản phẩm liên quan.
- Tình cờ cung cấp thông tin mà người dùng không biết có sự tồn tại của nó.



Nhu cầu tìm kiếm và vấn đề quá tải thông tin

Đối với nhà cung cấp dịch vụ

· Thu hút người dùng

→ tăng số clicks,
views, tỷ lệ từ xem
hàng đến mua hàng
→ tăng doanh số bán
hàng và lợi nhuận.

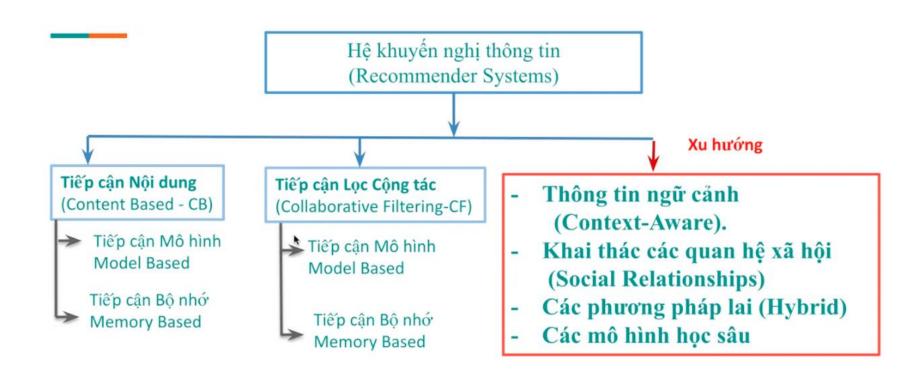


Phương pháp phổ biến

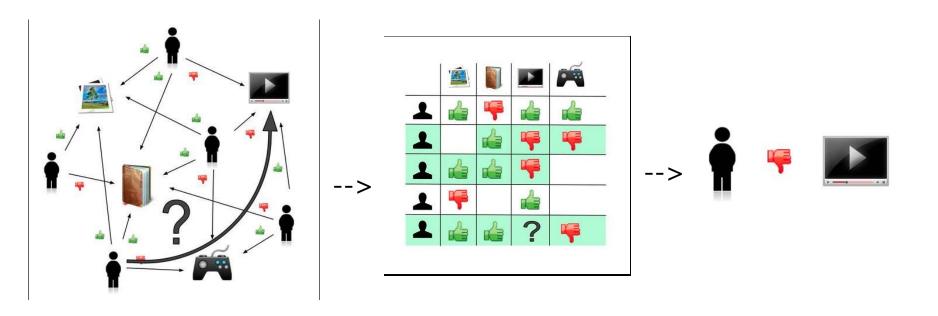
- · Lọc cộng tác
- · Lọc dựa trên nội dung
- Các phương pháp lai
- Phân tích mạng xã hội
- Các mô hình học sâu



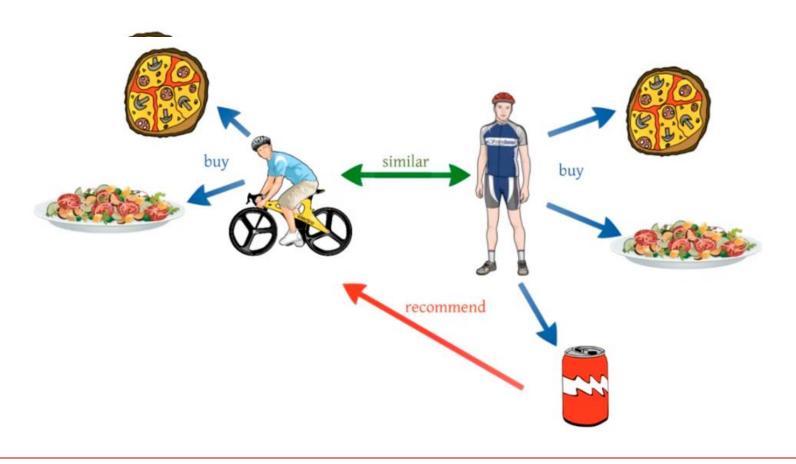
Cách tiếp cận phổ biến và xu hướng



Lọc cộng tác (Collaborative Filtering)

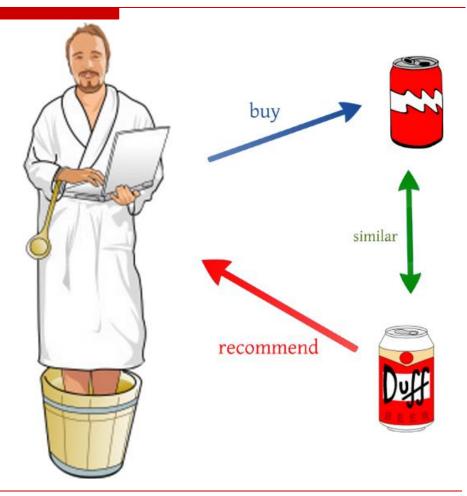


Lọc cộng tác (Collaborative Filtering)



Lọc dựa trên nội dung (Content-based Filtering)

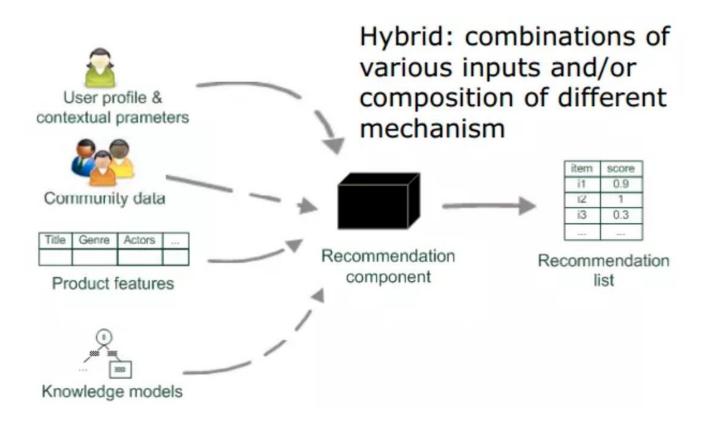
CB: "Show me more of the same what I've liked"



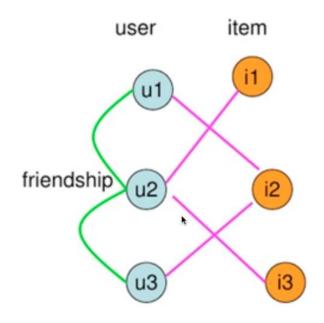
Lọc dựa tri thức (Knowledge-based Filtering)

- Đôi khi sở thích người dùng không thể xác định được từ thời điểm ban đầu, nhưng có thể xác định trong quá trình tương tác. (khảo sát người dùng trước một list câu hỏi chẳng hạn)
- Cách tiếp cận này phù hợp với cold-start hay người dùng ít thông tin.
- Ví dụ: hệ thống khuyến nghị nơi ẩm thực: tương tác biết anh/chị thích ăn món $X \rightarrow tư vấn đến nhà hàng Y$.

Phương pháp lai (Hybrid)



Tiếp cận mạng xã hội (Social Recommendation)



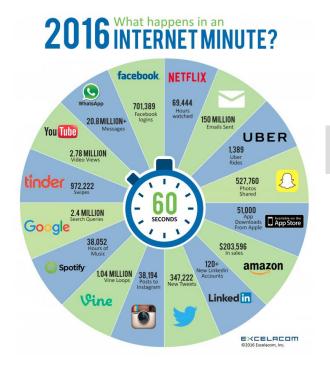
Ví dụ: U1 chơi thân U2, U2 chơi thân U3, mà U1 thích I2 nên đề xuất I2 cho U3



Thank you!

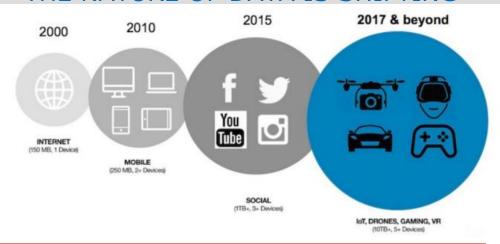
Bối cảnh ra đời các nghiên cứu liên quan





DATA IS being generated by people, machine, applications and a combination of all of these.

THE NATURE OF DATA IS SHIFTING



Theo báo cáo của IDC được tài trợ bởi Seagate, sẽ không có gì ngạc nhiên vào năm 2025, chúng ta sẽ bị tràn ngập bởi dữ liệu, **khoảng 163 ZB, gấp 10 lần so với năm 2016** (khoảng 16 ZB)



Name	Equal to:	Size in Bytes
Bit	1 bit	1/8
Nibble	4 bits	1/2 (rare)
Byte	8 bits	1
Kilobyte	1,024 bytes	1,024
Megabyte	1,024 kilobytes	1,048,576
Gigabyte	1,024 megabytes	1,073,741,824
Terrabyte	1,024 gigabytes	1,099,511,627,776
Petabyte	1,024 terrabytes	1,125,899,906,842,624
Exabyte	1,024 petabytes	1,152,921,504,606,846,976
Zettabyte	1,024 exabytes	1,180,591,620,717,411,303,424
Yottabyte	1,024 zettabytes	1,208,925,819,614,629,174,706,176

1 YB to như thế nào?



Các nghiên cứu ban đầu

❖ Recommendation System Bắt đầu thu hút nhiều sự quan tâm từ giữa thập niên 90, Resnick et al. [1994], Hill et al. [1995], Shardanand & Maes [1995].

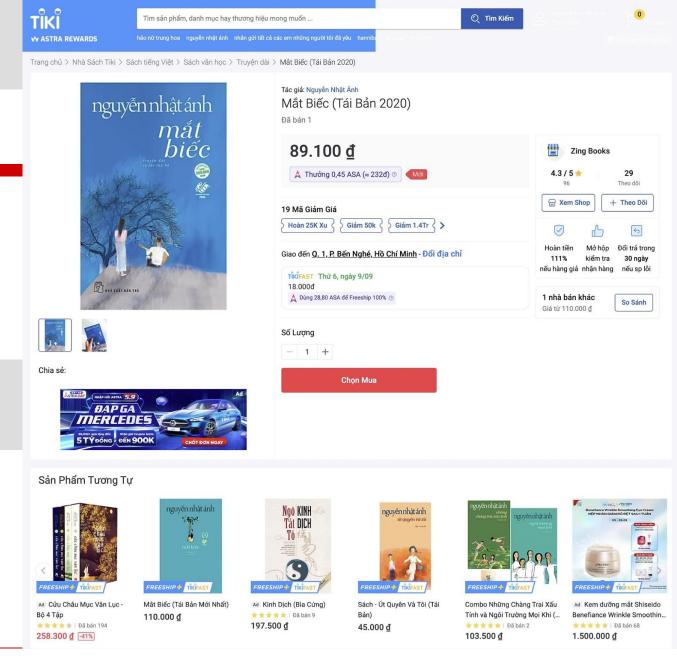
❖ Recommendation System là gì?

Khái niệm hệ khuyến nghị

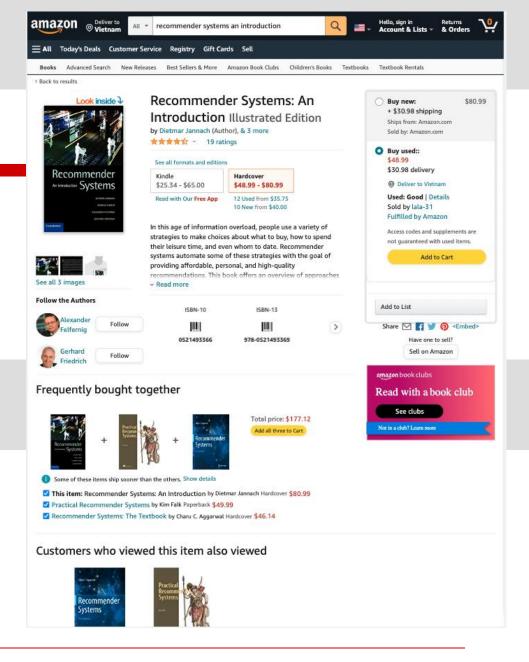
- · Ví dụ minh họa
- Định nghĩa
- Bài toán khuyến nghị thông tin

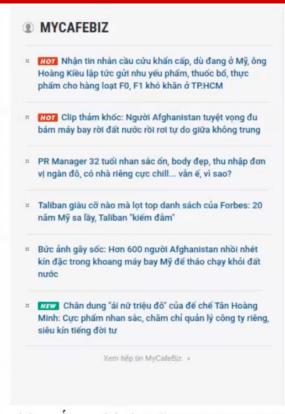


Minh họa khuyến nghị sản phẩm trên TIKI.VN



Minh họa khuyến nghị sản phẩm trên Amazon.com





Lịch đi học trở lại của học sinh cả nước: Có nhiều thay đổi I ĐOC THÊM

18-08-2021 - 09:31 AM | Thời sư



Lịch đi học trở lạ



18-08-2021 | **Bộ GD**8 17-08-2021 | **MỚI N**H 16-08-2021 | **Chính t**





NÓNG: TP.HCM công bố thời gian dự kiến bắt đầu năm học mới

Thời sự - 2 giờ trước

Mới đây, Sở GD-ĐT TP.HCM đã gửi báo cáo lên UBND thành phố về phương án bắt đầu năm học mới 2021-2022.



Chính thức: Hà Nội công bố thời gian đi học trở lại của học sinh

Tin hay nên đọc - 1 ngày trước

Mới đây, Sở GD-ĐT Hà Nội đã công bố thời gian đi học trở lại của học sinh trên khắp địa bán.



MỚI NHẤT: Lịch đi học trở lại của học sinh cả nước

Tin hay nên đọc - 1 ngày trước

Tính đến ngày 17/8, đã có gần 50 địa phương công bố thời gian tựu trưởng, khai giảng của học sinh trên khắp địa bàn.



Bộ GD&ĐT đề nghị địa phương tiếp nhận học sinh từ vùng dịch về quê học tập

Thời sư - 10 giờ trước

Bổ GD&ĐT vừa có văn bản gửi Chú tịch UBND các tính, TP trực thuộc T.Ư về việc tiếp nhận, tạo điều kiện cho học sinh học tập tại nơi cư trú do dịch COVID-19.

khuyến nghị tin tức Nguồn hình ảnh: cafebiz.vn

Hệ khuyến nghị là gì?

- Nên đọc tài liệu nào về hệ khuyến nghị?
- Làm khóa luận tốt nghiệp thì nên tìm ai hướng dẫn?
- Đầu tư gì là tốt nhất để hỗ trợ giáo dục trẻ em?
- Cuối tuần này cùng người yêu đi đâu, xem phim gì, đọc sách gì?
- Mua mac M2 3T thì nên mua ở đâu?
- - ☐ Cách bình thường: Người dùng sẽ search trên Internet
 - □ Như vậy có xem là khuyến nghị không? Không

Phát biểu bài toán khuyến nghị

Cho trước

- $U = \{u_1, u_2, u_3, ..., u_n\}$, không gian người dùng.
- $P = \{p_1, p_2, p_3, ..., p_m\}$, không gian đối tượng khuyến nghị
- Mục đích của hệ khuyến nghị là tìm hàm hữu ích $f: UxP \to R$, Và ước lượng giá trị của f(u,p) (với u thuộc U, p thuộc P).
- $Gi\acute{a}$ tri của f(u,p) cho biết u sẽ thích p nhiều hay ít, hay p hữu ích đối với u như thế nào.