**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM TP. HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

**A screenshot of a video game

Description automatically generated with medium confidence**

**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**HỆ THỐNG ĐỀ XUẤT SÁCH SỬ DỤNG COLLABORATIVE FILTERING**

**MÔN HỌC: HỆ TƯ VẤN THÔNG TIN**

**HỌC PHẦN: COMP131001**

**Thành phố Hồ Chí Minh, ngày 5 tháng 5 năm 2023**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM TP. HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

**A screenshot of a video game

Description automatically generated with medium confidence**

**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**HỆ THỐNG ĐỀ XUẤT SÁCH SỬ DỤNG COLLABORATIVE FILTERING**

**MÔN HỌC: HỆ TƯ VẤN THÔNG TIN**

**HỌC PHẦN: COMP131001**

|  |  |  |
| --- | --- | --- |
| **Giảng viên hướng dẫn:** Ths. Trần Thanh Nhã  **Nhóm INFOR**  **Họ và Tên MSSV** | | |
| Dư Thị Như Quỳnh | 46.01.104.151 |
| Nguyễn Lê Kim Khánh | 46.01.104.081 |
| Hồ Thị Ngọc Ánh | 4501104012 |
| Cao Thị Thanh Trâm | 4501104250 |

![Background pattern

Description automatically generated with low confidence](data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAAAJoAAABGCAYAAAA0A41DAAAAAXNSR0IArs4c6QAAAARnQU1BAACxjwv8YQUAAAAJcEhZcwAAEnQAABJ0Ad5mH3gAAADOSURBVHhe7dKxAYAwDMCw0P9/Boa+EE/S4gf8vL+BZecWVhmNhNFIGI2E0UgYjYTRSBiNhNFIGI2E0UgYjYTRSBiNhNFIGI2E0UgYjYTRSBiNhNFIGI2E0UgYjYTRSBiNhNFIGI2E0UgYjYTRSBiNhNFIGI2E0UgYjYTRSBiNhNFIGI2E0UgYjYTRSBiNhNFIGI2E0UgYjYTRSBiNhNFIGI2E0UgYjYTRSBiNhNFIGI2E0UgYjYTRSBiNhNFIGI2E0UgYjYTRSBiNhNEIzHwfzQSIu/IkawAAAABJRU5ErkJggg==)

**Thành phố Hồ Chí Minh, ngày 5 tháng 5 năm 2023**

**LỜI CẢM ƠN**

Lời nói đầu tiên nhóm chúng em xin gửi lời biết ơn sâu sắc đến thầy Trần Thanh Nhã - giảng viên hướng dẫn chúng em học phần Hệ tư vấn thông tin trong học kỳ này. Thầy đã truyền đạt cho chúng em những kiến thức bổ ích cũng như hướng dẫn chúng em vô cùng tận tình trong quá trình học.

Trong quá trình học tập và rèn luyện, chúng em đã được truyền đạt, cung cấp và trang bị những kiến thức cơ bản và chuyên sâu về recommender systems và machine learning. Qua từng buổi học, chúng em tiếp thu thêm nhiều kiến thức mới và học hỏi nhiều thông tin hữu ích giúp cho đồ án thêm hoàn thiện.

Mặc dù đã cố gắng để hoàn thiện bài báo cáo, tuy nhiên do hạn chế về thời gian và kiến thức nên vẫn không thể tránh khỏi những thiếu sót, chúng em rất mong nhận được sự thông cảm và đóng góp quý báu của thầy.

Nhóm chúng em xin chân thành cảm ơn!

****

**MỤC LỤC**

[DANH MỤC HÌNH ẢNH 1](#_Toc136425685)

[BẢNG PHÂN CÔNG 2](#_Toc136425686)

[MỞ ĐẦU 3](#_Toc136425687)

[1. CƠ SỞ LÝ THUYẾT 5](#_Toc136425688)

[1.1. Recommendation Systems (hệ thống gợi ý) 5](#_Toc136425689)

[1.1.1 Khái niệm 5](#_Toc136425690)

[1.1.2 Thành phần của Recommendation Systems 5](#_Toc136425691)

[1.1.3 Hướng tiếp cận 6](#_Toc136425692)

[1.2. Collaborative Filtering 8](#_Toc136425693)

[1.2.1 Giới thiệu thuật toán 8](#_Toc136425694)

[1.2.2. Neighborhood-based Collaborative filtering 9](#_Toc136425695)

[1.2.3. Matrix Fractorization 12](#_Toc136425696)

[2. XÂY DỰNG HỆ THỐNG 15](#_Toc136425697)

[2.1. Mô tả hệ thống 15](#_Toc136425698)

[2.1.1. Giới thiệu về dự án 15](#_Toc136425699)

[2.1.2. Tổng quan về dự án 16](#_Toc136425700)

[2.2. Phân tích và thiết kế hệ thống gợi ý 17](#_Toc136425701)

[2.3. Sử dụng Item-based Collaborative Filtering để gợi ý sách 18](#_Toc136425702)

[2.4. Bộ dữ liệu 20](#_Toc136425703)

****[2.4.1. Mô tả chi tiết bộ dữ liệu 20](#_Toc136425704)

[2.4.2. Data mining 23](#_Toc136425705)

[3. XÂY DỰNG ỨNG DỤNG 27](#_Toc136425706)

[3.1. Yêu cầu cài đặt trước một số công cụ sau: 27](#_Toc136425707)

[3.2. Cấu trúc tổ chức thư mục 27](#_Toc136425708)

[3.3. Xây dựng trang web 28](#_Toc136425709)

[3.3.1. Xây dựng thuật toán gợi ý sách 28](#_Toc136425710)

[3.3.2. Xây dựng hiện thị web 31](#_Toc136425711)

[4. WESITE GỢI Ý SÁCH 33](#_Toc136425712)

[6. KẾT LUẬN 37](#_Toc136425713)

[NGUỒN DỮ LIỆU 38](#_Toc136425714)

[TÀI LIỆU THAM KHẢO 38](#_Toc136425715)



# **DANH MỤC HÌNH ẢNH**

[Hình 1. Thành phần của Recommendation Systems 5](#_Toc135323972)

[Hình 2. Phân loại Recommender System 7](#_Toc135323973)

[Hình 3. Ý tưởng Collaborative filtering 8](#_Toc135323974)

[Hình 4. Neighborhood-based Collaborative filtering 9](#_Toc135323975)

[Hình 5. Các bước xây dựng thuật toán Neighborhood-based Collaborative filtering 10](#_Toc135323976)

[Hình 6. Minh họa Matrix Fractorization 12](#_Toc135323977)

[Hình 7. Mô hình training 17](#_Toc135323978)

[Hình 8. Item-based Collaborative filtering 18](#_Toc135323979)

[Hình 9. Data mining 24](#_Toc135323980)

[Hình 10. Giao diện chính của web 33](#_Toc135323981)

[Hình 11. Giao diện chức năng Recommend 33](#_Toc135323982)

[Hình 12. Giao diện hiện thị sách gợi ý 34](#_Toc135323983)

[Hình 13. Giao diện thông báo lỗi 34](#_Toc135323984)

# **BẢNG PHÂN CÔNG**

|  |  |  |  |
| --- | --- | --- | --- |
| **Họ tên** | **MSSV** | **Nhiệm vụ** | **Đánh giá mức độ hoàn thành** |
| Dư Thị Như Quỳnh | 46.01.104.151 | Soạn nội dung mục 1,2  Thực hiện ppt  Code thuật toán | 100% |
| Nguyễn Lê Kim Khánh | 46.01.104.081 | Tổng hợp báo cáo  Code thuật toán | 100% |
| Hồ Thị Ngọc Ánh | 4501104012 | Soạn nội dung mục 3  Thiết kế giao diện | 80% |
| Cao Thị Thanh Trâm | 4501104250 | Soạn nội dung mục 4,5  Thiết kế giao diện | 80% |

# **MỞ ĐẦU**

* **Lý do chọn đề tài**

Hệ thống gợi ý (Recommender Systems – RS) là một hệ thông giúp dự đoán thông minh sở thích của khách hàng và cung cấp những thông tin mà họ thực sự quan tâm dựa vào sở thích/thói quen của người trong quá khứ. Hệ thống gợi ý đã và đang được ứng dụng trong rất nhiều trong các lĩnh vực khác nhau như thương mại điện tử (hỗ trợ bán hàng trực tuyến), giải trí (gợi ý phim ảnh, bài hát,..), giáo dục đào tạo (gợi ý nguồn tài nguyên học tập, nghiên cứu,..),... Chính vì khả năng ứng dụng rộng rãi của nó, Hệ thống gợi ý mở ra nhiều tiềm năng trong việc nghiên cứu cũng như trong xây dựng và phát triển các hệ thống thực tế. Chính vì thế nhóm em đã chọn đề tài: “HỆ THỐNG GỢI ĐỀ XUẤT SÁCH SỬ DỤNG COLLABORATIVE FILTERING” để xây dựng một hệ thống gợi ý sách cho người dùng.

* **Mục tiêu nghiên cứu**

Hiểu được cách thức hoạt động của Recommendation Systems (Hệ tư vấn thông tin) và thuật toán Collaborative Filtering (Lọc cộng tác).

Ứng dụng vào thực tế giải quyết bài toán gợi ý, xây dựng hệ thống gợi ý sách với thuật toán Collaborative Filtering.

Cách tìm hiểu, tham khảo, học hỏi những kiến thức mới.

* **Đối tượng và phạm vi nghiên cứu**

Nghiên cứu về Recommendation Systems.

Nghiên cứu những kiến thức cơ bản của Recommendation Systems, Python.

* **Phương pháp nghiên cứu**

Sử dụng phương pháp nghiên cứu, phân tích, tổng hợp.

Thực hành những kiến thức liên quan đến đề tài.

* **Kết cấu đề tài**

Nội dung bài báo cáo bao gồm những nội dung sau:

Chương 1: Cơ sở lý thuyết

Chương 2: Xây dựng hệ thống

Chương 3: Sử dụng python thiết kế website gợi ý sách

# **1. CƠ SỞ LÝ THUYẾT**

## **1.1. Recommendation Systems (hệ thống gợi ý)**

### **1.1.1 Khái niệm**

Recommender System là một trong những ứng dụng phổ biến nhất của khoa học dữ liệu ngày nay. Được sử dụng để dự đoán đánh giá (ratting) hoặc sở thích (preference) của người dùng sẽ dành cho một mặt hàng. Nhằm cá nhân hóa trải nghiệm của người dùng, đề xuất được những mặt hàng người dùng quan tâm.[1]

Hệ thống gợi ý được áp dụng hầu hết trong mọi nền tảng trên internet, đặc biệt là đối với các công ty lớn như:

* Amazon sử dụng nó để đề xuất sản phẩm cho khách hàng.
* Netflix sử dụng gợi ý phim cho người dùng.
* Youtube sử dụng nó để đề xuất các video và quyết định video sẽ phát tiếp theo trên chế độ tự động phát.
* Facebook sử dụng nó để gợi ý kết bạn, đề xuất các trang để thích và mọi người theo dõi.

### **1.1.2 Thành phần của Recommendation Systems**

Các hệ thống gợi ý thường bao gồm bốn phần: cơ sở dữ liệu, giao diện người dùng – máy tính, các thuật toán và các thành phần khuyến nghị.[2]

Diagram

Description automatically generated

Hình 1. Thành phần của Recommendation Systems

Cơ sở dữ liệu: Đây là nơi lưu trữ thông tin về các mục tiêu khuyến nghị, ví dụ như danh sách các sách, bài viết, phim hay sản phẩm khác. Cơ sở dữ liệu cũng có thể chứa thông tin về người dùng, bao gồm lịch sử đọc, đánh giá và sở thích cá nhân.

Giao diện người dùng - máy tính: Đây là phần mà người dùng tương tác trực tiếp với hệ thống gợi ý. Giao diện người dùng cung cấp các chức năng như tìm kiếm, lựa chọn sở thích cá nhân và xem các gợi ý khuyến nghị.

Thuật toán: Đây là bộ quy tắc và quy trình tính toán được sử dụng để xác định các mục tiêu khuyến nghị. Thuật toán có thể sử dụng các phương pháp như lọc cộng tác (collaborative filtering), lọc nội dung (content-based filtering), học máy (machine learning) và các kỹ thuật khác để tạo ra các gợi ý phù hợp với người dùng.

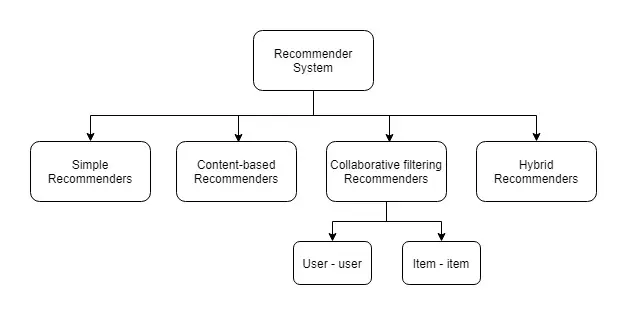
Các thành phần khuyến nghị: Các mục tiêu khuyến nghị cụ thể được hiển thị cho người dùng, ví dụ như danh sách các sách, bài viết, phim hay sản phẩm khác được gợi ý. Các thành phần này được chọn dựa trên kết quả của thuật toán và các yếu tố khác như sự tương đồng với sở thích cá nhân, độ phổ biến và sự tương tác của người dùng khác.

Các phần này hoạt động cùng nhau để tạo ra hệ thống gợi ý hiệu quả, giúp cung cấp các nội dung phù hợp và tăng trải nghiệm người dùng. Các công nghệ và phương pháp cụ thể có thể khác nhau tùy thuộc vào hệ thống gợi ý cụ thể và mục đích sử dụng của nó.

### **1.1.3 Hướng tiếp cận**

Hệ thống Recommender System có thể được chia làm 4 loại chính:

* Simple Recommenders: Đưa ra các đề xuất tổng quát cho mọi người dùng, dựa trên mức độ phổ biến và/hoặc phân loại của sản phẩm.
* Content-based Recommenders: Đề xuất các sản phẩm tương tự dựa trên một sản phẩm cụ thể dựa trên các đặc tính được gán nhãn (phân loại, nhà sản xuất,…)
* Collaborative filtering Recommenders: Dự đoán “ratings” hoặc “preference” mà người dùng sẽ đưa ra một mặt hàng dựa trên xếp hạng trước đây và sở thích của những người dùng khác.
* Hybrid Recommenders: Hybrid Filtering là sự kết hợp của hai giải thuật Content-based Filtering và Collabrative Filtering, Hybrid Fitering được sử dụng khi hệ thống Collabrative Filtering không có các hành vi (ratings), khi đó hệ thống sẽ sử dụng Content-based Filtering. Ngược lại, khi Content-based Filtering không có các feature cần thiết trong việc đánh giá thì hệ thống sẽ sử dụng Collaborative Fitering để thay thế.



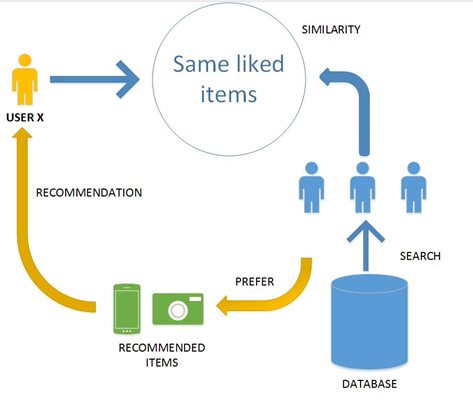
Hình 2. Phân loại Recommender System

## **1.2. Collaborative Filtering**

### **1.2.1 Giới thiệu thuật toán**

Collaborative filtering (lọc cộng tác) là thuật toán lọc tương tác tức là tìm ra sản phẩm mà khách hàng có khả năng ưa thích nhất dựa vào những sản phẩm mà những khách hàng khác có hành vi tương tự đã lựa chọn. Thuật toán sẽ không cần sử dụng thông tin sản phẩm là đầu vào cho dự báo rating. Đầu vào của thuật toán là một ma trận tiện ích (ultility matrix) chứa giá trị rating của các cặp (user, item). Mỗi cột là các rating mà một user đã rate và mỗi dòng là các rating của một item được rate. [3]

Ý tưởng thuật toán: Khi người dùng muốn tìm một cuốn sách mới để xem, thì thường hỏi bạn bè của mình cho các đề xuất. Người dùng có sự tin tưởng nhiều hơn vào các khuyến nghị từ những người bạn chia sẻ ( cùng thị hiếu, sở thích,…).



Hình 3. Ý tưởng Collaborative filtering

Collaborative Filtering sử dụng các tương tác và dữ liệu được hệ thống thu thập từ người dùng khác. Nhằm xác định mức độ tương quan giữa người dùng và các mục. Dựa trên sự tương quan này, collaborative filtering đưa ra các đề xuất cho người dùng về các mục mà họ có thể quan tâm.

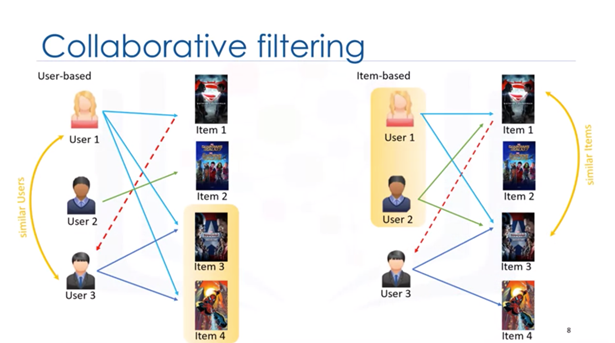
Phương pháp này có một số ưu điểm như đơn giản và dễ hiểu, dễ triển khai và cho kết quả tốt khi dữ liệu được phân bố đều trong các nhóm người dùng. Tuy nhiên, nó cũng có một số hạn chế như độ phức tạp tính toán tăng lên khi số lượng người dùng và sản phẩm tăng lên, và không đưa ra được gợi ý cho các sản phẩm mới hoặc những người dùng mới.

Có 2 phương pháp chính được sử dụng trong collaborative filtering bao gồm: Neighborhood-based collaborative Filtering và Matrix Fractorization.

### **1.2.2. Neighborhood-based Collaborative filtering**

Ở phương pháp này ta sẽ cần xây dựng ma trận hệ số tương quan của vector rating của các users để tìm ra nhóm users có cùng sở thích. Hệ số tương quan giữa các users càng lớn thì sở thích của họ càng giống nhau và trái lại thì họ càng có sở thích khác biệt.

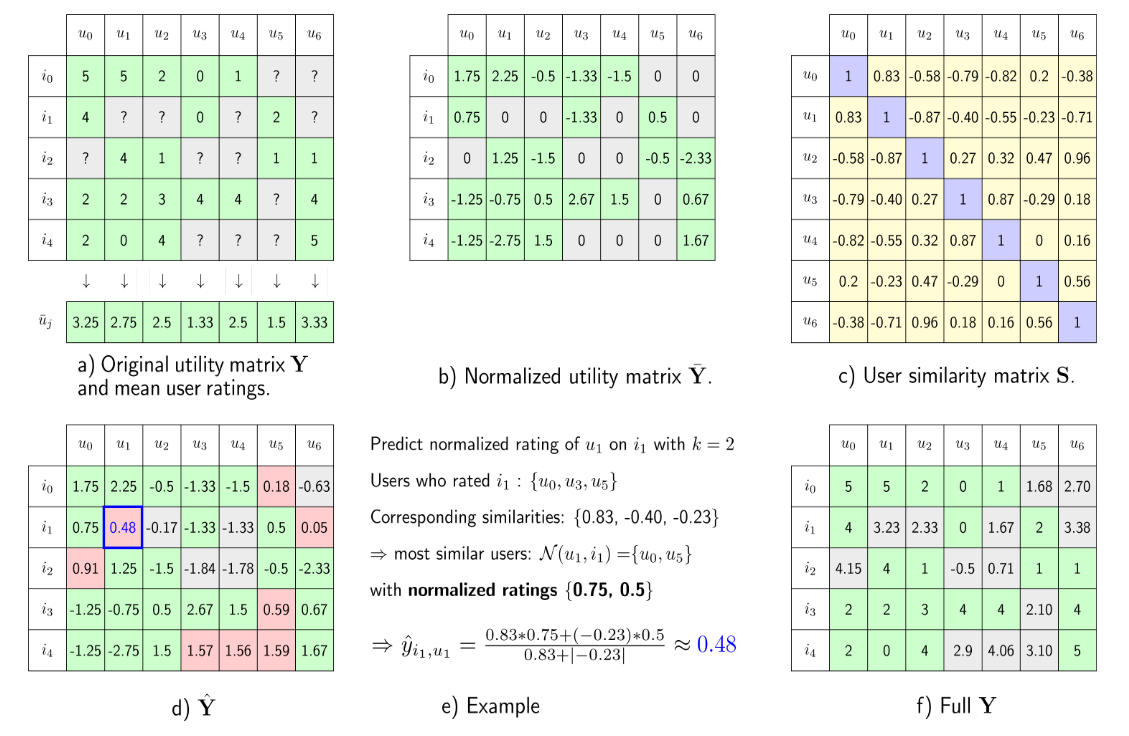
Thuật toán sẽ dự đoán giá trị rating tại một cặp (user, item) chưa được rate bằng cách tính tổng có trọng số các giá trị rating của users tương quan nhất với user đó mà đã rate item trên. Trọng số thông thường sẽ bẳng chính hệ số tương quan.



Hình 4. Neighborhood-based Collaborative filtering

* **Các bước xây dựng thuật toán Neighborhood-based Collaborative filtering**

Để xây dựng một thuật toán Neighborhood-based Collaborative filtering chúng ta cần trải qua các bước sau:



Hình 5. Các bước xây dựng thuật toán Neighborhood-based Collaborative filtering

Bước 1: Chuẩn hóa dữ liệu ở ma trận tiện ích **Y** bằng cách trừ đi ở mỗi cột (là các rating của cùng 1 user) trung bình giá trị rating của cột. Việc này là để loại bỏ sự khác biệt về mức độ cho điểm của các user. Vì ví dụ: Có một số user khó tính có thể cho điểm cao nhất là 3 nhưng user dễ tính thì điểm thấp nhất là 3. Khi đó nếu nhận định user khó tính không thích item (nếu ta coi 3 là điểm thấp) hoặc user dễ tính yêu thích item (nếu ta coi 3 là điểm cao) là không chuẩn xác. Chuẩn hóa giá trị rating nhằm mục đích đưa trung bình rating của các user sau khi chuẩn hóa về 0. Gía trị rating dương thể hiện user ưa thích item và trái lại âm sẽ là không thích, bằng 0 là trung lập.

Bước 2: Tính ma trận hệ số tương quan giữa các véc tơ cột. Ma trận tương quan thể hiện sự tương đồng trong hành vi mua sắm giữa các user. Từ ma trận tương quan ta có thể xác định ra các users có sở thích tương đồng nhất với một user xác định. Hệ số tương quan dương và càng gần 1 chứng tỏ 2 users có sở thích giống nhau. Hệ số tương quan âm là 2 users có hành vi trái ngược.

Bước 3: Dự báo rating của một user u cho một item i bằng cách xác định trên ma trận hệ số tương quan một tập S(u,k|i) gồm k users có giá trị tương quan lớn nhất đối với user u mà đã rate item i. Giá trị dự báo rating của user u sẽ được tính bằng tổng có trọng số của các rating trong tập k users tương quan nêu trên theo công thức bên dưới:

i,u =

Bước 4: Chuyển giá trị dự báo ở ma trận chuẩn hóa sang giá trị dự báo rating bằng cách cộng các giá trị ở ma trận chuẩn hóa với giá trị trung bình của mỗi cột.

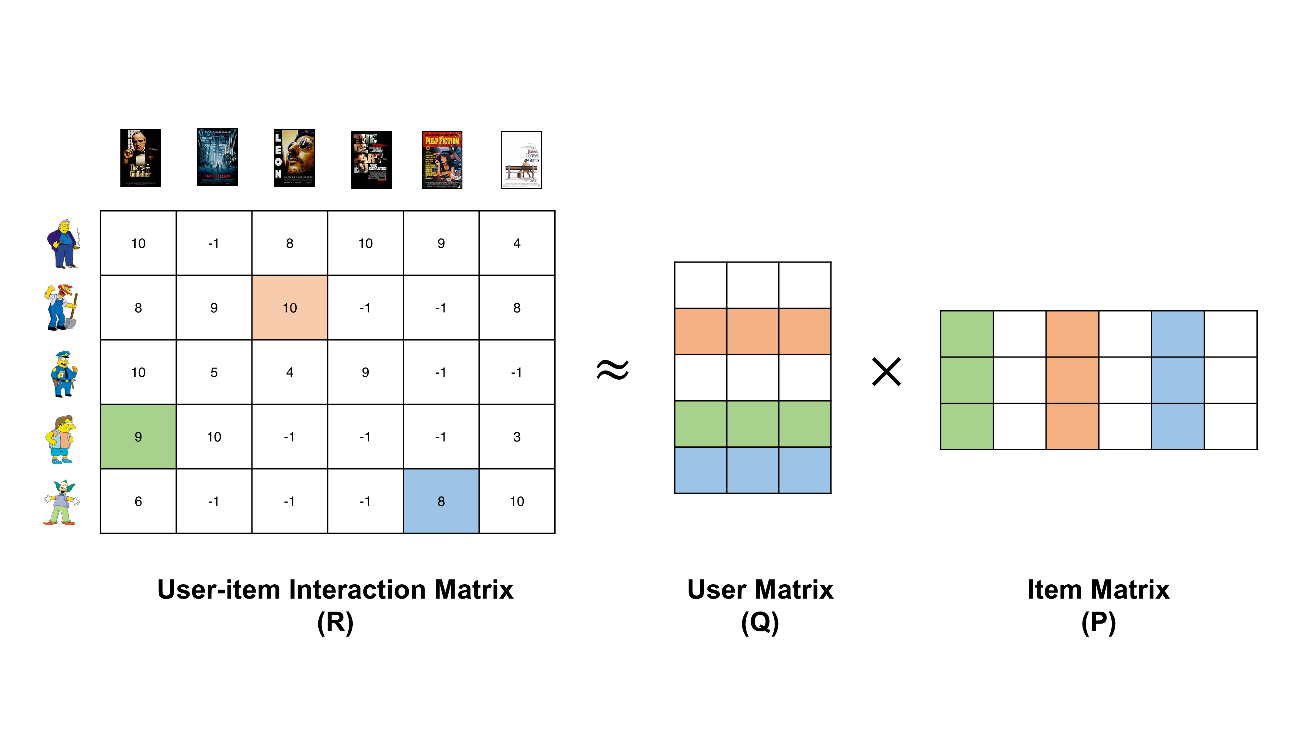
* **Hạn chế của phương pháp collaborative filtering:**

Thường phải lưu một ma trận hệ số tương quan với kích thước rất lớn. Việc này dẫn tới tốn tài nguyên lưu trữ và thời gian tính toán.

Ở trên ta đã lựa chọn việc chuẩn hóa theo chiều user. Ngoài ra, ta cũng có thể lựa chọn chuẩn hóa theo chiều item mà không làm thay đổi phương pháp bằng cách chuyển vị ma trận tiện tích Y. Việc lựa chọn chuẩn hóa theo chiều nào sẽ căn cứ trên kích thước theo chiều nào là lớn hơn. Thông thường số lượng user sẽ nhiều hơn item. Khi đó chuẩn hóa theo item sẽ có lợi thế hơn bởi: Kích thước ma trận hệ số tương quan giữa các user là nhỏ hơn nên tốn ít tài nguyên. Thêm nữa khi một user rating một item mới thì giá trị thay đổi về trung bình trên mỗi cột item là nhỏ hơn so với trường hợp chuẩn hóa theo user. Điều này dẫn tới ma trận hệ số tương quan ít thay đổi hơn và tần suất cập nhật cũng ít hơn.

### **1.2.3. Matrix Fractorization**

Bên cạnh thuật toán Neighborhood-based collaborative Filtering, một thuật toán khác cũng thuộc lớp các bài toán Collabrative filtering đó là Matrix fractorization. Thuật toán này thường mang lại độ chính xác cao hơn và đồng thời yêu cầu ít tài nguyên lưu trữ hơn. [4]

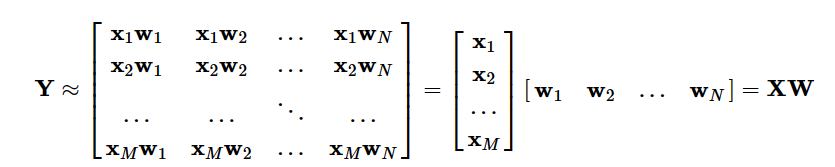


Hình 6. Minh họa Matrix Fractorization

Mỗi item được mô tả bằng một vector x được gọi là item profile. Trong phương pháp này, ta cần tìm một vector hệ số w tương ứng với mỗi user sao cho rating đã biết mà user đó cho item xấp xỉ với:



Mở rộng với Y là Utility matrix, giả sử đã được điền hết giá trị, ta có:

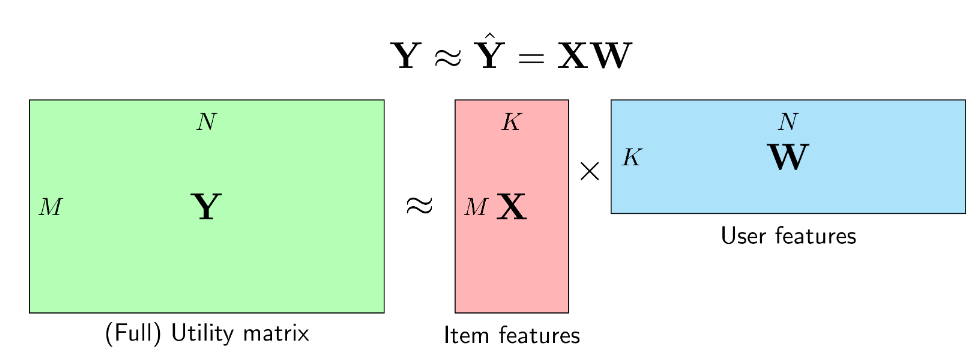


với M, N lần lượt là số items và số users.

Lưu ý, do X là được xây dựng dựa trên thông tin mô tả item và quá trình xây dựng này độc lập với quá trình đi tìm hệ số phù hợp cho mỗi user nên việc xây dựng item profile đóng vai trò quan trọng và có thể ảnh hưởng trực tiếp đến hiệu năng của mô hình. Thêm nữa, việc xây dựng mô hình riêng lẻ cho mỗi user dẫn đến kết quả chưa thực sự tốt vì không khai thác được đặc điểm giống nhau giữa các user.

Giả sử rằng ta không cần xây dựng trước các item profile mà ma trận này có thể huấn luyện đồng thời với ma trận trọng số, hay nói các khác bài toán này là bài toán tối ưu các ma trận X và W, trong đó X là ma trận của toàn bộ các item profiles, mỗi hàng tương ứng với 1 item. Còn W là ma trận của toàn bộ user models (các mô hình của users), mỗi cột tương ứng với một user. Chúng ta sẽ cố gắng xấp xỉ utility matrix Y ∈ RM∗N  bằng tích của hai ma trận con là X ∈ RM\*K và W ∈ R K∗N.

Trong đó, K được chọn thường nhỏ hơn rất nhiều so với M và N, và cả hai ma trận X và W đều phải có bậc (rank) không được vượt quá K



* **Ưu nhược điểm của Matrix factorization**

**Ưu điểm:**

* Khả năng dự đoán tốt: Matrix factorization CF có thể tạo ra các dự đoán chất lượng về sự quan tâm của người dùng đối với các mục tiêu chưa được xác định dựa trên các mẫu đánh giá hiện có.
* Không yêu cầu thông tin về người dùng và mục tiêu: Phương pháp này không đòi hỏi thông tin chi tiết về người dùng và mục tiêu. Nó chỉ cần ma trận đánh giá (người dùng-mục tiêu) để xây dựng mô hình.
* Xử lý hiệu quả với dữ liệu lớn: Matrix factorization CF thường dễ dàng mở rộng để xử lý dữ liệu lớn và các ma trận thưa (sparse matrix) bằng cách sử dụng các kỹ thuật như SVD (Singular Value Decomposition) hoặc các phương pháp tối ưu hóa như Stochastic Gradient Descent (SGD).

**Nhược điểm:**

* Vấn đề cold start: Matrix factorization CF không thể xử lý tốt khi có các người dùng hoặc mục tiêu mới mà không có đủ dữ liệu đánh giá để đưa ra dự đoán chính xác. Điều này gây khó khăn trong giai đoạn ban đầu khi không có đủ thông tin về người dùng hoặc mục tiêu mới.
* Mất mát thông tin: Matrix factorization CF giảm số chiều của ma trận ban đầu bằng cách phân tích thành hai ma trận thấp chiều. Trong quá trình này, một phần thông tin có thể bị mất mát, điều này có thể ảnh hưởng đến khả năng dự đoán chính xác.
* Độ phức tạp tính toán: Các thuật toán matrix factorization CF có thể tốn nhiều thời gian và tài nguyên tính toán, đặc biệt là khi xử lý dữ liệu lớn. Điều này làm cho nó khó khăn để triển khai trong môi trường thời gian thực.

# **2. XÂY DỰNG HỆ THỐNG**

## **2.1. Mô tả hệ thống**

### **2.1.1. Giới thiệu về dự án**

Hệ thống gợi ý dùng collaborative filtering là một bài toán quan trọng trong lĩnh vực recommend system. Với sự phát triển nhanh chóng của ngành công nghệ thông tin và số lượng người dùng sách ngày càng tăng, việc tìm kiếm và khám phá những sản phẩm phù hợp đã trở thành một thách thức. Collaborative filtering đã chứng minh sự hiệu quả của mình trong việc giải quyết bài toán gợi ý, đặc biệt là trong gợi ý sách.

Collaborative filtering tập trung vào sự tương quan giữa người dùng và các cuốn sách đã được đánh giá hoặc sử dụng bởi các người dùng khác. Thay vì phân tích nội dung và đặc tính của từng cuốn sách, collaborative filtering sử dụng thông tin về hành vi và sự tương tác của người dùng để xác định sự tương quan giữa người dùng và các cuốn sách. Phương pháp này đưa ra các đề xuất sách dựa trên sự tương quan giữa người dùng hiện tại và những người dùng khác có sở thích tương tự.

Collaborative filtering có nhiều ưu điểm đáng chú ý:

* Không đòi hỏi thông tin chi tiết về nội dung sách, giúp giảm bớt công sức thu thập và cung cấp thông tin.
* Collaborative filtering phù hợp cho cả người dùng mới mà không cần lịch sử đánh giá.
* Đảm bảo độ chính xác cao trong việc đưa ra các đề xuất phù hợp với sở thích của người dùng.
* Collaborative filtering dễ triển khai và mở rộng, giúp nhanh chóng xây dựng hệ thống gợi ý sách cho người dùng.

Trong bối cảnh ngày càng nhiều người dùng tìm kiếm những cuốn sách phù hợp với sở thích của mình, bài toán gợi ý sách dùng collaborative filtering đã đóng vai trò quan trọng. Phương pháp này không chỉ giúp người dùng khám phá và tìm kiếm sách một cách dễ dàng, mà còn mang đến trải nghiệm đọc sách cá nhân hóa và tăng cường tương tác trong cộng đồng đọc sách.

### **2.1.2. Tổng quan về dự án**

|  |
| --- |
| **Thông tin tổng quan**  **Tên dự án: Hệ thống đề xuất sách sử dụng Collaborative filtering** |
| **Mục tiêu:** Xây dựng trang web gợi ý sách làm rõ dữ liệu đầu vào và thuật toán Collaborative filtering |
| **Các bước thực hiện:**   * Phân tích thiết kế, xử lý dữ liệu * Lập trình * Biên soạn tài liệu * Huấn luyện sử dụng |
| **Đối tượng sử dụng:**  Người dùng có nhu cầu tìm kiếm những quyển sách đang được đánh giá cao hoặc tìm kiếm những tựa sách liên quan dựa trên một đầu sách đã có. |
| **Môi trường triển khai**  Bài toán sử dụng collaborative filtering để gợi ý sách triển khởi dưới môi trường của: Visual Studio Code  Ngôn ngữ sử dụng: Python và HTML. |

## **2.2. Phân tích và thiết kế hệ thống gợi ý**

Một hệ thống gợi ý sách thường được chia thành 3 quá trình chính:

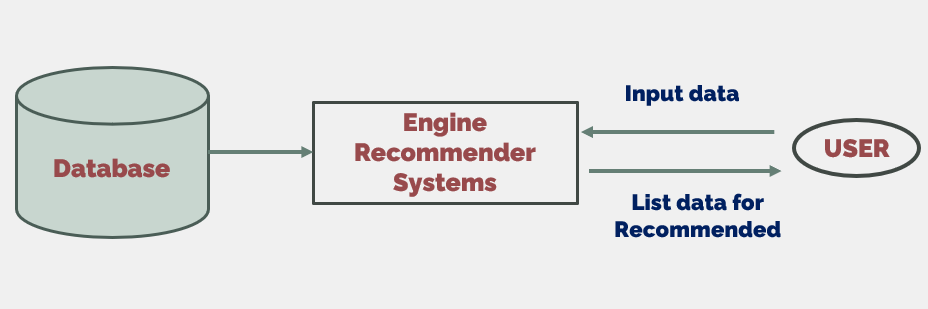
Bước 1: Tìm các đặc trưng (features) có ảnh hưởng đến việc đánh giá của người dùng, thông qua việc phân tích và thăm dò dữ liệu

Ở bài nghiên cứu của nhóm chúng em, các đặc trưng ảnh hưởng đến hệ thống gợi ý là lượt bình chọn (rating) của từng đầu sách dựa trên người dùng trước.

Bước 2: Phân tích và áp dụng giải thuật filtering phù hợp

Cho nên hệ thống gợi ý sách tiến hành phân tích dựa trên từng Items, qua ma trận tương tương đồng.

Bước 3: Tiến hành training mô hình

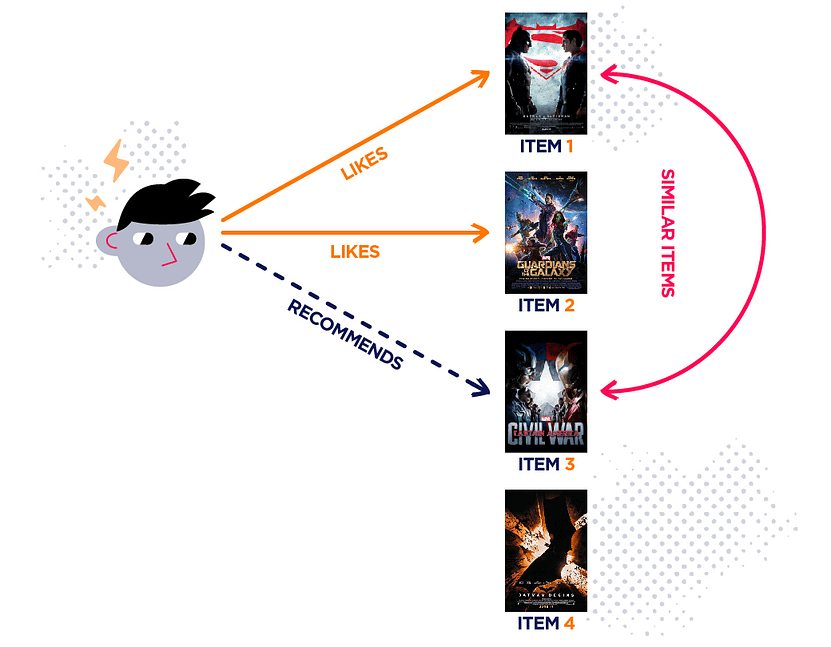


Hình 7. Mô hình training

### **2.3. Sử dụng Item-based** **Collaborative Filtering để gợi ý sách**

Item-Item Collaborative Filtering**:** sẽ đề xuất những cuốn sách đã được tất cả người dùng trong hệ thống xếp hạng theo cách tương tự.

Ví dụ: Những cuốn sách có kiểu xếp hạng giống nhau bởi tất cả người dùng.

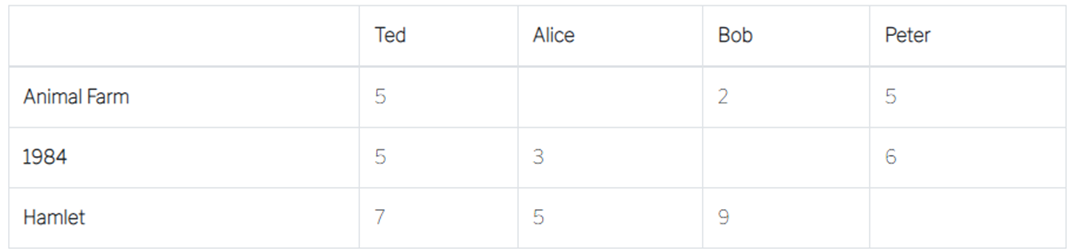


Hình 8. Item-based Collaborative filtering

Với Item-Item Collaborative Filtering, ta sẽ tìm kiếm những cuốn sách mới tương tự như những cuốn sách mà người dùng đã chọn được đánh giá cao.

**Ưu điểm của phương pháp này:**

* Thường có ít item hơn người dùng trong cơ sở dữ liệu, do đó, việc tạo ma trận tương tự (similar matrix) sẽ có xu hướng nhỏ hơn
* Người dùng mới sẽ được đề xuất rất nhanh mặc dù họ chỉ mới xếp hạng một vài mục.

**Xây dựng Item-based Collaborative Filtering gồm có 2 bước sau:**  
Bước 1: Sắp xếp dữ liệu thành ma trận item-user (MxN)  
Tại đây, ta có thể xem tất cả các xếp hạng cho một cuốn sách nhất định và mỗi hàng sách sau đó sẽ được coi là một vectơ N chiều trong đó N là số lượng người dùng. Ở đây, những cuốn sách đã được người dùng xếp hạng tương tự có thể cho thấy một cuốn sách phù hợp với một cuốn sách khác (ví dụ: Animal Farm và 1984 có các mẫu xếp hạng tương tự từ người dùng, cho thấy rằng nếu ai đó thích Animal Farm, họ có thể thích 1984).  


Bước 2: Xây dựng ma trận tương tự (similar matrix)  
Có nhiều cách để bạn có thể tìm ra sự giống nhau giữa hai cuốn sách, ở đây ta sẽ tìm sự giống nhau về cosine.  
  
Công thức tính đồng dạng Cosine:

Similatity () =   
  
Ta sẽ sử dụng công thức Cosine tính xếp hạng cho một cuốn sách cụ thể bằng cách sử dụng tổng trọng số xếp hạng của các cuốn sách tương tự khác.

Rating(U,Ii) =

Bây giờ chúng ta có một ma trận điểm tương đồng giữa tất cả các cuốn sách độc nhất (xem Animal Farm có điểm tương đồng cao nhất với 1984 như thế nào, vì vậy những người dùng thích Animal Farm cũng thích 1984, ngược lại những người dùng không thích Animal Farm cũng không thích 1984).  


## **2.4. Bộ dữ liệu**

### **2.4.1. Mô tả chi tiết bộ dữ liệu**

Tên bộ dữ liệu: Books dataset

Dữ liệu trực tuyến cho sách từ Amazon cùng với xếp hạng người dùng và người dùng đã mua chúng.

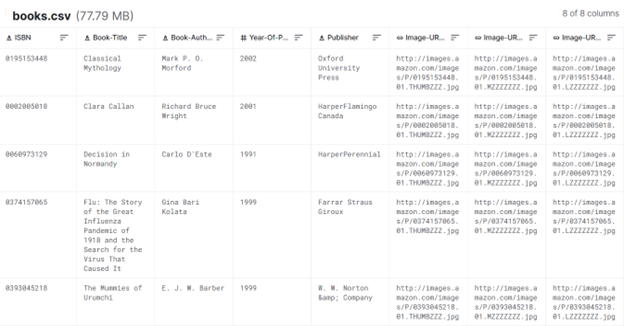
Bộ dữ liệu đã được biên soạn bởi CAI-Nicolas Ziegler vào năm 2004, và nó bao gồm ba bảng: users, ratings, books.

* **User.csv**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | STT | Tên Cột | Ý nghĩa | | 1 | User-ID | Mã người dùng | | 2 | Location | Nơi sinh sống của người dùng | | 3 | Age | Tuổi người dùng | |  |

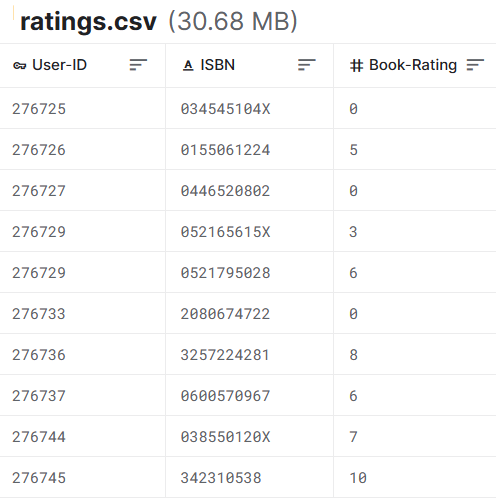
* **Books.csv**

|  |  |  |
| --- | --- | --- |
| STT | Tên Cột | Ý nghĩa |
| 1 | ISBN | ISBN viết tắt của International Standard Book Number, là một hệ thống mã số duy nhất được sử dụng để định danh các sách. |
| 2 | Book-Title | Tên sách |
| 3 | Book-Author | Tên tác giả |
| 4 | Year-Of-Publication | Năm xuất bản |
| 5 | Publisher | Nhà xuất bản |
| 6 | Image-URL-S | Kích cỡ hình ảnh của bìa sách \_ nhỏ |
| 7 | Image-URL-M | Kích cỡ hình ảnh của bìa sách \_ vừa |
| 8 | Image-URL-L | Kích cỡ hình ảnh của bìa sách \_ to |



* **Rating.csv**

|  |  |  |
| --- | --- | --- |
| STT | Tên Cột | Ý nghĩa |
| 1 | User-ID | Mã người dùng |
| 2 | ISBN | Mã sách được đánh giá |
| 3 | Book-Rating | Đánh giá của người dùng  Đánh giá được thể hiện trên thang điểm từ 1-10 (giá trị cao hơn biểu thị sự đánh giá cao hơn) và đánh giá ngầm được thể hiện bằng 0 |



### **2.4.2. Data mining**

Data mining – khai phá dữ liệu là quá trình phân loại, sắp xếp các tập hợp dữ liệu nhất định để xác định xu hướng, các mẫu và thiết lập các mối liên hệ hữu ích nhằm giải quyết các vấn đề nhờ phân tích dữ liệu. Mục tiêu của việc này là cho phép các doanh nghiệp có thể dự đoán được xu hướng tương lai, nhằm đưa ra các quyết định được hỗ trợ dữ liệu từ các tập dữ liệu khổng lồ.

Quá trình khai phá dữ liệu là một quá trình phức tạp bao gồm kho dữ liệu chuyên sâu cũng như các công nghệ tính toán.

**Các bước quan trọng khi khai phá dữ liệu bao gồm:**

Bước 1: Làm sạch dữ liệu – Trong bước này, dữ liệu được làm sạch sao cho không có dữ liệu thừa hay bất thường trong dữ liệu.

Bước 2: Tích hợp dữ liệu – Trong quá trình tích hợp dữ liệu, nhiều nguồn dữ liệu sẽ kết hợp lại thành một.

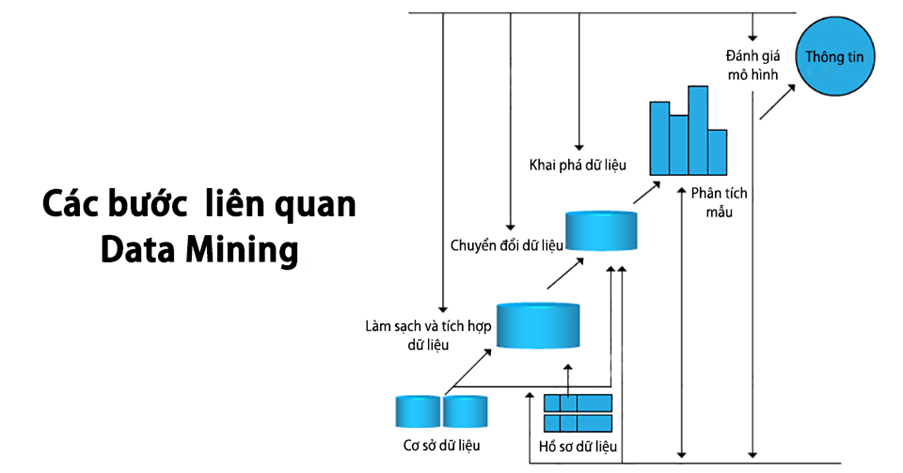
Bước 3: Lựa chọn dữ liệu – Trong bước này, dữ liệu được trích xuất từ cơ sở dữ liệu.

Bước 4: Chuyển đổi dữ liệu – Trong bước này, dữ liệu sẽ được chuyển đổi để thực hiện phân tích tóm tắt cũng như các hoạt động tổng hợp.

Bước 5: Khai phá dữ liệu – Trong bước này, chúng tôi trích xuất dữ liệu hữu ích từ nhóm dữ liệu hiện có.

Bước 6: Đánh giá mẫu – Chúng tôi phân tích một số mẫu có trong dữ liệu.

Bước 7: Trình bày thông tin – Trong bước cuối cùng, thông tin sẽ được thể hiện dưới dạng cây, bảng, biểu đồ và ma trận.



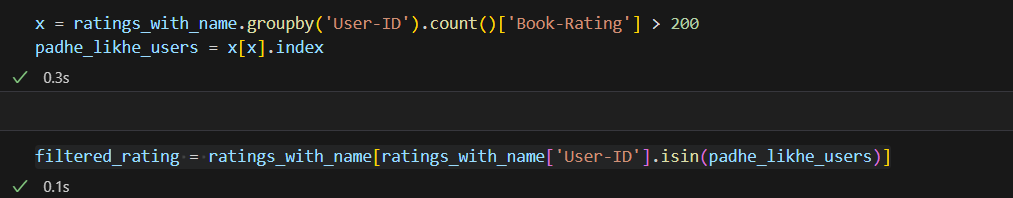
Hình 9. Data mining

**Tiến hành khai phá dữ liệu:**

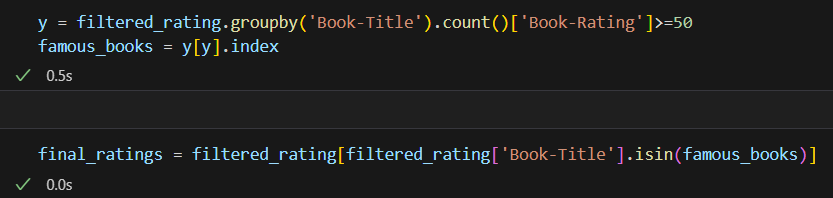
Bộ dữ liệu nhóm tiến hành nghiên cứu đã qua xử lý bước 1,2,3,4 nên chúng em tiến hành trích xuất dữ liệu, áp dụng trực tiếp vào bài toán gợi ý sách.

Bước 5: Trích xuất dữ liệu

* Trích xuất Users và Ratings của hơn 200

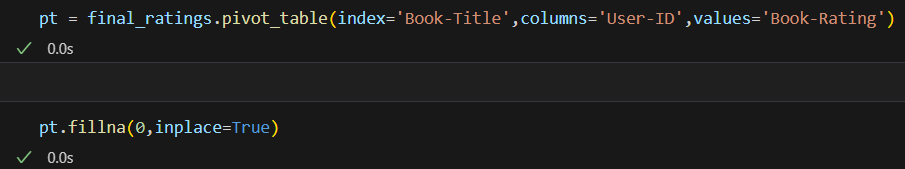
****

Trích xuất ID người dùng đã đưa ra hơn 200 xếp hạng và khi có id người dùng, sẽ chỉ trích xuất xếp hạng của id người dùng này từ khung dữ liệu xếp hạng.

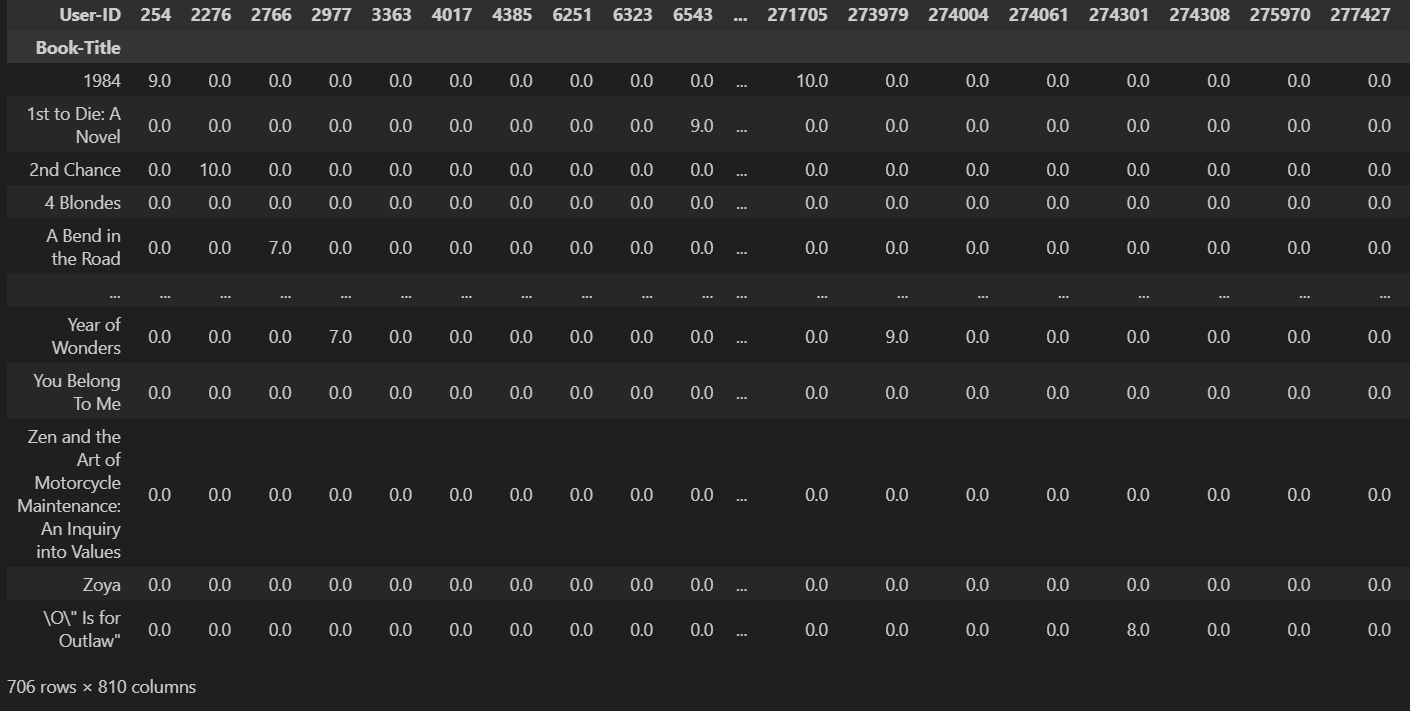
* Trích xuất những cuốn sách đã nhận được hơn 50 xếp hạng.

Đếm xếp hạng của từng cuốn sách, nhóm dữ liệu dựa trên tên sách và tổng hợp dựa trên xếp hạng. Cần phải loại bỏ các giá trị trùng lặp vì nếu cùng một người dùng đã xếp hạng cùng một cuốn sách nhiều lần thì điều đó sẽ tạo ra sự cố. Cuối cùng, có một bộ dữ liệu với người dùng đó đã xếp hạng hơn 200 cuốn sách và những cuốn sách đã nhận được hơn 50 xếp hạng. hình dạng của khung dữ liệu cuối cùng là 474007 hàng và 8 cột.

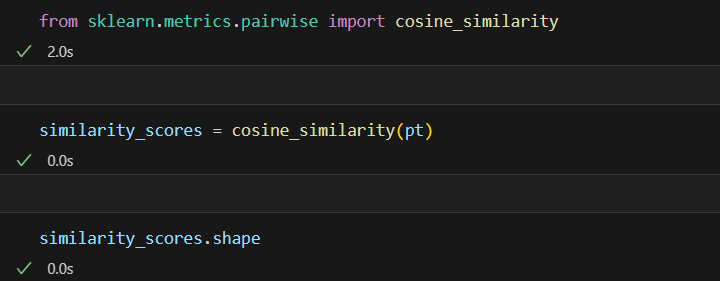
Bước 6: Tạo Pivot Table

****

Tạo một bảng tổng hợp trong đó các cột sẽ là ID người dùng, chỉ mục sẽ là tên sách và giá trị là xếp hạng. Và ID người dùng chưa xếp hạng bất kỳ cuốn sách nào sẽ có giá trị là NAN, vì vậy quy nó bằng 0.



Bước 7: Xây dựng hàm recommend





# **3. XÂY DỰNG ỨNG DỤNG**

## **3.1. Yêu cầu cài đặt trước một số công cụ sau:**

* Python 3.6 trở lên: [Download Python | Python.org](https://www.python.org/downloads/)
* Jupiter notebook (cài đặt trong cmd)

pip install notebook

* Cài đặt thư viện cần thiết cho Python: numpy, pandas, sklearn

pip install numpy

pip install pandas

pip install sklearn

* Các tools IDE như Visual Studio Code

## **3.2. Cấu trúc tổ chức thư mục**

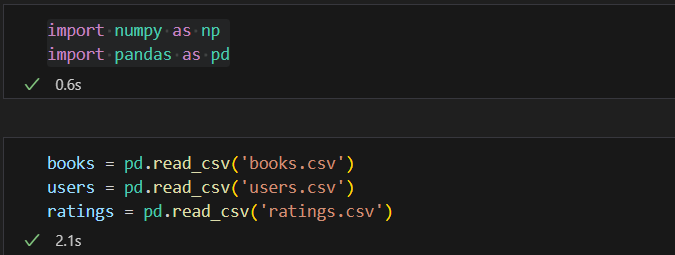
|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Thư mục giao diện web:**   * index.html: giao diện chính * recommend.html: giao diện gợi ý sách khi nhập tên sách * error.html: thông báo lỗi khi không có dữ liệu | **Thư mục dữ liệu:**   * Books.csv: dữ liệu sách * Ratings.csv: dữ liệu đánh giá sách * Users.csv: dữ liệu người dùng | **Danh mục các file chạy web:**   * app.py: định nghĩa và chạy ứng dụng web * book-recommender-system.ipynb: thuật toán gợi ý sách * Procflie: c ung cấp một cách cấu hình để quản lý quy trình và triển khai ứng dụng web |

## **3.3. Xây dựng trang web**

### **3.3.1. Xây dựng thuật toán gợi ý sách**

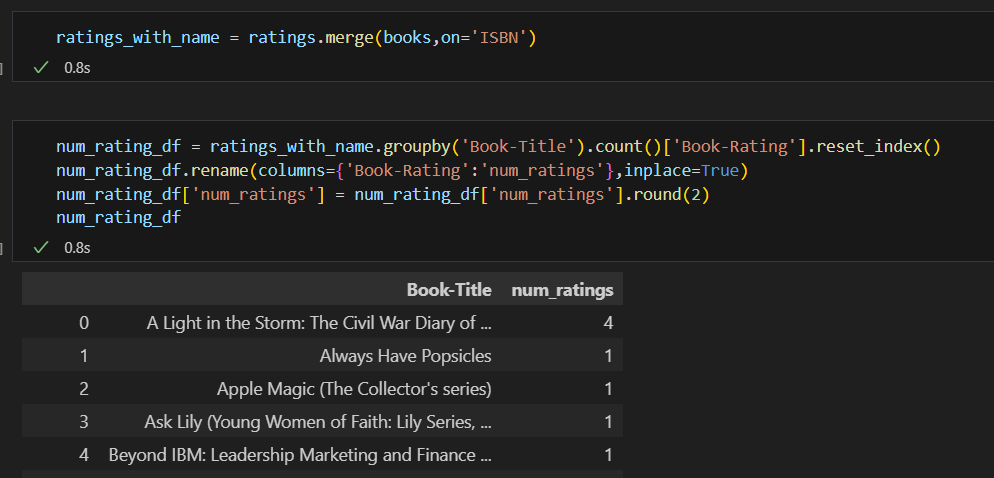
* **Đọc dữ liệu đầu vào**

Đọc dữ liệu đầu vào gồm 3 tệp: books.csv, users.csv, ratings.csv

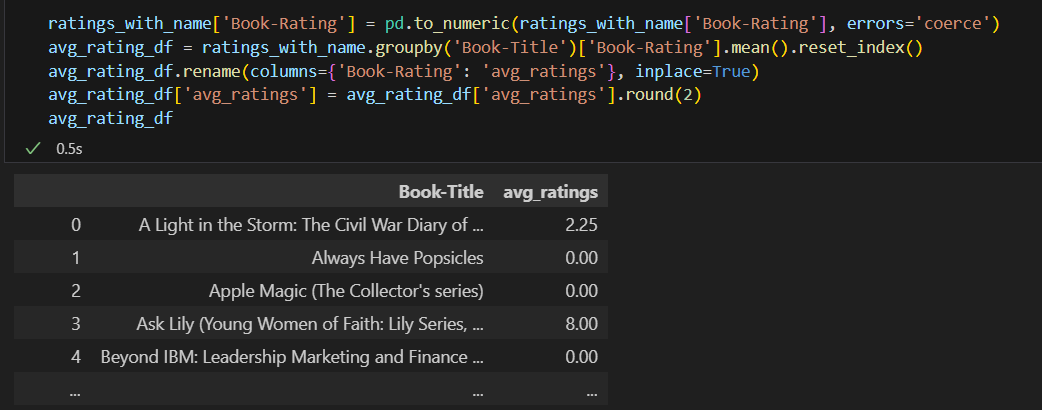
****

* **Gợi ý sách dựa trên mức độ phổ biến**

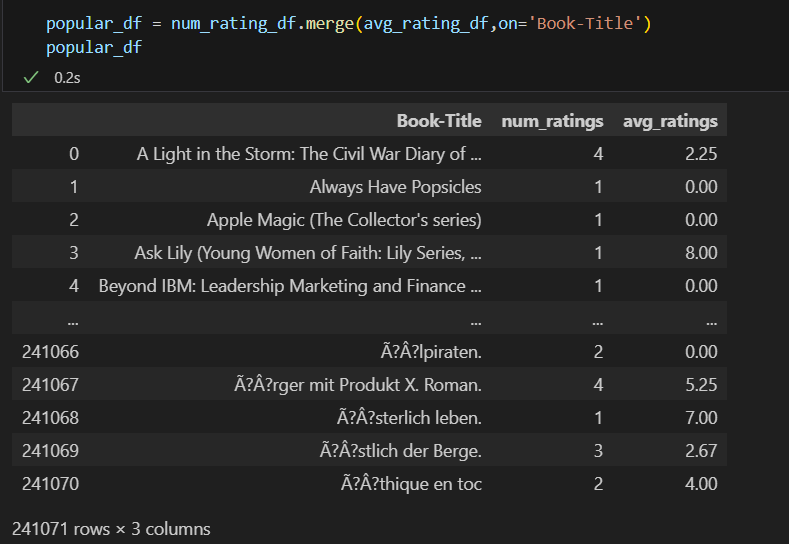
Tạo bảng mới gồm có Book-Title (tên sách), num\_ratings (đánh giá)

****

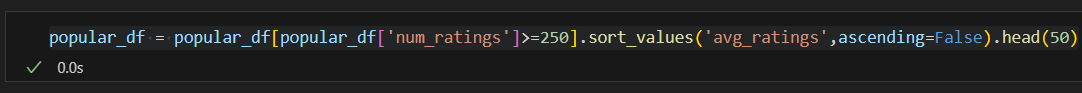
Tạo bảng chứa thông tin về tiêu đề sách ('Book-Title') và giá trị trung bình đánh giá ('avg\_ratings') cho mỗi tiêu đề sách làm mới vị trí các phần tử từ lớn đến bé.



Tạo bảng đánh giá độ phổ biến chứa thông tin về số lượng đánh giá và giá trị trung bình đánh giá được kết hợp lại cho cùng một tiêu đề sách.



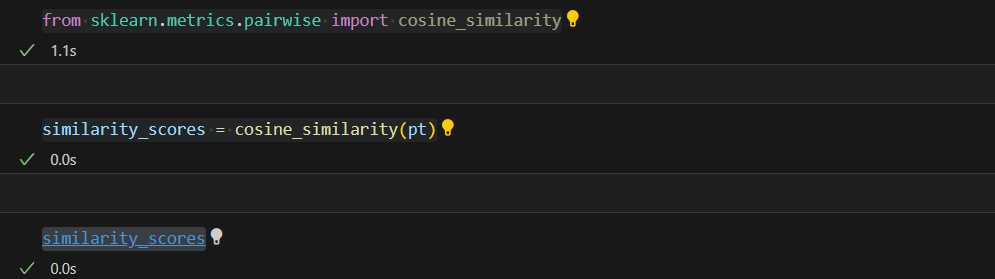
Câu lên thực hiện lọc thông tin về 50 tiêu đề sách phổ biến nhất. Để được xem là phổ biến, một tiêu đề sách cần có số lượng đánh giá ít nhất là 250, giá trị trung bình đánh giá từ cao đến thấp.



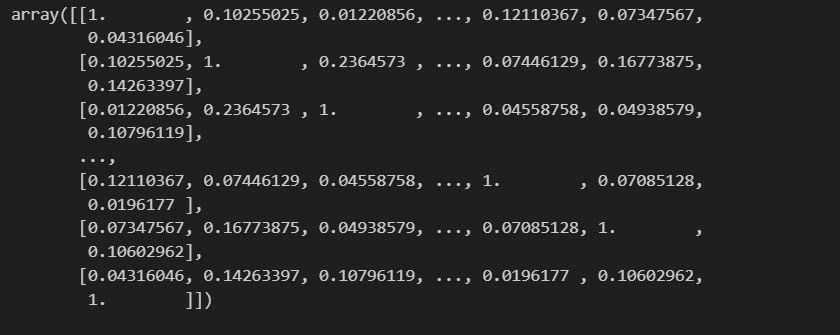
* **Xử lý Collaborative filtering**

Dữ liệu dầu vào là các trích xuất đã xử lý ở mục 2.4.2

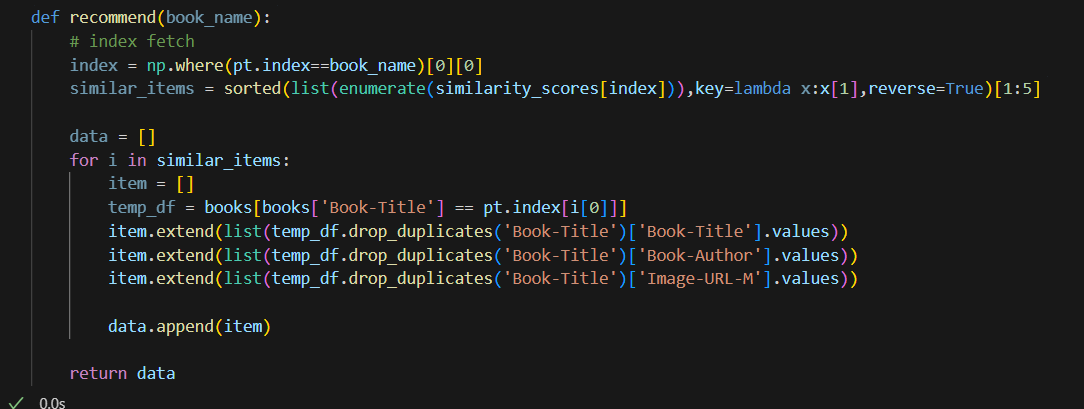
Tiến hành xử lý các dữ liệu đã được nhận



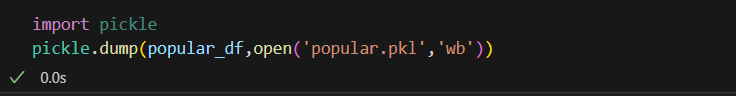
Thu được ma trận được sử dụng để tìm kiếm các tiêu đề sách tương tự nhau dựa trên các đặc trưng đánh giá của người dùng. Càng gần giá trị 1 thể hiện mức độ tương tự cao hơn giữa các tiêu đề sách, trong khi giá trị càng gần 0 thể hiện sự khác biệt lớn hơn giữa chúng.

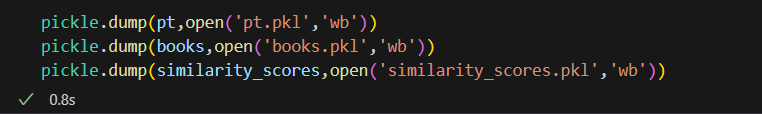


Tạo hàm recommend() để tạo một danh sách các cuốn sách tương tự dựa trên tên sách đầu vào. Kết quả trả về là một danh sách chứa thông tin về các cuốn sách tương tự, bao gồm tên sách, tác giả và URL hình ảnh.



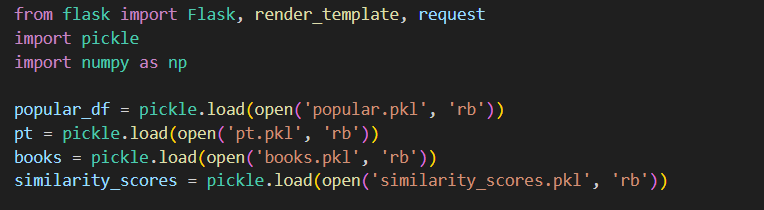
Sau khi tính toán từ các dữ liệu thô, các bảng kết quả sẽ được lưu lại bằng file .pkl sử dụng thư viện pickle.





### **3.3.2. Xây dựng hiện thị web**

Sử dụng các thư viện cần thiết. Lấy dữ liệu từ các file đã lưu trước hệ thống sử dụng thư viện Pickle

****

Trong thư mục template có chứa các tệp html

Trỏ tới bằng render\_template('index.html'), có các biến được truyền vào như book\_name, author, image, votes, và rating. Các biến này lấy giá trị từ cột tương ứng trong popular\_df, chứa thông tin về các cuốn sách phổ biến.

Trong trang'index.html', các giá trị được truyền vào có thể được sử dụng để hiển thị thông tin về các cuốn sách phổ biến trên giao diện người dùng, bao gồm tên sách, tác giả, URL hình ảnh, số lượt đánh giá và đánh giá trung bình.

****

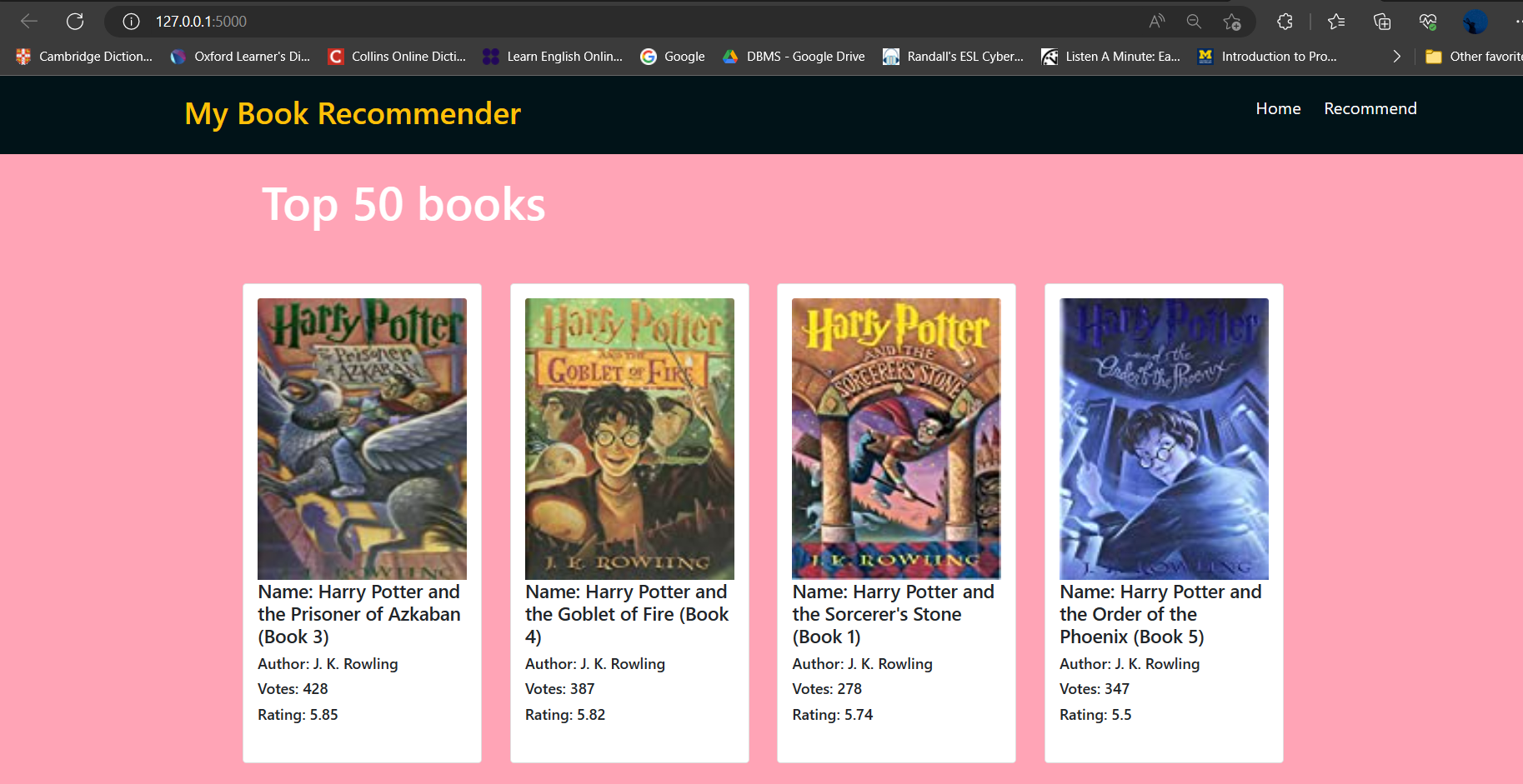
Trong‘recommend.html’ với dữ liệu để hiển thị danh sách các cuốn sách tương tự lên trang gợi ý sách. Nếu không tìm thấy dữ liệu, ta trả về template HTML error.html với thông báo lỗi tương ứng.

****

1. **WESITE GỢI Ý SÁCH**

**Giao diện chính**

Hiện thị top 50 quyển sách có lượt đánh giá tốt nhất theo thứ tự giảm dần

****

Hình 10. Giao diện chính của web

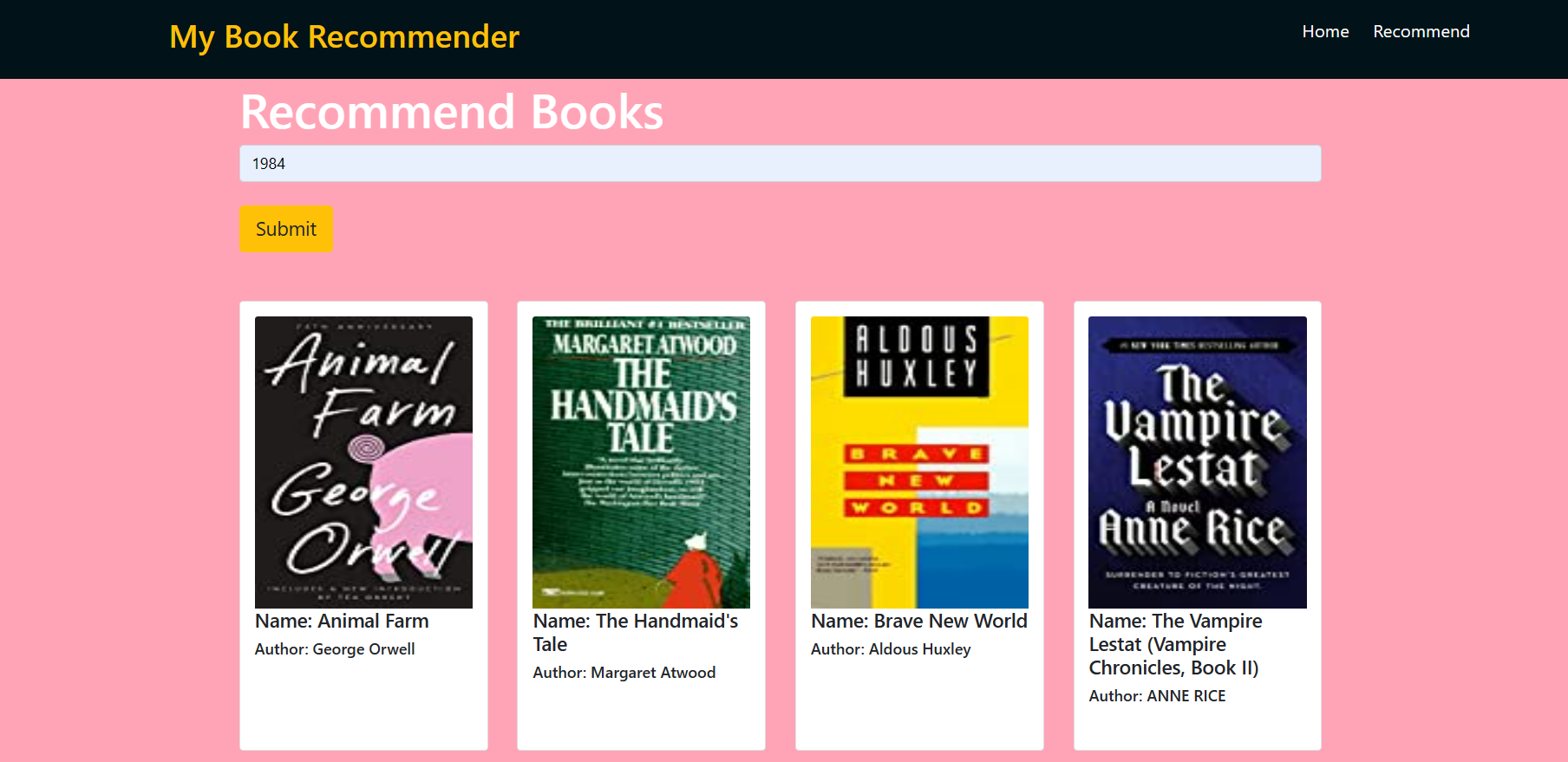
**Giao diện Recommend**

Khi chọn chức năng Recommend sẽ hiện thị giao diện



Hình 11. Giao diện chức năng Recommend

Khi thực hiện tìm kiếm tên sách, giao diện sẽ hiện thị những tựa sách liên quan.

****

Hình 12. Giao diện hiện thị sách gợi ý

Khi thực hiện tìm kiếm tên sách không có trong dữ liệu, sẽ hiện thị giao diện thông báo lỗi

****

****

Hình 13. Giao diện thông báo lỗi

1. **CÁC CÔNG TRÌNH NGHIÊN CỨU LIÊN QUAN**

Ngày nay các hệ thống gợi ý được áp dụng rộng rãi trên nhiều lĩnh vực nhằm mang đến trải nghiệm cá nhân hóa tốt nhất cho người dùng. Đặc biệt, với lượng sách lớn được xuất bản mỗi năm, hệ thống gợi ý sách là hệ thống được quan tâm nhiều. Dưới đây là một số bài viết liên quan đến hệ thống gợi ý sách

Bài báo của Missi Hikmatyar, Ruuhwan được công bố năm 2020, với tiêu đề "Book Recommendation System Development Using User-Based Collaborative Filtering", mô tả về việc phát triển một hệ thống gợi ý sách dựa trên phương pháp User-Based Collaborative Filtering. Hệ thống gợi ý sách trong thư viện Đại học Perjuangan ở Tasikmalaya, Indonesia. Mục đích của hệ thống là giúp người dùng thư viện tìm kiếm sách phù hợp với từng cá nhân, nghiên cứu trên sự tương đồng giữa các thành viên trong cộng đồng người dùng dựa trên mẫu mượn sách và nhóm dựa trên chủ đề của sách được tìm kiếm. Quá trình tìm kiếm sách gợi ý được hiển thị từ cao đến thấp và cung cấp các đầu sách thích hợp cho người dùng dựa trên từng cá nhân. [8]

"Book Recommendation using Collaborative Filtering",2023 của tác giả Prasanta Kumar Sahoo, Venkat Ramana, S Dhanish, A Nikith Kumar tập trung vào ứng dụng của phương pháp lọc cộng tác trong việc giới thiệu sách cho người dùng. Nhấn mạnh rằng việc lựa chọn sách phù hợp trở nên khó khăn do sự đa dạng và số lượng sách có sẵn. Sử dụng ma trận tương tác người dùng-sản phẩm và ma trận vector để xây dựng hệ thống giới thiệu sách, sử dụng đo đạc tương đồng cosine để đề xuất sách tương tự nhất cho người dùng.[9] Nhóm chúng em thực hiện nghiên cứu tương tự trong hệ thống gợi ý sách.

Bài viết "College Library Personalized Recommendation System Based on Hybrid Recommendation Algorithm" (2019), của tác giả Yonghong Tian, Bing Zheng, Yanfang Wang, Yue Zhang, Qi Wu. Đề tài nghiên cứu này tập trung vào việc áp dụng hệ thống gợi ý cá nhân trong việc giới thiệu sách trong thư viện đại học. Vấn đề hiện tại là thư viện đại học hàng năm đưa ra một lượng lớn sách mới, dẫn đến việc người dùng phải mất nhiều thời gian để chọn sách. Đồng thời, nhiều sách không được sử dụng hiệu quả, gây lãng phí tài nguyên của thư viện. Vấn đề này được gây ra bởi hiện tượng "quá tải thông tin". Để giải quyết vấn đề này, thư viện cần dựa vào cơ chế lọc thông tin. Cơ chế lọc thông tin được chia thành hai loại: cơ chế tìm kiếm và cơ chế gợi ý. Cơ chế tìm kiếm sử dụng từ khóa để giúp người dùng nhanh chóng tìm thấy sách liên quan, trong khi cơ chế gợi ý tự động đề xuất sách cho người dùng. Hệ thống gợi ý cá nhân nhằm dự đoán sở thích dựa trên quan tâm, hành vi và thông tin khác của người dùng. Gợi ý cá nhân không chỉ cung cấp nhu cầu của người dùng, mà còn giúp người dùng khám phá và khám phá sở thích mới. Ứng dụng hệ thống gợi ý trong thư viện đại học giải quyết vấn đề khó khăn trong việc chọn sách và cải thiện tỷ lệ sử dụng tài nguyên của thư viện. Hệ thống gợi ý kết hợp dựa trên sinh viên đại học, và cải tiến cơ chế lọc cộng tác bằng Hybrid Recommendation. Hybrid Recommendation cung cấp một cách tiếp cận linh hoạt và mạnh mẽ để tăng cường chất lượng gợi ý. Bằng cách kết hợp nhiều phương pháp gợi ý, có thể khắc phục nhược điểm của từng phương pháp đơn lẻ và mang lại gợi ý cá nhân hóa chính xác hơn cho người dùng.[10]

1. **KẾT LUẬN**

Hệ thống đề xuất sách sử dụng Collaborative Filtering là một giải pháp để gợi ý các cuốn sách tương tự dựa trên hành vi đọc sách của người dùng. Hệ thống này hoạt động bằng cách xây dựng một ma trận đánh giá, trong đó các hàng đại diện cho các cuốn sách và các cột đại diện cho các người dùng.

Đầu tiên, dữ liệu đánh giá sách được tiền xử lý để loại bỏ nhiễu và tính toán đánh giá trung bình cho mỗi cuốn sách. Sau đó, chỉ những người dùng có hành vi đọc sách đáng tin cậy (có hơn 200 đánh giá) được lựa chọn để xây dựng ma trận đánh giá cuối cùng. Chỉ những cuốn sách phổ biến (có hơn 50 đánh giá) được lựa chọn để tính toán độ tương tự giữa chúng.

Tiếp theo, ma trận đánh giá được sử dụng để tính toán độ tương tự cosine giữa các cuốn sách. Độ tương tự cosine đo lường sự tương đồng giữa các cuốn sách dựa trên hành vi đọc sách của người dùng. Kết quả là một ma trận độ tương tự.

Cuối cùng, khi người dùng yêu cầu gợi ý sách dựa trên một cuốn sách cụ thể, hệ thống tìm kiếm cuốn sách trong ma trận đánh giá và sử dụng độ tương tự để tìm ra các cuốn sách tương tự. Các cuốn sách tương tự được trả về và hiển thị trên giao diện người dùng.

Hệ thống đề xuất sách sử dụng Collaborative Filtering có ưu điểm là dựa trên thông tin thực tế từ hành vi đọc sách của người dùng và có thể tạo ra những gợi ý cá nhân hóa. Tuy nhiên, hệ thống này cũng có nhược điểm là cần có đủ dữ liệu đánh giá từ người dùng để tạo ra các gợi ý chính xác, và có thể gặp khó khăn khi xử lý các vấn đề như hiện tượng "lạc hướng theo hàng loạt" khi người dùng chỉ đánh giá một số ít cuốn sách. Hệ thống còn đơn giản, chưa khắc phục được một số lỗi tìm kiếm gợi ý sách bằng từ khóa.

# **NGUỒN DỮ LIỆU**

**Dữ liệu:** [Books Dataset | Kaggle](https://www.kaggle.com/datasets/saurabhbagchi/books-dataset)

# **TÀI LIỆU THAM KHẢO**

[1] Vihar Kurama, 2022, What Is Collaborative Filtering: A Simple Introduction, website Builtin

[What Is Collaborative Filtering: A Simple Introduction | Built In](https://builtin.com/data-science/collaborative-filtering-recommender-system)

[2] Phạm Đình Khánh, 2019, Collaborative và Content-based filtering, website Khoa học dữ liệu - Khanh's blog

[Khoa học dữ liệu (phamdinhkhanh.github.io)](https://phamdinhkhanh.github.io/2019/11/04/Recommendation_Compound_Part1.html#221-neighborhood-based-collaborative-filtering)

[3] Hoàng Đinh, 2021, Xây dựng Collaborative Filtering RS [Recommender System cơ bản - Phần 3], website Viblo

[Xây dựng Collaborative Filtering - Recommender System cơ bản (viblo.asia)](https://viblo.asia/p/xay-dung-collaborative-filtering-rs-recommender-system-co-ban-phan-3-Az45bMqolxY)

[4] Vũ Hữu Tiệp, 2017, Matrix Factorization Collaborative Filtering, website Machinelearningcoban

[Machine Learning cơ bản (machinelearningcoban.com)](https://machinelearningcoban.com/2017/05/31/matrixfactorization/)

[5] Yehuda Koren,Robert Bell and Chris Volinsky, 2019, Matrix Factorization Techniques For Recommender Systems

[Recommender-Systems-[Netflix].pdf (datajobs.com)](https://datajobs.com/data-science-repo/Recommender-Systems-%5BNetflix%5D.pdf)

[6] Hoàng Đinh, 2022, Data Mining - Khai phá dữ liệu - [Data Science Series], website Viblo

[Data Mining - Khai phá dữ liệu - [Data Science Series] - Viblo](https://viblo.asia/p/data-mining-khai-pha-du-lieu-data-science-series-1Je5EAX45nL)

[7] Kiên Nguyễn, 2019, Hướng dẫn từng bước lập trình web với Python, website TopDev

[Hướng dẫn từng bước lập trình web với Python | TopDev](https://topdev.vn/blog/lap-trinh-web-voi-python/)

[8] Missi Hikmatyar, Ruuhwan, 2020, Book Recommendation System Development Using User-Based Collaborative Filtering

[(PDF) Book Recommendation System Development Using User-Based Collaborative Filtering (researchgate.net)](https://www.researchgate.net/publication/340652227_Book_Recommendation_System_Development_Using_User-Based_Collaborative_Filtering?_sg=fwfSSZkDqeyvo3ApzBj0yzOCVrTdbesLMkXb6Jf-8JOCzlqWCL3rPTrHgH3F-QUIZLP4pCw67F-i4EM)

[9] Prasanta Kumar Sahoo, Venkat Ramana, S Dhanish, A Nikith Kumar, 2023, Book Recommendation using Collaborative Filtering

[(PDF) Book Recommendation using Collaborative Filtering (researchgate.net)](https://www.researchgate.net/publication/370444164_Book_Recommendation_using_Collaborative_Filtering)

[10] Yonghong Tian, Bing Zheng, Yanfang Wang, Yue Zhang, Qi Wu, 2019, College Library Personalized Recommendation System Based on Hybrid Recommendation Algorithm, website Sciencedirect, tr.490 – tr.494

[College Library Personalized Recommendation System Based on Hybrid Recommendation Algorithm - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S2212827119307401)