

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
ĐẠI HỌC KINH TẾ TP HỒ CHÍ MINH  
TRƯỜNG CÔNG NGHỆ VÀ THIẾT KẾ



## ĐỒ ÁN MÔN HỌC

ĐỀ TÀI:

UEH LIBRARY BOOK RECOMMENDATION:  
ỨNG DỤNG GIẢI THUẬT DI TRUYỀN TRONG VIỆC  
XÂY DỰNG HỆ THỐNG KHUYẾN NGHỊ

**Học phần:** Trí Tuệ Nhân Tạo

**Nhóm Sinh Viên:**

1. HOÀNG ĐỨC DÂN
2. LÊ ĐỨC HẬU
3. TRƯƠNG THỊ HỒNG MAI
4. VÕ MINH NGUYỄN

**Chuyên Ngành:** KHOA HỌC DỮ LIỆU

**Khóa:** K47

**Giảng Viên:** TS. Đặng Ngọc Hoàng Thành

TP. Hồ Chí Minh, Ngày 5 tháng 8 năm 2023

## LÝ DO CHỌN ĐỀ TÀI

Trong thực tế, hệ thống khuyến nghị sách đã trở thành một phần quan trọng trong nhiều thư viện và nền tảng kỹ thuật số. Đối với các trang web thương mại điện tử lớn như Amazon, hệ thống khuyến nghị sách giúp cải thiện trải nghiệm mua sắm của khách hàng bằng cách đề xuất những sản phẩm tương tự hoặc liên quan dựa trên hành vi mua hàng của họ. Điều này giúp tăng cơ hội bán hàng và cải thiện sự hài lòng của khách hàng.

Đối với các thư viện số và nền tảng kỹ thuật số, hệ thống khuyến nghị sách cung cấp một công cụ hữu ích để giúp người dùng khám phá những đầu sách mới và thú vị dựa trên sở thích và hành vi đọc sách của họ. Thay vì phải tìm kiếm và chọn lựa từ hàng ngàn đầu sách có sẵn, người dùng có thể nhận được những gợi ý chính xác và cá nhân hóa để tiết kiệm thời gian và nỗ lực.

Tuy nhiên, UEH Digital Smart Library hiện tại chưa có hệ thống khuyến nghị sách tích hợp. Điều này có thể là một hạn chế đối với trải nghiệm của người dùng khi sử dụng thư viện kỹ thuật số này. Thiếu hệ thống khuyến nghị sách có thể làm giảm khả năng khám phá sách mới và mang lại trải nghiệm đọc sách không tối ưu cho người dùng.

Việc thiếu hệ thống khuyến nghị sách cũng có thể dẫn đến việc người dùng không nhận thấy được giá trị của thư viện kỹ thuật số và chuyển sang sử dụng các nền tảng khác có tích hợp tính năng khuyến nghị sách. Điều này có thể ảnh hưởng đến sự phát triển và ứng dụng của UEH Digital Smart Library trong việc hỗ trợ và phục vụ người dùng một cách tốt nhất.

Do đó, việc xây dựng và triển khai hệ thống khuyến nghị sách là một bước tiến quan trọng trong việc nâng cao trải nghiệm của người dùng và tăng tính ứng dụng và hiệu quả của UEH Digital Smart Library. Hệ thống này sẽ giúp cải thiện việc khám phá sách, tăng khả năng phục vụ của thư viện và thu hút sự quan tâm và sử dụng của cộng đồng đọc sách và sinh viên tại trường Đại học Kinh tế TP.HCM.

Nhóm nghiên cứu đã chọn đề tài "**Hệ thống khuyến nghị sách sử dụng giải thuật di truyền, lọc tương tác và đa lớp**" trong nghiên cứu này vì một số lý do quan trọng.

Trước tiên, hệ thống khuyến nghị sách đang trở nên ngày càng phổ biến và quan trọng trong thế giới số hóa. Trong thời đại thông tin vô tận, người dùng thường gặp khó khăn

trong việc tìm kiếm và chọn lựa sách phù hợp với sở thích và nhu cầu của họ. Hệ thống khuyến nghị sách giúp giải quyết vấn đề này bằng cách cung cấp các đề xuất sách cá nhân hóa dựa trên hành vi và tương tác của người dùng với hệ thống.

Thứ hai, nhóm nghiên cứu đã quyết định áp dụng mô hình BLI<sub>GA</sub> từ một bài nghiên cứu đã được chứng minh là hiệu quả trong việc khuyến nghị sách. Bằng cách sử dụng mô hình đã có sẵn, chúng tôi tiết kiệm thời gian và tập trung vào việc tùy chỉnh mô hình cho bối cảnh cụ thể của UEH Digital Smart Library.

Thứ ba, việc thu thập dữ liệu của sinh viên là rất quan trọng để đảm bảo tính chính xác và hiệu quả của hệ thống khuyến nghị. Chúng tôi đã tiến hành thu thập dữ liệu để hiểu rõ hơn về sở thích và yêu cầu của người dùng khi sử dụng thư viện kỹ thuật số của trường. Dữ liệu này sẽ đóng vai trò quan trọng trong việc cải thiện mô hình và đưa ra những đề xuất sách chính xác hơn.

Thứ tư, chúng tôi cũng đặc biệt quan tâm đến việc xây dựng giao diện cho UEH Digital Smart Library. Giao diện tương tác dễ sử dụng và thân thiện đối với người dùng là yếu tố quan trọng giúp hệ thống khuyến nghị được áp dụng và sử dụng rộng rãi. Chúng tôi đã tập trung vào việc xây dựng một giao diện trực quan và dễ dàng sử dụng để giúp người dùng tận dụng những đề xuất sách một cách hiệu quả và hấp dẫn.

Cuối cùng, chúng tôi đã chọn đề tài này vì nó cho phép thực hiện một dự án nghiên cứu ứng dụng thực tế và đánh giá hiệu quả của các phương pháp dự báo trong bối cảnh thực tế của UEH Digital Smart Library. Điều này là rất quan trọng để đảm bảo tính ứng dụng và thực tiễn của kết quả nghiên cứu trong lĩnh vực khuyến nghị sách.

Tóm lại, lựa chọn đề tài "**Hệ thống khuyến nghị sách sử dụng giải thuật di truyền, lọc tương tác và đa lớp**" là một quyết định có ý nghĩa và mang lại giá trị cao trong việc nghiên cứu và ứng dụng trong lĩnh vực khuyến nghị sách trong thời đại số hóa ngày nay.

# MỤC LỤC

<b>LÝ DO CHỌN ĐỀ TÀI.....</b>	<b>2</b>
<b>CÁC THUẬT NGỮ TRONG BÀI.....</b>	<b>7</b>
<b>DANH MỤC HÌNH ẢNH.....</b>	<b>8</b>
<b>DANH MỤC BẢNG.....</b>	<b>9</b>
<b>DANH MỤC TỪ VIẾT TẮT .....</b>	<b>10</b>
<b>PHẦN 1. TỔNG QUAN.....</b>	<b>11</b>
1. Giới thiệu về hệ thống khuyến nghị (Recommendation System) .....	11
2. Phân loại của hệ thống khuyến nghị (Recommendation System).....	11
2.1. <i>Collaborative Filtering</i> .....	11
2.2. <i>Content-Based Filtering</i> .....	12
3. Hybrid Filtering.....	13
<b>PHẦN 2. COLABORATIVE FILTERING .....</b>	<b>15</b>
1. Giới thiệu về Colaborative Filtering .....	15
2. Phân loại Colaborative Filtering .....	17
2.1. <i>Tiếp cận dựa trên bộ nhớ (Memory-based Collaborative Approach)</i> .....	17
2.2. <i>Colaborative Filtering dựa trên người dùng (User-based Collaborative Filtering)</i> .....	17
2.3. <i>Colaborative Filtering dựa trên vật phẩm (Item-based Collaborative Filtering)</i> .....	17
2.4. <i>Tiếp cận dựa trên mô hình (Model-based Collaborative Approach)</i> .....	18
3. Ứng dụng Item-Based Collaborative Filtering cho hệ thống khuyến nghị (Recommendation System) .....	20
3.1. <i>Lựa chọn vùng lân cận/ hàng xóm</i> .....	20
3.2. <i>Dự đoán xếp hạng</i> .....	20
<b>PHẦN 3. GIẢI THUẬT DI TRUYỀN.....</b>	<b>22</b>
1. Giới thiệu về giải thuật di truyền .....	22
2. Cách hoạt động của giải thuật di truyền .....	22
2.1. <i>Khởi tạo</i> .....	22
2.2. <i>Fitness Function</i> .....	22
2.3. <i>Lựa chọn (Selection)</i> .....	22
2.4. <i>Sinh sản (Reproduction)</i> .....	23

2.5. Thay thế (Replacement) .....	23
2.6. Chấm dứt (Termination) .....	23
<b>PHẦN 4. ỨNG DỤNG GIẢI THUẬT DI TRUYỀN VÀO HỆ THỐNG KHUYẾN NGHỊ .....</b>	<b>24</b>
1. Các nghiên cứu trước đây .....	24
2. Sử dụng đa tiêu chí và thông tin ngữ nghĩa trong hệ thống khuyến nghị.....	25
3. Hệ thống khuyến nghị dựa trên di truyền được đề xuất .....	26
3.1. Biểu diễn di truyền và khởi tạo quần thể.....	29
3.2. Cấp độ lọc.....	30
a. Tiêu chí lọc tương quan ngữ nghĩa .....	30
b. Tiêu chí lọc dựa trên sự hài lòng .....	32
c. Lựa chọn danh sách các mặt hàng tốt nhất.....	33
3.3. Các toán tử di truyền và tiêu chí kết thúc.....	34
<b>PHẦN 5. CÁC KẾT QUẢ THỰC NGHIỆM.....</b>	<b>36</b>
1. Bộ dữ liệu dùng cho thí nghiệm.....	36
3. Kết quả hệ thống cho giá trị ngẫu nhiên.....	37
4. Các tiêu chí đánh giá .....	39
4.1. Mean Absolute Error (MAE):.....	39
4.2. Precision: .....	40
4.3. Recall: .....	40
4.4. F1-measure: .....	40
5. Kết quả đánh giá.....	40
6. Điều chỉnh siêu tham số.....	41
<b>PHẦN 6. THIẾT KẾ GIAO DIỆN CHO HỆ THỐNG.....</b>	<b>42</b>
1. Xây dựng giao diện cho hệ thống .....	42
1.1. Qt Designer & PyQt5.....	42
1.2. Ý tưởng thiết kế giao diện cho hệ thống.....	42
2. Thiết kế giao diện. ....	43
2.1. Nút chức năng .....	43
2.2. Chức năng.....	43
a. Tìm kiếm tên sách.....	44
b. Đánh giá sách. ....	44

<i>c. Hiện thị danh sách khuyến nghị.....</i>	<i>45</i>
<i>d. Thông báo chờ.....</i>	<i>45</i>
<i>e. Thông báo người dùng chưa nhập đánh giá:.....</i>	<i>46</i>
3. Kết quả giao diện .....	46
<b>PHẦN 7. KẾT LUẬN.....</b>	<b>49</b>
1. Các kết quả đạt được .....	49
2. Những hạn chế và hướng phát triển .....	50
<b>PHỤ LỤC .....</b>	<b>51</b>
<b>TƯ LIỆU THAM KHẢO.....</b>	<b>88</b>

## CÁC THUẬT NGỮ TRONG BÀI

## DANH MỤC HÌNH ẢNH

Hình 1: Minh họa lọc User to User Filtering trong Collaborative Filtering .....	12
Hình 2: Minh họa cho phương pháp Content-Based Filtering .....	13
Hình 3: Minh họa cho phương pháp Hybrid Filtering .....	14
Hình 4: Minh họa User-based Collaborative Filtering .....	17
Hình 5: Minh họa Item-based Collaborative Filtering .....	18
Hình 6: Một số ví dụ về Matrix Factorization .....	19
Hình 7: Mô hình xây dựng hệ thống khuyến nghị dựa trên di truyền .....	27
Hình 8: Quy trình thực hiện hệ thống khuyến nghị dựa trên giải thuật di truyền .....	29
Hình 9. Biểu đồ phân phối các giá trị của Bộ dữ liệu 1 .....	36
Hình 10. Biểu đồ phân phối các giá trị của Bộ dữ liệu 1 .....	37
Hình 11. Biểu đồ phân phối các giá trị của Bộ dữ liệu 3 .....	37
Hình 13. Giá trị trung bình của dự đoán qua từng thể hệ với AU ngẫu nhiên .....	39
Hình 14. UEH Smart Digital Library .....	43
Hình 15. Mô phỏng giao diện chờ của hệ thống .....	47
Hình 16. Mô phỏng giao diện đánh giá sách của hệ thống .....	47
Hình 17. Mô phỏng giao diện cho kết quả trả về của hệ thống .....	48
Hình 18. Giá trị hàm mục tiêu qua các thế hệ của các tổ hợp siêu tham số .....	58



## **DANH MỤC BẢNG**

Bảng 1. Thông tin về các bộ dữ liệu sử dụng .....	36
Bảng 2. Kết quả đánh giá cho các bộ dữ liệu .....	41
Bảng 3. Thống kê kết quả cho các tổ hợp siêu tham số .....	87

## DANH MỤC TỪ VIẾT TẮT

# PHẦN 1. TỔNG QUAN

## 1. Giới thiệu về hệ thống khuyến nghị (Recommendation System)

Recommendation System hoặc Recommender Systems là một lớp con của hệ thống lọc thông tin tìm cách dự đoán “xếp hạng” hoặc “ưu tiên” mà người dùng sẽ cung cấp cho một mặt hàng (chẳng hạn như âm nhạc, sách hoặc phim ảnh) hoặc yếu tố xã hội (ví dụ: con người hoặc nhóm) mà họ chưa xem xét, sử dụng một mô hình được xây dựng từ các đặc điểm của một mục (phương pháp tiếp cận dựa trên nội dung) hoặc môi trường xã hội của người dùng (lọc cộng tác phương pháp tiếp cận). Mặc dù nhiều cách tiếp cận khác nhau để hệ thống gợi ý đã được phát triển trong vài năm qua năm, sự quan tâm trong lĩnh vực này vẫn còn cao do nhu cầu ngày càng tăng về các ứng dụng thực tế, có khả năng cung cấp các đề xuất được cá nhân hóa và giải quyết quá tải thông tin. Những nhu cầu ngày càng tăng này đặt ra một số quan trọng những thách thức đối với các hệ thống tư vấn và để đối phó với những bài toán nhiều kỹ thuật nâng cao được đề xuất như nội dung tăng cường lọc cộng tác, lọc dựa trên phân cụm, kết hợp sự tương đồng dựa trên vật phẩm và người dùng và nhiều hơn. Bất chấp những tiến bộ này, các hệ thống tư vấn vẫn yêu cầu cải tiến và do đó trở thành một lĩnh vực nghiên cứu phong phú. Trong bài báo này, trước khi thảo luận về những hạn chế chính của phương pháp đề xuất, khảo sát toàn diện về phương pháp khuyến nghị được cung cấp. Cuộc thảo luận của cách tiếp cận khác nhau và những hạn chế của họ một cách thích hợp do đó cung cấp các khả năng nghiên cứu trong tương lai trong hệ thống khuyến nghị.

## 2. Phân loại của hệ thống khuyến nghị (Recommendation System)

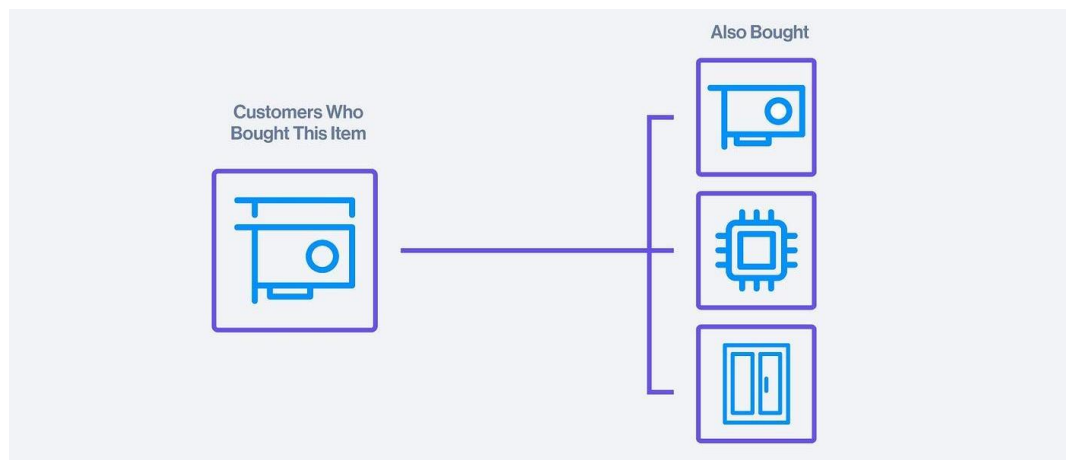
Tùy thuộc vào các sản phẩm hoặc dịch vụ mà doanh nghiệp cung cấp, các hệ thống khuyến nghị khác nhau có thể được áp dụng. Một số ví dụ về các hệ thống khác nhau là:

### 2.1. Collaborative Filtering

Collaborative Filtering tập trung vào sự giống nhau giữa những người dùng và vật phẩm khác nhau. Những người dùng có chung sở thích giống nhau sẽ có nhiều khả năng quan tâm đến các sản phẩm tương tự khác. Những điểm tương đồng này có thể cải thiện đề xuất cho tất cả người dùng trong tập dữ liệu và tiếp tục tìm hiểu khi sản phẩm mới ra mắt thị trường.

Có một số loại lọc Collaborative Filtering:

- **User to Product Filtering** là phương pháp đơn giản nhất trong tất cả các phương pháp lọc, trong đó thuật toán sẽ tìm kiếm các mặt hàng tương tự mà người tiêu dùng đã mua hoặc thích trước đó. Thể loại, giá cả, danh mục mặt hàng, v.v. đều là những danh mục ảnh hưởng đến quá trình lọc.
- **User to User Filtering** dùng hoạt động bằng cách tìm những người tiêu dùng có chung sở thích và đề xuất các sản phẩm và dịch vụ dựa trên những gì người dùng trông giống anh ta đã chọn. Thuật toán như vậy yêu cầu tính toán và tài nguyên cao vì thuật toán sẽ cần so sánh tất cả người dùng trong thời gian thực.



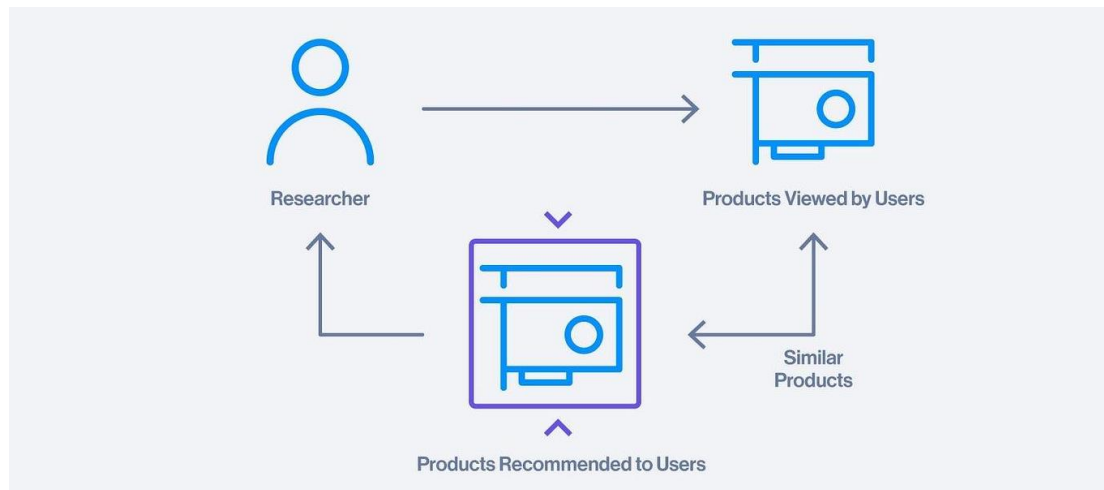
Hình 1: Minh họa lọc User to User Filtering trong Collaborative Filtering

## 2.2. Content-Based Filtering

Content-Based Filtering dựa trên nội dung đánh giá mức độ giống nhau của các sản phẩm. hệ thống khuyến nghị sẽ đề xuất các sản phẩm có phân loại tương tự với người dùng đã tương tác trước đó.

Ví dụ: Nếu ba bộ phim đã xem gần đây nhất bao gồm thể loại hài, hệ thống sẽ đề xuất các bộ phim hoặc chương trình hài tương tự khác. Những đề xuất như vậy cũng bắt buộc đối với hình ảnh sản phẩm sử dụng Xử lý hình ảnh hoặc Xử lý ngôn ngữ tự nhiên để khớp các mặt hàng có hình thức, tiêu đề hoặc mô tả tương tự.

Lưu ý rằng các đề xuất dựa trên sự tương đồng sẽ gặp phải vấn đề bắt đầu nguội (**the cold start problem**). Sự cố khởi động nguội xảy ra khi không có đủ dữ liệu tùy chọn. hệ thống khuyến nghị không thể đề xuất chính xác các tùy chọn tuyệt vời khi được triển khai lần đầu trên nền tảng vì cần có thời gian để thu thập và đào tạo.



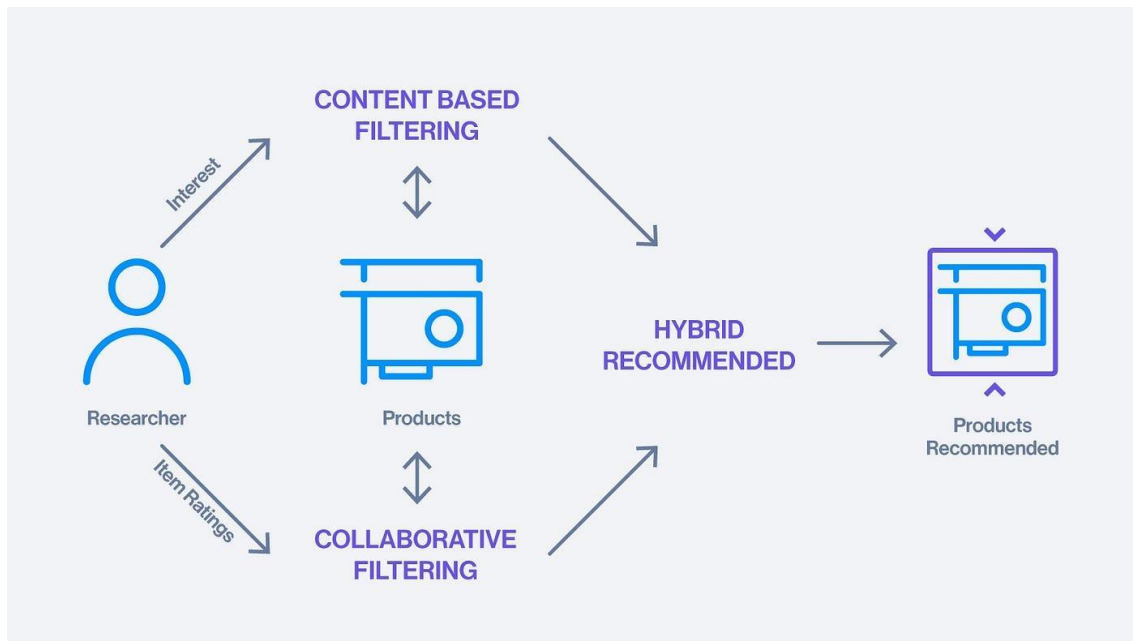
Hình 2: Minh họa cho phương pháp Content-Based Filtering

### 3. Hybrid Filtering

Hybird Filtering sử dụng cả Collaborative Filtering và Content-Based Filtering, sử dụng các ưu điểm của nhau.

Một số nghiên cứu so sánh hiệu suất của hệ thống Hybird Filtering với riêng Collaborative Filtering và Content-Based Filtering đã chỉ ra rằng hệ thống Hybird có độ chính xác cao hơn.

Kết hợp cả hai thuật toán có thể loại bỏ nhiều vấn đề như vấn đề khởi động nguội (**the cold start problem**) và giúp thu thập dữ liệu nhanh chóng. Nhiều trang web yêu thích của chúng tôi như Google, Youtube và Netflix sử dụng bộ lọc kết hợp trong hệ thống khuyến nghị của họ.



Hình 3: Minh họa cho phương pháp Hybrid Filtering

## PHẦN 2. COLABORATIVE FILTERING

### 1. Giới thiệu về Colaborative Filtering

Có hai loại hệ thống khuyến nghị, lọc dựa trên nội dung (**Content-based Filtering**) và lọc cộng tác (**Collaborative Filtering**). Tính năng lọc dựa trên nội dung sử dụng các thuật toán máy học để dự đoán và đề xuất các mặt hàng mới nhưng tương tự cho người dùng. Nó sử dụng các tính năng của vật phẩm để nhóm các vật phẩm tương tự lại với nhau.

Collaborative Filtering chỉ sử dụng các tương tác trong quá khứ giữa khách hàng và sản phẩm họ đã sử dụng để đề xuất mặt hàng mới. Các tính năng của mục không quan trọng vì các tương tác giữa người dùng và mục được sử dụng và được lưu trữ trong ma trận tương tác giữa người dùng và mục.

Trong Collaborative Filtering, tất cả người dùng đều được xem xét và những người có sở thích và thị hiếu tương tự được sử dụng để đề xuất các sản phẩm mới và cụ thể cho khách hàng chính. Nó giúp các công ty và khách hàng bắt kịp xu hướng.

Hai loại tương tác giữa người dùng và sản phẩm được ghi lại:

- Đầu tiên là thông qua các tìm kiếm trực tiếp và các hành động phản hồi ngầm như nhấp chuột, lịch sử đặt hàng, phát nội dung nhất định, v.v.
- Thứ hai là thông qua phản hồi trực tiếp từ người dùng. Ví dụ: xếp hạng một bộ phim mà họ đã xem theo thang điểm từ 1 đến 5 sao, thích hoặc không thích một video trên YouTube, 'gắn dấu sao' cho một album hoặc danh sách phát trên Spotify để đánh dấu nó là mục yêu thích, v.v.

### Tại sao chúng ta cần hệ thống khuyến nghị?

Trở lại năm 2006, Netflix đã đưa ra giải thưởng để giải quyết một vấn đề đơn giản đã tồn tại trong nhiều năm. Đó là tìm thuật toán hợp tác tốt nhất để dự đoán xếp hạng của người dùng cho những bộ phim mà họ chưa xem, dựa trên xếp hạng trước đó của các bộ phim khác.

Ngày nay, những gã khổng lồ thương mại điện tử tiếp tục cố gắng giải quyết vấn đề này theo cách tốt hơn bằng cách quan sát hành vi trong quá khứ của người dùng để dự đoán những thứ khác mà người dùng đó sẽ thích. Tại sao? Bởi vì biết trước những gì sẽ cung

cấp có thể thúc đẩy lợi nhuận của họ bằng cách tăng doanh số bán hàng và nâng cao trải nghiệm của khách hàng.

Các đề xuất cũng giúp khách hàng khám phá các sản phẩm và ưu đãi mới mà họ không tìm kiếm một cách rõ ràng, do đó đẩy nhanh quá trình tìm kiếm. Điều này cho phép các công ty gửi các bản tin được cá nhân hóa qua email cung cấp các chương trình truyền hình, phim, sản phẩm và dịch vụ mới phù hợp hơn với họ.

Một trong những lợi thế quan trọng nhất của các thuật toán đề xuất hiện đại là khả năng nhận phản hồi ngầm và đề xuất nội dung/sản phẩm mới, do đó luôn cập nhật các tùy chọn của khách hàng. Điều này cho phép các doanh nghiệp tiếp tục phục vụ khách hàng ngay cả khi thị hiếu của họ thay đổi theo thời gian.

### **Ma trận tương tác người dùng – sản phẩm (User-item Interaction Matrix)**

Trong Collaborative Filtering, chúng tôi bỏ qua các tính năng của một sản phẩm riêng lẻ. Thay vào đó, chúng tôi tập trung vào một nhóm người tương tự đang sử dụng mặt hàng đó và đề xuất các mặt hàng khác mà nhóm đó thích.

Những người dùng tương tự được chia thành các cụm nhỏ và được đề xuất các mặt hàng mới theo sở thích của cụm đó. Hãy hiểu điều này với một ví dụ đề xuất phim dễ dàng:

Users	Movie 1	Movie 2	Movie 3	Movie 4
User 1	5	4		5
User 2	4		3	
User 3		1		2
User 4	1	2		

*Image source: GeeksforGeeks*

Những gì chúng ta có thể suy ra từ ma trận người dùng – sản phẩm này là:

- Người dùng 1 và 2 thích Phim 1. Vì Người dùng 1 rất thích phim 2 và 4 nên có nhiều khả năng Người dùng 2 cũng thích như vậy.
- Người dùng 1 và 3 có thị hiếu trái ngược nhau.
- Cả người dùng 3 và 4 đều không thích Movie 2, vì vậy có khả năng cao là Người dùng 4 cũng sẽ không thích Movie 4.
- Người dùng 3 có thể không thích Phim 1.



Đây là logic đằng sau việc sử dụng ma trận tương tác giữa ma trận người dùng – sản phẩm - để tìm các nhóm người dùng tương tự thông qua Collaborative Filtering.

## 2. Phân loại Collaborative Filtering

Hai loại phương pháp tiếp cận Collaborative Filtering là:

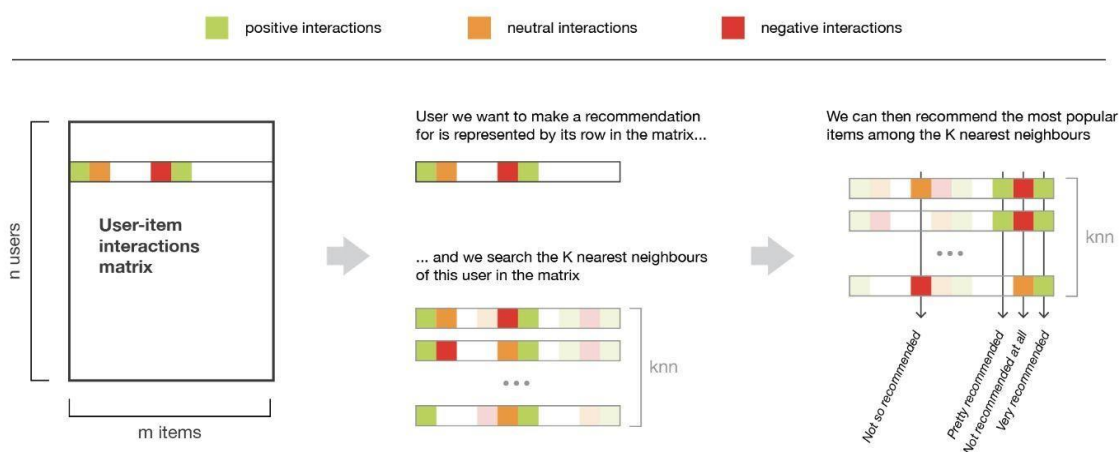
### 2.1. Tiếp cận dựa trên bộ nhớ (Memory-based Collaborative Approach)

Trong Collaborative Filtering dựa trên bộ nhớ, chỉ ma trận tương tác giữa người dùng và sản phẩm được sử dụng để đưa ra đề xuất mới cho người dùng. Toàn bộ quá trình dựa trên xếp hạng và tương tác trước đó của người dùng.

Lọc dựa trên bộ nhớ bao gồm 2 phương pháp: Collaborative Filtering dựa trên người dùng (**User-based Collaborative Filtering**) và Collaborative Filtering dựa trên sản phẩm (**Item-based Collaborative Filtering**)

### 2.2. Collaborative Filtering dựa trên người dùng (User-based Collaborative Filtering)

Để đề xuất các đề xuất mới cho một người dùng cụ thể, một nhóm người dùng tương tự (hàng xóm gần nhất) được tạo dựa trên tương tác của người dùng tham chiếu. Các sản phẩm phổ biến nhất trong nhóm này, nhưng mới đối với người dùng mục tiêu, được sử dụng cho các đề xuất.

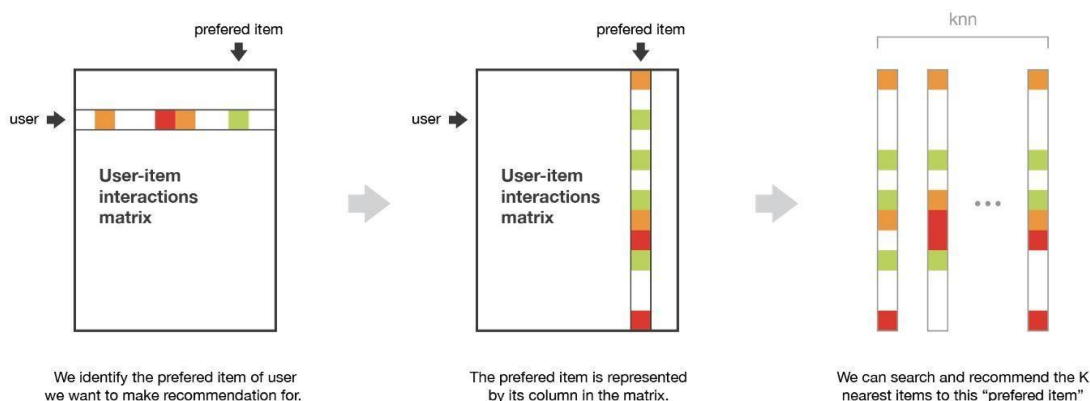


Hình 4: Minh họa User-based Collaborative Filtering

Image source: Towards Data Science

### 2.3. Collaborative Filtering dựa trên vật phẩm (Item-based Collaborative Filtering)

Trong lọc dựa trên sản phẩm, các đề xuất mới được chọn dựa trên các tương tác cũ của người dùng mục tiêu. Đầu tiên, tất cả các sản phẩm mà người dùng đã thích đều được xem xét. Sau đó, các sản phẩm tương tự được tính toán và tạo thành các cụm (hàng xóm gần nhất). Các sản phẩm mới từ các cụm này được đề xuất cho người dùng.



Hình 5: Minh họa Item-based Collaborative Filtering

Image source: Towards Data Science

## 2.4. Tiếp cận dựa trên mô hình (Model-based Collaborative Approach)

Trong cách tiếp cận dựa trên mô hình, các mô hình máy học được sử dụng để dự đoán và xếp hạng các tương tác giữa người dùng và các mục mà họ chưa tương tác. Các mô hình này được đào tạo bằng cách sử dụng thông tin tương tác đã có sẵn từ ma trận tương tác bằng cách triển khai các thuật toán khác nhau như hệ số hóa ma trận, học sâu, phân cụm, v.v.

### Hệ số hoá ma trận (Matrix Factorization)

Hệ số hóa ma trận được sử dụng để tạo các tính năng tiềm ẩn bằng cách phân tách ma trận tương tác người dùng – sản phẩm thừa thớt thành hai ma trận nhỏ hơn và dày đặc của các thực thể người dùng và sản phẩm.

Quay trở lại ví dụ về phim ở trên, giả sử chúng ta có một ma trận thừa thớt gồm 4 người dùng và 4 phim, xếp hạng từ 1 đến 5:

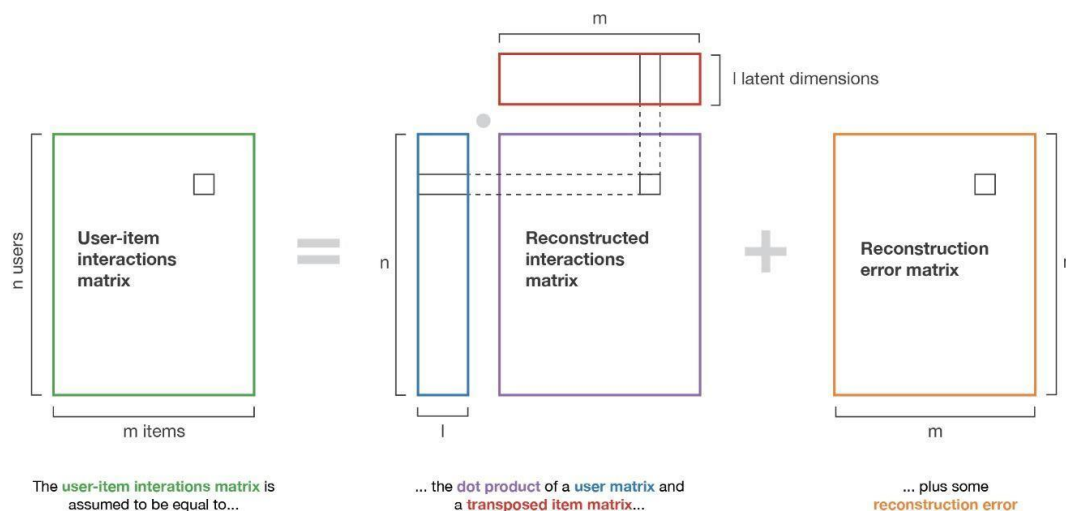
Users	Movie 1	Movie 2	Movie 3	Movie 4
User 1	5	4		5
User 2	4		3	
User 3		1		2
User 4	1	2		

Image source: *GeeksforGeeks*

Vì không phải tất cả các bộ phim đều được mọi người dùng xem và đánh giá, nên chúng tôi kết thúc bằng một ma trận thưa thớt. Để tạo một mô hình cho ma trận của chúng ta, chúng ta có thể giả định rằng:

- Tồn tại một số tính năng tiềm ẩn có thể phân biệt giữa phim hay và phim dở.
- Các tính năng này có thể giúp chúng tôi hiểu các lựa chọn của người dùng (giá trị càng cao, tùy chọn càng cao).

Chúng ta không cung cấp các tính năng này một cách rõ ràng mà để mô hình khám phá các tính năng hữu ích và tạo ma trận người dùng và vật phẩm của nó. Vì các tính năng được học và không được cung cấp, chúng có mối tương quan và ý nghĩa toán học nhưng không có sự hiểu biết trực quan.



Hình 6: Một số ví dụ về Matrix Factorization

Image source: *Towards Data Science*

### 3. Ứng dụng Item-Based Collaborative Filtering cho hệ thống khuyến nghị (Recommendation System)

Kỹ thuật này sử dụng ma trận *Sản phẩm x Sản phẩm* để lưu trữ điểm tương đồng giữa các mục. Trong thực tế, hệ thống sẽ khuyến nghị các sản phẩm tương tự nhất với một tập hợp các sản phẩm đã được xếp hạng với số điểm cao bởi người dùng tích cực. Thật vậy, đánh giá dự đoán phụ thuộc vào giá trị tương tự giữa sản phẩm và hàng xóm của nó: sự tương đồng càng tăng, xếp hạng dự đoán càng cao tương tự như hàng xóm của xếp hạng.

Giống như kỹ thuật CF dựa trên người dùng, quy trình CF dựa trên sản phẩm có thể được tóm tắt thành ba các bước: tính toán độ tương tự, lựa chọn vùng lân cận và xếp hạng sự dự đoán.

#### 3.1. Lựa chọn vùng lân cận/ hàng xóm

Giống như kỹ thuật dựa trên người dùng, có hai đề xuất phương pháp: Số tối đa (top-k) của hàng xóm ([Shardanand và Maes, 1995](#)) và ngưỡng tương quan ([Resnick et al., 1994](#)). Các kỹ thuật top-k chọn k mục gần nhất, trong đó k biểu thị số hạng mục. Kỹ thuật ngưỡng tương quan thiết lập một ngưỡng và duy trì các sản phẩm có điểm tương đồng, với sản phẩm có sẵn vượt quá ngưỡng.

#### 3.2. Dự đoán xếp hạng

Để dự đoán xếp hạng, một số biện pháp đã được đề xuất. Các các biện pháp phổ biến hơn như sau: Điểm số Z ([Ricci et al., 2010](#)), Tổng trọng số ([Sarwar et al., 2001](#)) và Lấy trung bình là trung tâm tổng hợp/Mean-Centered aggregation ([Aggarwal, 2016](#)). Công thức tổng trọng số như sau:

$$\widetilde{r}_{ui} = \frac{\sum_{v \in N_u^i} Sim_{uv} * r_{vi}}{\sum_{v \in N_u^i} |Sim_{uv}|}$$

Trong đó,

$N_u^i$  là tập hợp những người hàng xóm (top-k) giống nhau nhất cho người dùng u và đã xếp hạng mục i; v là người dùng thuộc về  $N_u^i$  và  $Sim_{uv}$  là giá trị tương tự giữa người dùng u và v.

Ngoài ra, chức năng dự đoán tập trung vào trung bình được đề xuất bởi ([Aggarwal, 2016](#)) là một biện pháp tổng hợp phổ biến. phương trình 2 chương trình công thức lấy trung bình giới hạn trung tâm (chúng tôi giữ các biến giống như phương trình trên):

$$\widetilde{r_{uu}} = r_u^- + \frac{\sum_{v \in N_u^i} Sim_{uv} * (r_{vi} - r_v^-)}{\sum_{v \in N_u^i} |Sim_{uv}|}$$

trong đó  $r_u$  và  $r_v$  là xếp hạng trung bình do người dùng u và v chỉ định tương ứng.

## PHẦN 3. GIẢI THUẬT DI TRUYỀN

### 1. Giới thiệu về giải thuật di truyền

Thuật toán di truyền bắt đầu với một tập các giải pháp (ngẫu nhiên) gọi là quần thể. Mỗi cá thể trong quần thể được gọi là bộ nhiễm sắc thể trong đó đại diện cho một giải pháp. Mỗi cá thể được đánh giá bởi Fitness Function. các thế hệ là Các nhiễm sắc thể phát triển thông qua các lần lặp. Số lần lặp tối đa có thể là được xác định trước hoặc một điều kiện dừng chẳng hạn như tìm một chức năng tập thể dự mong muốn có thể được chỉ định. Một trong những tính năng quan trọng nhất của GA là đại diện cho vấn đề với một chuỗi các ký hiệu được gọi là gen.

Biểu diễn thường được hình thành bởi nhị phân, mảng có giá trị thực hoặc giá trị nguyên. mã hóa nhị phân xảy ra bởi chuỗi 1-0 và cần thêm dung lượng bộ nhớ trong khi tính toán. Mã hóa số nguyên chứa thực các giá trị đại diện cho **giải pháp [6]**. một ma trận biểu diễn được sử dụng trong công trình nghiên cứu này. Các toán tử di truyền phổ biến (chọn lọc, lai chéo và đột biến) được sử dụng trong thí nghiệm.

### 2. Cách hoạt động của giải thuật di truyền

Các thuật toán di truyền sử dụng chu kỳ thế hệ tiến hóa để tạo ra các giải pháp chất lượng cao. Họ sử dụng các hoạt động khác nhau làm tăng hoặc thay thế dân số để cung cấp giải pháp phù hợp được cải thiện. Các thuật toán di truyền tuân theo các giai đoạn sau để giải các bài toán tối ưu phức tạp:

#### 2.1. Khởi tạo

Thuật toán di truyền bắt đầu bằng cách tạo ra một quần thể ban đầu. Dân số ban đầu này bao gồm tất cả các giải pháp có thể xảy ra cho vấn đề đã cho. Kỹ thuật phổ biến nhất để khởi tạo là sử dụng các chuỗi nhị phân ngẫu nhiên.

#### 2.2. Fitness Function

Fitness Function giúp thiết lập thể lực của tất cả các cá nhân trong quần thể. Nó chỉ định điểm số thể lực cho mỗi cá nhân, điều này xác định thêm xác suất được chọn để sinh sản. Điểm fitness càng cao thì cơ hội được chọn để sinh sản càng cao.

#### 2.3. Lựa chọn (Selection)

Trong giai đoạn này, các cá thể được chọn để sinh sản con cái. Các cá thể được chọn sau đó được sắp xếp theo cặp hai con để tăng cường khả năng sinh sản. Những cá nhân này truyền gen của họ cho thế hệ tiếp theo.

Mục tiêu chính của giai đoạn này là thiết lập khu vực có nhiều cơ hội tạo ra giải pháp tốt nhất cho vấn đề (tốt hơn thế hệ trước). Thuật toán di truyền sử dụng kỹ thuật lựa chọn tỷ lệ phù hợp để đảm bảo rằng các giải pháp hữu ích được sử dụng để tái tổ hợp.

#### 2.4. Sinh sản (Reproduction)

Giai đoạn này liên quan đến việc tạo ra một dân số trẻ em. Thuật toán sử dụng các toán tử biến thể được áp dụng cho tập hợp gốc. Hai thao tác chính trong giai đoạn này bao gồm lai ghép và đột biến.

- **Crossover:** Toán tử này hoán đổi thông tin di truyền của hai bố mẹ để sinh sản con cái. Nó được thực hiện trên các cặp bố mẹ được chọn ngẫu nhiên để tạo ra quần thể con có kích thước bằng quần thể bố mẹ.
- **Đột biến:** Toán tử này thêm thông tin di truyền mới vào quần thể con mới. Điều này đạt được bằng cách lật một số bit trong nhiễm sắc thể. Đột biến giải quyết vấn đề tối thiểu cục bộ và tăng cường đa dạng hóa. Hình ảnh sau đây cho thấy cách đột biến được thực hiện.

#### 2.5. Thay thế (Replacement)

Thay thế thế hệ diễn ra trong giai đoạn này, là sự thay thế dân số già bằng dân số trẻ mới. Dân số mới bao gồm các điểm thể dục cao hơn so với dân số cũ, đây là dấu hiệu cho thấy một giải pháp cải tiến đã được tạo ra.

#### 2.6. Chấm dứt (Termination)

Sau khi thay thế đã được thực hiện, tiêu chí dừng được sử dụng để cung cấp cơ sở cho việc chấm dứt. Thuật toán sẽ kết thúc sau khi đạt được giải pháp phù hợp với ngưỡng. Nó sẽ xác định giải pháp này là giải pháp tốt nhất trong dân số.

## PHẦN 4. ỨNG DỤNG GIẢI THUẬT DI TRUYỀN VÀO HỆ THỐNG KHUYẾN NGHỊ

### 1. Các nghiên cứu trước đây

Trong RS, GA đã được sử dụng ở ba khía cạnh: phân cụm (KIM & AHN, 2008; Zhang & you Chang, 2006), mô hình lai (Gao & Li, 2008; Ho et al., 2007) và sử dụng GA mà không yêu cầu thông tin bổ sung do mô hình lai cung cấp (Alhijawi & Kilani, 2016; Bobadilla, Ortega, Hernando, & Alcalá, 2011). Thông thường, GA được sử dụng để tìm độ đo tương tự tối ưu. Tiêu, La, Chen và Li (2015) đã đề xuất một hệ thống CF dựa trên vật phẩm có tên là ItemCFGa. Hệ thống này liên quan đến một chức năng tương tự mới sử dụng xếp hạng trung bình cho từng người dùng thay vì xếp hạng trung bình tổng thể cho tất cả người dùng. Bobadilla et al. (2011) đã trình bày một RS mới mà đã sử dụng GA để tìm hàm tương tự tối ưu. Hàm tương tự được chọn bởi GA cung cấp chất lượng tốt hơn và kết quả nhanh hơn so với các số liệu truyền thống khác. Một RS dựa trên GA khác có tên là SimGen được phát triển bởi Alhijawi và Kilani (2016).

SimGen tính toán ma trận tương tự giữa người dùng và người dùng bằng GA. Đồng thời, nó không sử dụng bất kỳ điểm tương đồng phổ biến nào. Các biện pháp như tương quan Pearson và độ tương tự dựa trên vector cosine. Mỗi cá nhân trong dân số đại diện cho một người dùng-người dùng ma trận tương tự chứa các giá trị tương tự trong phạm vi  $[0-1]$ .

Gupta, Shivhare và Sharma (2015) đã đề xuất một kỹ thuật phân cụm c-means mờ và sử dụng độ tương tự có trọng số GA của Bobadilla et al. (2011). Trong hệ thống của họ, các giá trị cụm được tạo của fuzzy c-means được chuyển đến GA để tìm sự tương đồng

giữa các giá trị được phân cụm và để có được các biện pháp tương tự tối ưu. KIM và AHN (2008) đã đề xuất RS sử dụng GA để chọn hạt giống ban đầu tối ưu hoặc phụ tối ưu cho kỹ thuật phân cụm K-neighbor. Bên cạnh đó, Wang, Yu, Feng, Wang, 2014 đã sử dụng một phương pháp phân cụm K-means cải tiến kết hợp với GA để xây dựng hệ thống khuyến nghị phim dựa trên mô hình kết hợp. GA tối ưu hóa vector trọng số của các tính năng để tính toán mức độ tương tự giữa những người dùng.

Salehi (2014) đã đề xuất một phương pháp đề xuất dựa trên tính năng tiềm ẩn giúp cải thiện chất lượng đề xuất và giải quyết vấn đề thừa thớt. Họ đã sử dụng GA như một nhiệm



vụ học tập có giám sát để tối ưu hóa các tính năng tiềm ẩn' trọng số cho mỗi người học dựa trên dữ liệu đánh giá lịch sử. Một RS lai mới cho tài liệu học tập được phát triển bởi Gao & Li (2008) bao gồm nhiều RS. Hệ thống của họ sử dụng GA để dự báo hiệu suất của từng RS. Lv, Hu, và Chen (2016a) đã phát triển RS dựa trên vật phẩm phụ thuộc vào các tính năng của vật phẩm với các trọng số tương ứng. Kim, Kim, Lee và Ahn (2010) đã phát triển một nhạc mới RS kết hợp kỹ thuật lọc dựa trên nội dung với các phương pháp GA tương tác. RS được đề xuất có khám phá sở thích của người dùng. Họ đã sử dụng GA để lấy các bản nhạc phù hợp để giới thiệu cho AU. Giả, Đình, Lưu, Zhang và Zhang (2014) đã đề xuất thuật toán đề xuất CF giúp cải thiện độ chính xác của thuật toán truyền thống. Họ đã sử dụng GA để chọn trọng số và giá trị ngưỡng một cách nhanh chóng để tính toán độ đo tương tự. Thuật toán đề xuất sử dụng hai biến pháp để tạo đề xuất: độ tương tự và độ tin cậy.

Hầu hết, GA đã được sử dụng để tìm số liệu tương tự tối ưu hoặc để tìm trọng số tối ưu của các tính năng dẫn đến thước đo tương tự tối ưu (Bobadilla và cộng sự, 2011; Gupta và cộng sự, 2015; Lv và cộng sự, 2016a; Salehi, 2014; Xiao và cộng sự, 2015). Ngoài ra, GA có đã được sử dụng để chọn các hạt giống ban đầu (K) thích hợp nhất cho thuật toán phân cụm K-mean (KIM & AHN, 2008). Bên cạnh đó, các GA đã được áp dụng để tìm ma trận tương đồng giữa người dùng và người dùng tối ưu (Alhijawi & Kilani, 2016) và để dự báo hiệu suất của nhiều RS (Gao & Li, 2008). Những đóng góp gần đây cho lĩnh vực nghiên cứu RS dựa trên di truyền nhằm mục đích tìm ra các mục thích hợp để hình thành danh sách khuyến nghị.

## **2. Sử dụng đa tiêu chí và thông tin ngữ nghĩa trong hệ thống khuyến nghị**

Các RS truyền thống đưa ra đề xuất cho người dùng chỉ dựa trên một tiêu chí hoặc một chức năng tiện ích. Giả định phổ biến của CF là hai người dùng  $u_1$  và  $u_2$  giống nhau nếu cả hai người dùng đưa ra các đánh giá khá giống nhau cho các mục cụ thể. Do đó, RS có thể dự đoán các mục chưa được xếp hạng cho người dùng  $u_1$  dựa trên xếp hạng của người dùng  $u_2$  đối với các mục này và ngược lại. Tuy nhiên, người dùng  $u_1$  có thể đưa ra xếp hạng dựa trên các tiêu chí cụ thể khác với tiêu chí mà người dùng  $u_2$  đã thông qua khi xếp hạng mục này. Ví dụ: người dùng  $u_1$  và người dùng  $u_2$  cả hai đều cho điểm 4 cho cuốn sách B\*. Người dùng  $u_1$  thích B\* vì tác giả Tác giả1 đã viết cuốn sách này trong khi người dùng  $u_2$  đã ra lệnh cho B\* vì cuốn sách này đã kể một câu chuyện hành động.

Trong khi đó, RSs mục tiêu chỉ n cho phép người dùng đưa ra xếp hạng cho n đặc điểm ngữ nghĩa của mục.

Công việc được đề xuất sử dụng GA để tìm danh sách đề xuất phù hợp thay vì tìm các mục để soạn danh sách khuyến nghị.  $BLI_{GA}$  sử dụng GA để chọn danh sách các mặt hàng đáp ứng sở thích của AU. Do đó, mỗi cá nhân trong dân số đại diện cho một danh sách đề xuất ứng cử viên cho AU. GA chọn một cá thể để giới thiệu gen của nó cho AU tùy thuộc vào trên các cấp độ đa lọc phân cấp. Theo hiểu biết tốt nhất của chúng tôi, chúng tôi chưa tìm thấy trong tài liệu bất kỳ công trình nào trong lĩnh vực nghiên cứu RS tập trung vào chất lượng của toàn bộ danh sách đề xuất.

Bên cạnh đó,  $BLI_{GA}$  áp dụng tư tưởng đa tiêu chí ở hai khía cạnh khác nhau:

- $BLI_{GA}$  sử dụng ý tưởng của Shambour và Lu (2011) để tính toán giá trị tương tự ngữ nghĩa dựa trên mục của một mục mục tiêu nhất định chống lại các mặt hàng khác. Sự giống nhau về ngữ nghĩa đã được tính toán dựa trên các đặc điểm ngữ nghĩa phổ biến.
- $BLI_{GA}$  tạo đề xuất dựa trên nhiều tiêu chí lọc, mỗi tiêu chí liên quan đến một tiêu chí khác nhau. Như vậy, mỗi cá nhân có các giá trị đa thể dục đã được coi là xếp hạng do GA đưa ra để đo lường mức độ phù hợp của cá nhân. Tất cả các tài liệu được đánh giá hoạt động trong các phương pháp dựa trên di truyền sử dụng một kỹ thuật lọc (tức là một chức năng tập thể dục) để chọn ra phương pháp tốt nhất cá nhân.  $BLI_{GA}$  lọc các cá nhân theo thứ bậc để chọn ra người tốt nhất dựa trên nhiều chức năng thể dục.

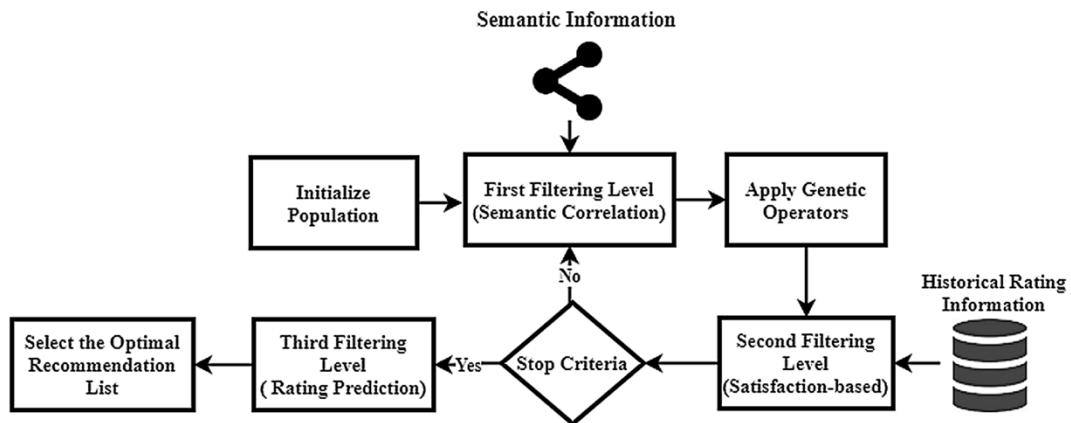
### **3. Hệ thống khuyến nghị dựa trên di truyền được đề xuất**

Alhijawi (2020) đã đề xuất  $BLI_{GA}$  để giải quyết hai vấn đề sparsity và cold start.  $BLI_{GA}$  có 3 hàm mục tiêu, lần lượt là: thông tin ngữ nghĩa của vật phẩm, độ tương đồng về sự ưa thích giữa người dùng AU và các người dùng khác, và cuối cùng là đánh giá dự đoán của các vật phẩm trong danh sách.

Ba lớp lọc trên nhằm đảm bảo  $BLI_{GA}$  có ba đặc điểm như sau:

1. Các vật phẩm thuộc về  $BLI_{GA}$  có độ tương quan cao về đặc điểm ngữ nghĩa.
2. Các hàng xóm của AU hài lòng với các vật phẩm thuộc về  $BLI_{GA}$ .

3. Các vật phẩm thuộc về AU có đánh giá dự đoán cao.



Hình 7: Mô hình xây dựng hệ thống khuyến nghị dựa trên di truyền

### Quy trình thực hiện hệ thống khuyến nghị dựa trên giải thuật di truyền:

**1. Khởi tạo dân số (Dòng 1).** Ban đầu, BLI<sub>GA</sub> lấp đầy dân số bằng M danh sách đề xuất được tạo ngẫu nhiên, iniPop. Mỗi cá nhân chứa các mục ngẫu nhiên chưa được AU xếp hạng trước đó.

**2. Mức lọc đầu tiên (Dòng 3-4).** Đầu tiên, BLI<sub>GA</sub> đánh giá các cá thể dựa trên tiêu chí tương quan ngữ nghĩa để chọn ra top những cá thể tốt nhất (bestMem) ở bước này.

**3. Áp dụng toán tử di truyền (Dòng 5-7).** Từ bestMem, một nhóm cha mẹ được chọn để sinh con bằng cách áp dụng các toán tử lai ghép và đột biến. Mỗi đứa trẻ được tạo ra sẽ được đánh giá theo mối tương quan ngữ nghĩa. Tất cả những đứa trẻ có giá trị tương quan ngữ nghĩa lớn hơn cá thể xấu nhất thuộc về bestMem đều được thêm vào dân số hiện tại.

**4. Mức lọc thứ hai (Dòng 8-9).** Tiếp theo, GA đánh giá từng cá nhân dựa trên tiêu chí hài lòng của người dùng. Các cá nhân tốt nhất hàng đầu (bestMem) được nắm bắt tùy thuộc vào mức độ lọc dựa trên sự hài lòng.

**5. Kiểm tra tiêu chí dừng (Dòng 2).** BestMem thế hệ hiện tại được chuyển sang thế hệ tiếp theo nextGen. BLI<sub>GA</sub> lặp lại các bước này cho một số thế hệ được xác định trước, maxGen.

**6. Cấp lọc thứ ba (Dòng 11).** Ở thế hệ trước, GA đánh giá các cá nhân dựa trên tiêu chí lọc xếp hạng được dự đoán. Mức lọc xếp hạng dự đoán là mức lọc chiếm ưu thế. Do đó, cá nhân tốt nhất được chọn làm  $BLI_{GA}$  dựa trên tiêu chí lọc xếp hạng dự đoán.

**7. Chọn danh sách khuyến nghị tối ưu (Dòng 12).**  $BLI_{GA}$  nắm bắt các mục của cá nhân được chọn và giới thiệu nó cho AU. Lưu ý rằng quy mô dân số chủ yếu giảm.

---

	<b>Input:</b> ma trận user-item $U_i$ và ma trận item-category $I$
	<b>Output:</b> danh sách khuyến nghị
1	Tạo quần thể ban đầu (iniPop) chứa $M$ cá thể.
2	for currentGen , maxGen, thực hiện
3	Tính tương quan ngữ nghĩa ( $S_{emCorrRate(ind)}$ ) của mỗi cá nhân trong iniPop.
4	Chọn ra những cá nhân xuất sắc nhất (topX) tốt nhất (bestMem.
5	Chọn bố mẹ
6	Áp dụng toán tử chéo với xác suất CrossOverP.
7	Áp dụng toán tử đột biến với xác suất đột biến
8	Tính toán mức độ tương đồng dựa trên sự hài lòng ( $S_{imilarity(ind)}$ ).
9	Chọn ra những cá nhân xuất sắc nhất (topX) tốt nhất (bestMem).
10	end
11	Đánh giá từng cá nhân trong bestMem dựa trên xếp hạng dự đoán ( $predictS_{atD(ind, x)}$ ).
12	Chọn cá nhân tốt nhất (recommendMem) và đề xuất các mặt hàng của nó cho AU.

---

Hình 8: Quy trình thực hiện hệ thống khuyến nghị dựa trên giải thuật di truyền

Lưu ý rằng quy mô dân số chủ yếu giảm từ thế hệ này sang thế hệ khác. Trong khi đó, quần thể ban đầu chứa  $M$  cá nhân và thế hệ sau có bestMem trong đó  $|bestMem| < M$ .

Mục 3.1 trình bày chi tiết liên quan đến quần thể ban đầu và sự biểu hiện di truyền. Mục 3.2 minh họa việc đánh giá quy trình và các cấp độ; tiêu chí lọc ngữ nghĩa, lọc dựa trên sự hài lòng và tiêu chí lọc xếp hạng dự đoán. Hơn nữa, trong này phần trình bày chi tiết liên quan đến quá trình lựa chọn danh sách khuyến nghị. Mục 3.3 trình bày các toán tử di truyền, chiến lược chấm dứt và thay thế.

### 3.1. Biểu diễn di truyền và khởi tạo quần thể

BLI<sub>GA</sub> sử dụng GA làm công cụ tối ưu hóa để chọn danh sách đề xuất mong muốn cho AU. Như vậy, mỗi cá nhân đại diện cho một vectơ của các mục có thể ( $\overrightarrow{RecL}$ ). Mỗi gen là một mục có thể được chọn ngẫu nhiên từ tập hợp các mục chưa được xếp hạng trước đó bởi AU. BLI<sub>GA</sub> tạm thời xóa mục đã chọn khỏi tập hợp các mục chưa được AU xếp hạng trước đó (tức là trong khi khởi tạo cá nhân hiện tại). Bước này đảm bảo rằng cá nhân không bao gồm các mục trùng lặp. Giá trị gen đề cập đến ID của vật phẩm.

BLI<sub>GA</sub> khởi tạo một quần thể chứa M danh sách đề xuất ứng cử viên khác nhau có thể có cho AU. Chính thức, dân số và cá nhân được xác định như sau:

$$Population = [\overrightarrow{RecL_1} \overrightarrow{RecL_2} : \overrightarrow{RecL_M}] = [i_1^1, i_2^1, i_3^1, \dots, i_N^1 \ i_1^2, i_2^2, i_3^2, \dots, i_N^2 : \\ i_1^M, i_2^M, i_3^M \dots, i_N^M]$$

BLI<sub>GA</sub> đảm bảo rằng dân số chứa danh sách đề xuất duy nhất bằng cách sử dụng cơ chế xác thực phụ thuộc vào tổng hợp gen. Các gen của mỗi cá thể mới được tạo ra được so sánh với các cá thể có tổng kết gen tương tự. Ví dụ: giả sử rằng quần thể chứa bốn cá thể sau và một cá thể mới  $ind_n(I_9, I_3, I_2, I_1)$  là sản xuất. BLI<sub>GA</sub> sẽ so sánh cá thể mới với chỉ  $ind_1$  và  $ind_3$  vì ba cá thể này có tổng kết gen giống nhau là 15. Trong trường hợp cá thể mới chứa ít nhất một cá thể duy nhất thì nó sẽ được bổ sung vào quần thể.

$$ind_1 \ ind_2 \ ind_3 \ ind_4 \ (I_2 \ I_7 \ I_5 \ I_{10} \ I_5 \ I_3 \ I_2 \ I_1 \ I_4 \ I_{10} \ I_7 \ I_8)$$

### 3.2. Cấp độ lọc

Trong BLI<sub>GA</sub>, GA đánh giá các cá nhân theo thứ bậc dựa trên ba tiêu chí lọc để đảm bảo rằng danh sách đề xuất được chọn (BLI) đáp ứng ba đặc điểm chính:

1. Các mục được đề xuất có các đặc điểm ngữ nghĩa chung.
2. Những người hàng xóm của AU hài lòng với một số mặt hàng được đề xuất.
3. Xếp hạng dự đoán của các mục được đề xuất càng cao càng tốt.

Mỗi tiêu chí lọc được liên kết với một chức năng thể dục. Do đó, ba chức năng tập thể dục đã được xem xét trong phân cấp đánh giá của BLI<sub>GA</sub>. GA đưa ra xếp hạng cho từng cá nhân theo các tiêu chí lọc đó để chọn BLI của AU. Lọc đầu tiên tiêu chí phụ thuộc vào tương quan ngữ nghĩa item-item ( $SemSimI(p, q)$ ) trong đó  $p, q \in ind$ . Sự giống nhau về ngữ nghĩa dựa trên mục của một mục nhất định so với các mục khác được tính toán dựa trên các tính năng ngữ nghĩa phổ biến. Cấp độ lọc thứ hai đo lường sự hài hòa mức độ (tức là mức độ tương đồng dựa trên sự hài lòng) giữa AU và những người dùng khác đã xếp hạng cho ít nhất một mặt hàng thuộc về cá nhân  $ind$ . Sự tương đồng dựa trên sự hài lòng phụ thuộc vào ma trận đánh giá mục người dùng. Tiêu chí lọc thứ ba nhằm mục đích tìm ra tiềm năng các mục yêu thích có xếp hạng dự đoán cao sẽ được chọn làm BLI.

#### a. Tiêu chí lọc tương quan ngữ nghĩa

Mức độ lọc này đã được chọn dựa trên giả định rằng mọi người có xu hướng mua các mặt hàng liên quan đến những gì họ đã mua. Do đó, mọi người mua các mặt hàng có các đặc điểm ngữ nghĩa chung. Về cơ bản, BLI nên chứa một tập hợp các mục có ngữ nghĩa cao mức độ tương quan. BLI<sub>GA</sub> áp dụng ý tưởng được đề xuất bởi Shambour và Lu (2011) để tính toán sự giống nhau về ngữ nghĩa dựa trên mục của một mặt hàng nhất định so với các mặt hàng khác. Mỗi mục đại diện bởi một vector các tính năng ngữ nghĩa (Eq. (2)). Mỗi đặc điểm ngữ nghĩa đại diện bởi một vector của các số nguyên ( $\vec{V}_c$ ). Chẳng hạn, đặt  $V_{c1}$  đại diện cho đặc trưng thể loại và  $v_{c1}^i = (1, 10, 15, 20)$  c Tôi 1 thì mặt hàng này thuộc về các danh mục: 1, 10, 15 và 20. Hệ số Jaccard (Eq. (3)) đã được sử dụng để đo lường mối tương quan ngữ nghĩa giữa các mục. Giá trị tương quan ngữ nghĩa nằm trong khoảng  $[0,1]$ . Trong khi đó,  $SemSimI(p, q) = 1$  chỉ ra rằng mục p và mục q thuộc về cùng một đặc điểm ngữ nghĩa. Trong khi,  $SemSimI(p, q) = 0$  chỉ ra rằng mục p và mục q hoàn toàn thuộc về các đặc trưng ngữ nghĩa khác nhau.

$$\vec{V}_i = (V_{c1}, V_{c2}, V_{c3}, \dots, V_{cn}) \quad SemSimI(p, q) = \frac{F_{11}}{F_{10} + F_{01} + F_{11}}$$

Trong đó,

- F11 đề cập đến tổng số các tính năng trong đó cả hai mục, p và q, thuộc về nó.

Công thức

$$F_{11} = \sum_{f=1}^n |V_{cf}^p \wedge V_{cf}^q|$$

- F10 là tổng số thuộc tính trong đó p thuộc và q không thuộc. Công thức,

$$F_{10} = \sum_{f=1}^n |V_{cf}^p - V_{cf}^q|$$

- F01 đề cập đến tổng số đối tượng địa lý trong đó q thuộc về nó và p không thuộc về nó. Công thức,

$$F_{01} = \sum_{f=1}^n |V_{cf}^q - V_{cf}^p|$$

BLIGA tổng hợp sự giống nhau về ngữ nghĩa của mục-mục giữa mỗi hai mục thuộc về cá nhân để tính toán mục đầu tiên giá trị phù hợp (tức là xếp hạng dựa trên ngữ nghĩa (Eq. (7))) của ind. Mức độ tương đồng về ngữ nghĩa giữa các mục càng cao thì mức độ tương đồng càng cao. Đánh giá dựa trên ngữ nghĩa của cá nhân. BLIGA tính toán mối tương quan ngữ nghĩa giữa các mục trong chế độ ngoại tuyến để giảm thời gian cho những tính toán này. Trong quá trình xử lý trực tuyến, giá trị fitness của cá nhân được tính toán.

$$SemCorrRating(ind) = \sum_{p,q \in ind} SemSiml(p, q)$$

#### **b. Tiêu chí lọc dựa trên sự hài lòng**

Mức độ lọc dựa trên sự hài lòng đã được chọn dựa trên giả định rằng những người thích cùng một mặt hàng có khả năng để cảm thấy tương tự đối với các mặt hàng khác. Do đó, BLIGA tìm kiếm một danh sách chứa các mục tương quan về mặt ngữ nghĩa cao, trong đó những người dùng có cùng sở thích như AU hài lòng với một số mục đó.

Sự tương đồng dựa trên sự hài lòng là sự tương đồng dựa trên người dùng đo lường mức độ hài hòa giữa AU và những người dùng khác đã cho xếp hạng trên ít nhất một mục thuộc về cá nhân ind. Đối với mục đích tính toán, mỗi người dùng được biểu diễn dưới dạng một vectơ số nguyên (Eq. (8)) trong đó giá trị của  $v_{ik}^u$  đề cập đến xếp hạng nhất định của người dùng u trên mục ik. BLIGA tính xếp hạng dựa trên sự hài lòng (Eq. (9)) sử dụng kết hợp tuyến tính bao gồm phép đo tương quan Pearson (Phương trình (10)) và Jaccard nhị phân (Phương trình (12)). Pearson phép đo tương quan phụ thuộc vào các mục thường được xếp hạng giữa hai người dùng bất kể số lượng xếp hạng nhất định của mỗi người dùng. Do đó, điều này có thể góp phần tính toán độ tương tự cao không công bằng và không đáng tin cậy giữa hai người dùng trong trường hợp họ có một tỷ lệ nhỏ của các mặt hàng thường được đánh giá. Chẳng hạn, giả sử người dùng ua đã xếp hạng trên tổng số 266 mặt hàng, người dùng ub đã xếp hạng trên một tổng cộng có 32 sản phẩm và cả hai người dùng đều xếp hạng giống nhau cho 32 sản phẩm đó. Sau đó, sẽ không đáng tin cậy khi cho rằng người dùng ua giống 100% với người dùng ub vì tỷ lệ các mặt hàng được xếp hạng chung giữa cả hai người dùng là 12,1%. Do đó,  $Psim(AU, u)$  được điều chỉnh bởi hệ nhị phân Hệ số Jaccard để xem xét tỷ lệ các mục được đánh giá chung giữa những người dùng (tức là  $v_{ik}^{ux} > 0 \wedge v_{ik}^{uz} > 0$ ). Lưu ý rằng BLIGA tính toán cả giá



trị tương quan Pearson và hệ số Jaccard ở chế độ ngoại tuyến và chỉ tính toán giá trị thể chất của cá nhân trong quá trình xử lý trực tuyến.

$$\vec{V}_u = (v_{i1}^u, v_{i2}^u, \dots, v_{ig}^u)$$

$$SatRating(ind) = \sum_{u \in U} SatSimU(AU, u)$$

$$SatSimU(AU, u) = Psim(AU, u) * Jaccard(AU, u)$$

$$Psim(AU, u) = \frac{\sum_{i \in I} (r_i^{AU} - r_{AU}^-)(r_i^u - r_u^-)}{\sqrt{\sum_{i \in I} (r_i^{AU} - r_{AU}^-)^2} \sqrt{\sum_{i \in I} (r_i^u - r_u^-)^2}}$$

$$Jaccard(AU, u) = \frac{|\vec{V}_{AU} \cap \vec{V}_u|}{|\vec{V}_{AU} \cup \vec{V}_u|}$$

Trong đó,

I là nhóm các mục được đánh giá bởi cả người dùng AU và u.

- $r_{AU,i}$  là tỷ lệ người dùng AU trên vật phẩm i.
- $r_{AU}^-$  là giá trị xếp hạng trung bình của người dùng AU.
- $r_{u,i}$  là đánh giá của người dùng u về mục i.
- $r_u^-$  là giá trị xếp hạng trung bình của người dùng u.
- U đề cập đến tập hợp người dùng đã xếp hạng cho ít nhất một mục thuộc về ind.

### c. Lựa chọn danh sách các mặt hàng tốt nhất

Ở thể hệ trước, quần thể chứa một số cá thể ưu tú nhất. Mỗi cá nhân bao gồm các mục thỏa mãn hai điều kiện:

1. Tương quan về mặt ngữ nghĩa.
2. Những người hàng xóm của AU đã đưa ra xếp hạng về một số mặt hàng đó.

Tuy nhiên, cả hai điều kiện đều không đảm bảo rằng AU có thể thích những mặt hàng đó hơn. Do đó, BLI<sub>GA</sub> dự đoán mức độ hài lòng của AU với từng danh sách giới thiệu ứng viên. Tổng trọng số đã điều chỉnh của Resnick (Eq. (14)) đã được sử dụng để dự đoán

xếp hạng của các mặt hàng. BLIGA sử dụng xếp hạng dự đoán để tính toán xếp hạng dựa trên dự đoán, thứ ba giá trị fitness, (Eq. (13)) của cá nhân ind.

$$\begin{aligned} predictSatRatng(ind, AU) &= \sum_{i \in ind} P_{AU,i} P_{AU,i} \\ &= r_{AU}^- + \frac{\sum_{u \in U^+} (r_i^u - r_u^-) * Psim(AU, u)}{\sum_{u \in U^+} Psim(AU, u)}, \end{aligned}$$

BLIGA coi những người dùng Top-k tương tự như AU (tức là  $U^+$ ), người đã xếp hạng cho ít nhất một mục thuộc về cá nhân trong quá trình dự đoán. Do đó, đối với mỗi danh sách đề xuất ứng cử viên, một nhóm hàng xóm khác nhau được sử dụng trong bước dự đoán. BLI đề cập đến cá nhân có xếp hạng dựa trên dự đoán cao nhất. BLIGA đối mặt với thách thức khi nhiều cá nhân có cùng xếp hạng dựa trên dự đoán cao nhất. Để xử lý một thách thức như vậy, BLIGA trả về cá nhân trong đó xếp hạng dựa trên sự hài lòng là cao nhất. Trong trường hợp nhiều cá nhân có cùng mức đánh giá dựa trên mức độ hài lòng cao nhất, BLIGA sẽ trả về cá nhân có mức đánh giá cao nhất. Xếp hạng dựa trên ngữ nghĩa cao nhất. Trong trường hợp nhiều cá nhân có cùng mức đánh giá dựa trên ngữ nghĩa cao nhất, BLIGA sẽ chọn ngẫu nhiên một cá nhân và trả lại nó.

### 3.3. Các toán tử di truyền và tiêu chí kết thúc

BLIGA áp dụng các toán tử di truyền phổ biến: chọn lọc, lai ghép và đột biến. GA kết thúc khi đạt đến mức tối đa số thế hệ (maxGen). Các tính năng của toán tử di truyền được sử dụng là:

- **Lựa chọn (Selection):** Phương pháp lựa chọn bánh xe roulette (RW) đã được sử dụng để chọn cha mẹ để giao phối (ParentPool). RW cho cao xác suất lựa chọn cho các cá nhân khỏe mạnh và xác suất lựa chọn thấp cho các cá nhân xấu. Tuy nhiên, cả tốt và xấu các cá nhân có cơ hội được chọn nhưng những cá nhân tốt có xác suất cao hơn.
- **Crossover:** Kỹ thuật lai ghép một điểm được áp dụng để sinh ra hai con mới sau mỗi vòng lai ghép. Các xác suất chéo (CrossOverP) được điều chỉnh bằng thực nghiệm. Trong khi đó, một số ngẫu nhiên giữa  $[1, (0,8 \times |ParentPool|)]$  là được chọn để xác định số lượng thao tác chuyển đổi trên mỗi thế hệ. Kỹ thuật lai ghép được cho ăn với hai cá thể trong cái mà; (1) cá nhân đầu tiên có xác suất CrossOverP

được chọn từ ParentPool, (2) cá nhân thứ hai là ngẫu nhiên được chọn từ ParentPool. BLI<sub>GA</sub> sử dụng Thuật toán 2 để đảm bảo rằng cá nhân không bao gồm các mục trùng lặp. Cá nhân với các mục trùng lặp đã được loại trừ.

- **Đột biến:** BLI<sub>GA</sub> sử dụng kỹ thuật đột biến Thống nhất. Loại đột biến này chỉ có thể được sử dụng cho các gen nguyên và float. Đơn giản, đứa trẻ được tạo ra bằng cách thay thế giá trị của gen được chọn từ bố mẹ bằng một giá trị ngẫu nhiên thống nhất được chọn giữa giới hạn trên và giới hạn dưới của gen đó. Giá trị của gen mới được chọn ngẫu nhiên từ một nhóm các vật phẩm chứa các mục mà AU không nhìn thấy trước đó và không được bao gồm trong cá thể mẹ. Bước này đảm bảo rằng các cá nhân không bao gồm các mục trùng lặp. Xác suất đột biến (mutP) được điều chỉnh bằng thực nghiệm. Mỗi cá nhân trong ParentPool có xác suất mutP được chọn làm đầu vào cho hoạt động đột biến.

Các cá thể tốt nhất của bố mẹ và con cái (bestMem) ở thế hệ hiện tại được chọn lọc để chuẩn bị cho quần thể thế hệ tiếp theo. Do áp dụng phương pháp lọc thứ bậc nên số lượng cá thể trong quần thể không cố định từ thế hệ này sang thế hệ khác.

## PHẦN 5. CÁC KẾT QUẢ THỰC NGHIỆM

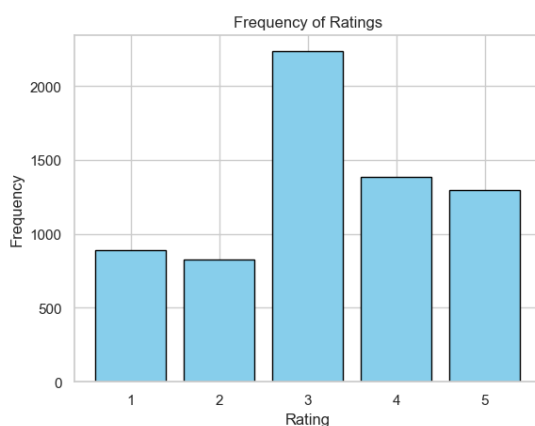
### 1. Bộ dữ liệu dùng cho thí nghiệm

Bộ dữ liệu đã được dùng cho các thí nghiệm ở dưới. Bảng 1 cho thấy các thông tin về các bộ dữ liệu.

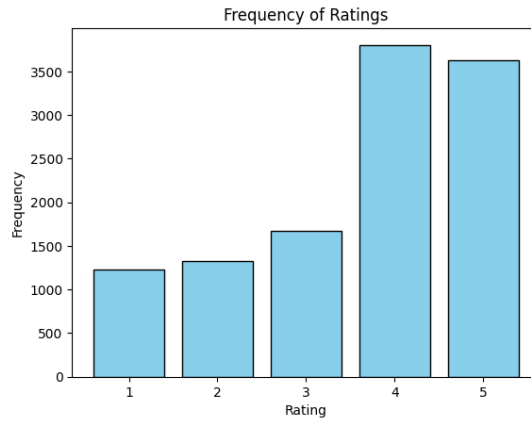
	Bộ dữ liệu Google	Bộ dữ liệu Kaggle	Bộ dữ liệu UEH
Số người dùng	100	250	49
Số sách	200	145	10
Số thể loại	348	145	10
Số đánh giá có	6636	11665	282
Thang đánh giá	1-5	1-5	1-5
Số đánh giá thiếu	13364	24585	208
Độ thừa thớt của dữ liệu	66.82%	67.82%	42.45%

*Bảng 1. Thông tin về các bộ dữ liệu sử dụng*

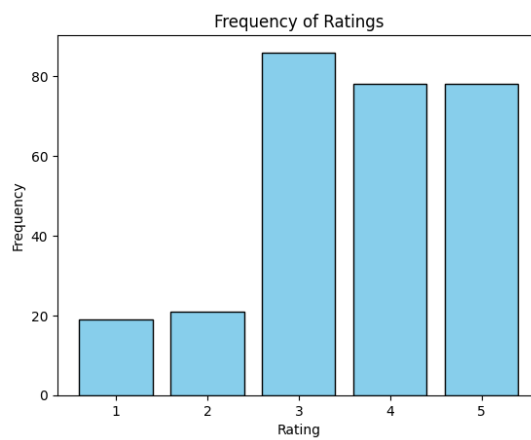
Các hình 9, 10 và 11 cho thấy phân phối tần suất của các nhãn đánh giá. Có thể thấy, các bộ dữ liệu đều có phân bố không đều. Điều này có thể sẽ ảnh hưởng giá trị của hàm mục tiêu khi thực hiện cho từng bộ giá trị.



*Hình 9. Biểu đồ phân phối các giá trị của Bộ dữ liệu 1*



Hình 10. Biểu đồ phân phối các giá trị của Bộ dữ liệu 1



Hình 11. Biểu đồ phân phối các giá trị của Bộ dữ liệu 3

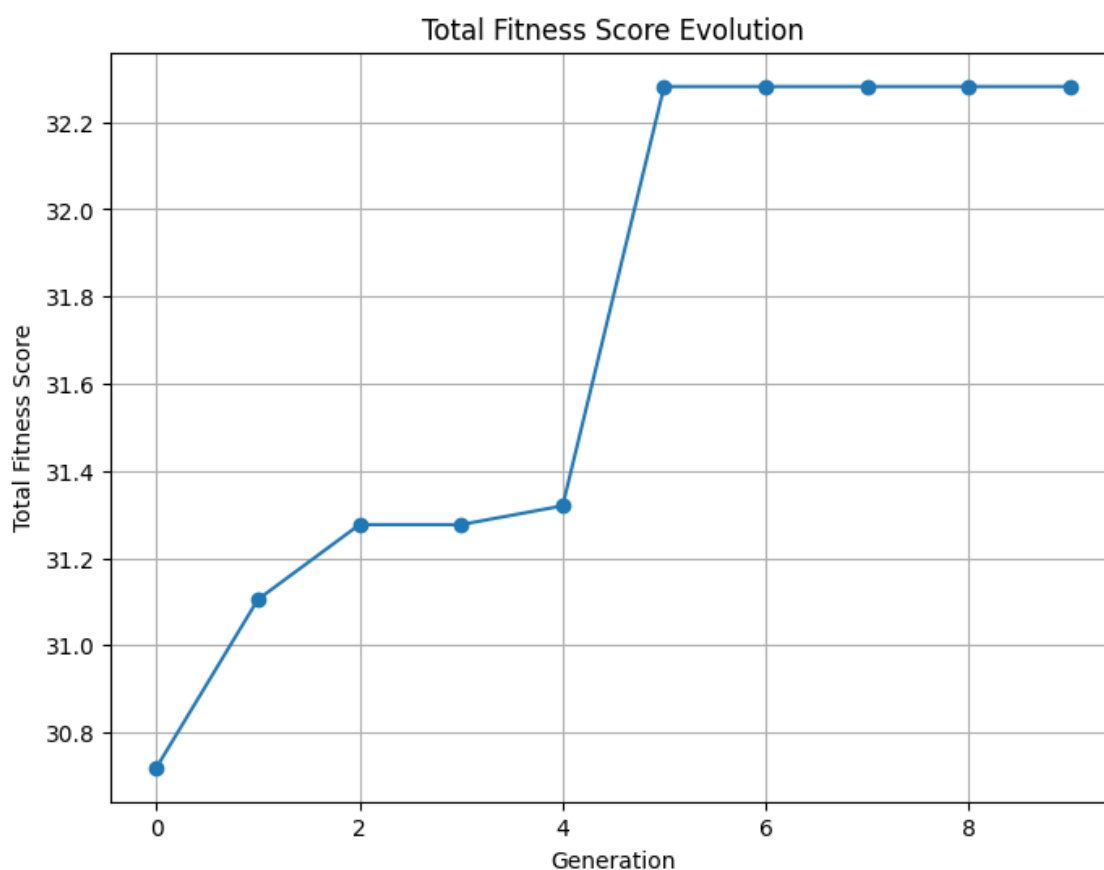
### 3. Kết quả hệ thống cho giá trị ngẫu nhiên

- **Bộ siêu tham số được sử dụng:** Bộ siêu tham số dùng để thực hiện khuyến nghị cho người dùng AU ngẫu nhiên cho bộ dữ liệu Google: `pop_size = 50`, `mutation_rate = 0.2`, `num_generations = 10`, `crossover_func = "one_point"`, `mutation_func = "swap"`, `no_rec = 10`.

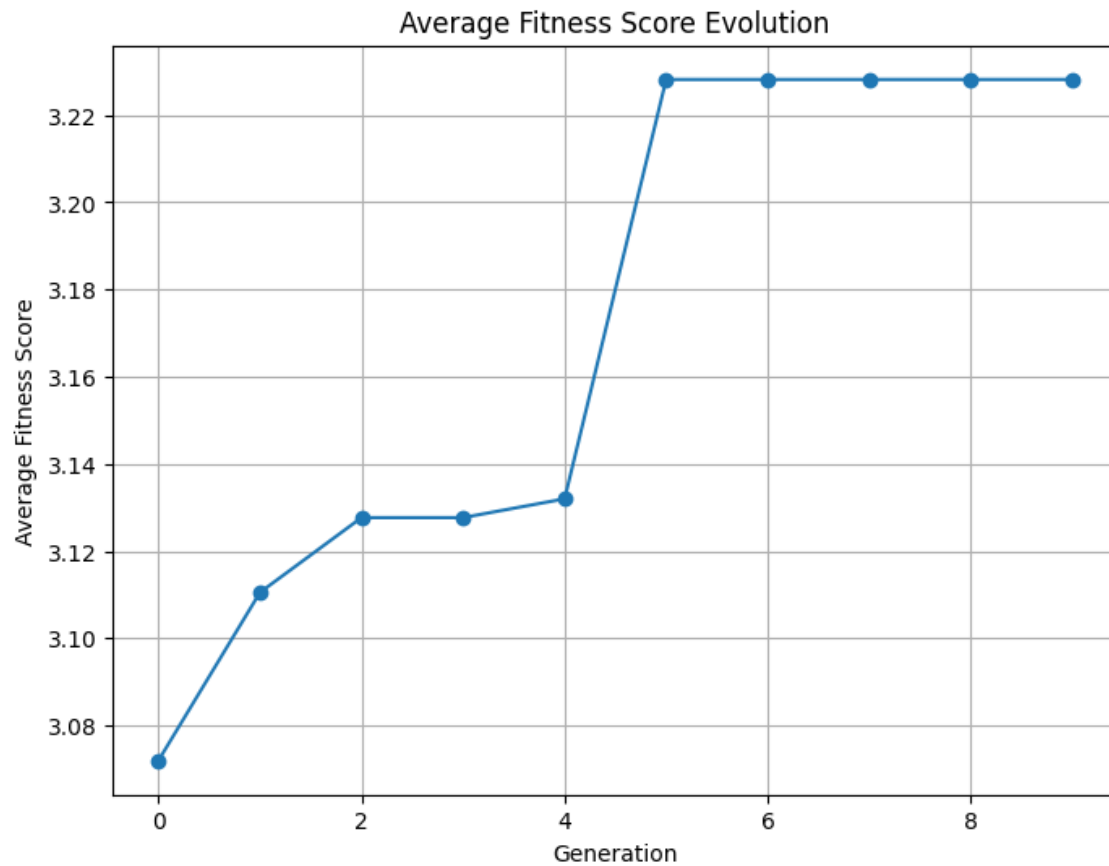
Trong đó:

- `pop_size`: Số lượng cá thể ban đầu của quần thể
- `mutation_rate`: Tỷ lệ đột biến
- `num_generations`: Số lượng thế hệ
- `crossover_func`: Phép lai tạo sử dụng
- `mutation_func`: Phép đột biến sử dụng
- `no_rec`: Số lượng vật phẩm khuyến nghị

- **Kết quả dự đoán so với thể hệ:** Hình 12 và hình 13 lần lượt cho thấy tổng giá trị dự đoán và giá trị dự đoán trung bình cho danh sách khuyến nghị tốt nhất qua từng thể hệ. Có thể thấy, giá trị dự đoán được cải thiện qua các thể hệ cho người dùng AU thử nghiệm. Trong đó, thuật toán tìm được hàm mục tiêu tối ưu nhất từ thể hệ thứ 5.



Hình 12. Giá trị hàm mục tiêu qua từng thể hệ với AU ngẫu nhiên



Hình 13. Giá trị trung bình của dự đoán qua từng thế hệ với AU ngẫu nhiên

#### 4. Các tiêu chí đánh giá

Để đo độ chính xác của các dự đoán, một số tiêu chí đánh giá đã được sử dụng. Đối với mỗi người dùng trong tập dữ liệu, nhóm coi người dùng đó là người dùng chủ động (AU) mới và tính toán các dự đoán cho từng mục. Sau đó, nhóm so sánh các dự đoán với các giá trị thực tế cho từng mục đã được đánh giá bởi người dùng. Cuối cùng, nhóm tính MAE bằng cách lấy trung bình chênh lệch tuyệt đối giữa các dự đoán và giá trị thực tế trên tất cả các mục đã được đánh giá. Với phương pháp này, độ hiệu quả của thuật toán không bị ảnh hưởng bởi sự ngẫu nhiên của giá trị của người dùng AU và các phép toán được thực hiện chỉ trên các giá trị của từng tập dữ liệu.

##### 4.1. Mean Absolute Error (MAE):

MAE đo lường trung bình chênh lệch tuyệt đối giữa các dự đoán và giá trị thực tế cho tất cả các mục được đánh giá bởi một người dùng.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\text{Actual Value}_i - \text{Predicted Value}_i|$$

#### 4.2. Precision:

Precision đo lường tỷ lệ các mục phù hợp được dự đoán chính xác (true positives) trên tổng số các mục phù hợp được dự đoán.

$$Precision = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

#### 4.3. Recall:

Recall đo lường tỷ lệ các mục phù hợp được dự đoán chính xác (true positives) trên tổng số các mục phù hợp thực tế.

$$Recall = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

#### 4.4. F1-measure:

F1-measure là trung bình điều hòa của Precision và Recall và cung cấp một chỉ số duy nhất cân bằng giữa Precision và Recall.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Trong đó:

- **True Positives (TP)** đại diện cho vật phẩm có đánh giá được dự đoán lớn hơn ngưỡng và giá trị thực sự lớn hơn ngưỡng,
- **False Positives (FP)** đại diện cho vật phẩm có đánh giá được dự đoán lớn hơn ngưỡng nhưng giá trị thực sự không lớn hơn ngưỡng,
- **False Negatives (FN)** đại diện cho vật phẩm có đánh giá được dự đoán không lớn hơn ngưỡng nhưng giá trị thực sự lớn hơn ngưỡng.

### 5. Kết quả đánh giá

Bảng 2 cho thấy kết quả đánh giá cho các bộ dữ liệu sử dụng. Có thể thấy,  $BLI_{GA}$  đem lại độ dự đoán tương đối tốt cho các bộ dữ liệu. Cụ thể hơn, Precision cao nhất đạt 97.7% cho bộ dữ liệu 3. Tương tự, bộ dữ liệu 1 có Precision cao hơn các chỉ số khác, với 90.9%. Bộ dữ liệu 1 có MAE thấp nhất, với 0.2802.

Tiêu chí đánh giá	Bộ dữ liệu Google	Bộ dữ liệu Kaggle	Bộ dữ liệu UEH
-------------------	-------------------	-------------------	----------------



Ngưỡng	3	3	3
MAE	0.2802	0.2763	0.3180
Precision	0.9090	0.8376	0.9770
Recall	0.8396	0.9189	0.8765
F1-Score	0.8729	0.8764	0.9240

*Bảng 2. Kết quả đánh giá cho các bộ dữ liệu*

## 6. Điều chỉnh siêu tham số

Để đánh giá hiệu năng của BLI<sub>GA</sub>, nhóm thử nghiệm với các tổ hợp siêu tham số trong bảng 2. Có tổng cộng 48 tổ hợp, với tổ hợp {pop\_size = 50, mutation\_rate = 0.1, num\_generations = 15, crossover\_func = “uniform”, mutation\_func = “swap”, no\_rec = 5} đem lại kết quả dự đoán cao nhất. Với máy tính có các thông số sau: 4GB RAM; 128GB bộ nhớ, vi xử lý thế hệ 4 Intel® Core™ i5, tổ hợp siêu tham số mang lại tốc độ xử lý nhanh nhất là {pop\_size = 20, mutation\_rate = 0.1, num\_generations = 5, crossover\_function = “uniform”, mutation\_function = “swap”, no\_rec = 5}. Ngoài ra, trong hầu hết các trường hợp được thử nghiệm, hàm mục tiêu đạt giá trị tối ưu và ổn định từ thế hệ thứ 5.

## PHẦN 6. THIẾT KẾ GIAO DIỆN CHO HỆ THỐNG

### 1. Xây dựng giao diện cho hệ thống

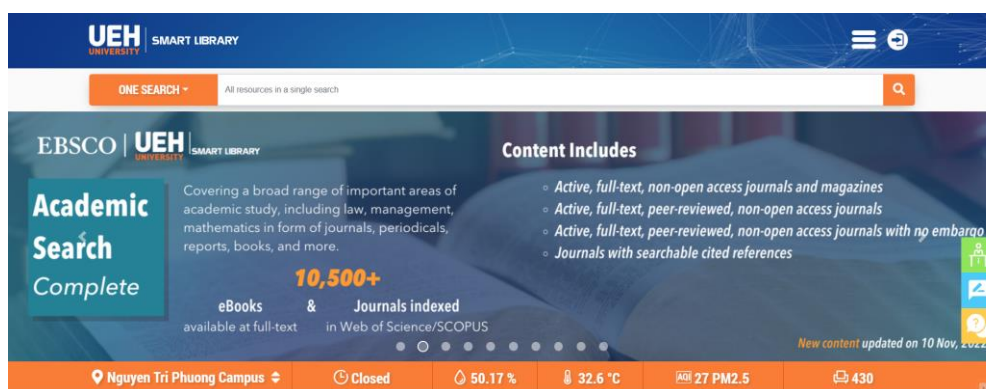
#### 1.1. Qt Designer & PyQt5

Tại chương này, nhóm nghiên cứu tập trung vào việc thiết kế giao diện cho hệ thống bằng cách sử dụng phần mềm Qt Designer phối với thư viện PyQt5 trong môi trường lập trình Python. Với Qt Designer - công cụ thiết kế giao diện người dùng được cung cấp bởi Qt, giúp dễ dàng và nhanh chóng tạo các thành phần giao diện như: nút, hộp văn bản, tiện ích, và kết nối chúng với mã Python. Bên cạnh đó, PyQt5 - thư viện Python liên kết mạnh mẽ với Qt, sẽ hỗ trợ cung cấp các lớp và phương thức cần thiết để xây dựng giao diện tương tác và tích hợp chúng với logic ứng dụng.

Lựa chọn phương pháp thiết kế giao diện bằng Qt Designer kết hợp với PyQt5 là vì sự tiện ích và hiệu quả mà hai công cụ này mang lại. Qt Designer cho phép nhóm nghiên cứu tạo giao diện người dùng một cách trực quan và dễ dàng. Sau khi hoàn thành thiết kế, PyQt5 cung cấp các phương thức liên kết giao diện với mã Python, tạo nên sự tương tác mạnh mẽ giữa giao diện và các chức năng của hệ thống.

#### 1.2. Ý tưởng thiết kế giao diện cho hệ thống

Với mục đích ban đầu của nhóm nghiên cứu là ứng dụng Giải thuật Di truyền trong việc xây dựng Hệ thống Khuyến nghị vào thực tiễn, cụ thể là thư viện điện tử thông minh của trường Đại học Kinh tế TP. HCM - Đại học UEH (UEH Smart Digital Library). Nên giao diện chính của bài nghiên cứu sẽ được lấy ý tưởng từ đó, việc áp dụng các yếu tố giao diện từ UEH Digital Library sẽ mang đến sự thân thiện và sự quen thuộc cho người dùng, đặc biệt là sinh viên, từ đó nâng cao khả năng tiếp nhận và sử dụng của hệ thống.



Hình 14. UEH Smart Digital Library

Như vậy, việc lựa chọn thiết kế giao diện bằng Qt Designer kết hợp với PyQt5, cùng với việc áp dụng ý tưởng từ UEH Smart Digital Library, sẽ tạo nên một giao diện chuyên nghiệp, tiện ích và thân thiện cho hệ thống, cung cấp trải nghiệm người dùng cao và tối ưu hóa khả năng ứng dụng thực tế cho cộng đồng sinh viên.




## 2. Thiết kế giao diện.

### 2.1. Nút chức năng

Thanh tìm kiếm:

ON SEARCH

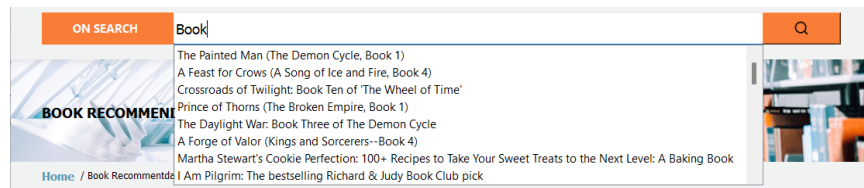
Q

Tên nút	Hình ảnh	Chức năng	Triển khai
<b>Search</b>		Khi nhấn nút, hệ thống sẽ giúp người dùng tìm kiếm danh sách dựa trên từ khóa mà họ đã nhập vào thanh tìm kiếm	Sau khi người dùng đã nhập từ khóa vào thanh tìm kiếm và nhấn nút <b>Search</b> , hệ thống sẽ thực hiện tìm danh sách dựa trên cơ sở dữ liệu và hiển thị kết quả lên giao diện
<b>Submit</b>		Khi nhấn nút, hệ thống sẽ tự động gợi ý 10 quyển sách tương thích nhất với người dùng dựa trên lịch sử đánh giá trước đó	Sau khi người dùng nhập đánh giá cho một hoặc nhiều quyển sách và nhấn <b>Submit</b> để hệ thống khuyến nghị các quyển sách phù hợp
<b>Rating</b>		Người dùng có thể đánh giá từng quyển sách bằng cách nhấn vào các ngôi sao từ 1 tới 5, với các mức điểm tương ứng	Sau khi người dùng nhấn chọn vào quyển sách bất kỳ trên thanh tìm kiếm, hộp chức năng đánh giá sẽ hiện ra. Tại đây, khi người dùng nhấn vào ngôi sao, hệ thống sẽ lưu trữ điểm đánh giá tương ứng cho quyển sách đó.

### 2.2. Chức năng

#### a. Tìm kiếm tên sách.

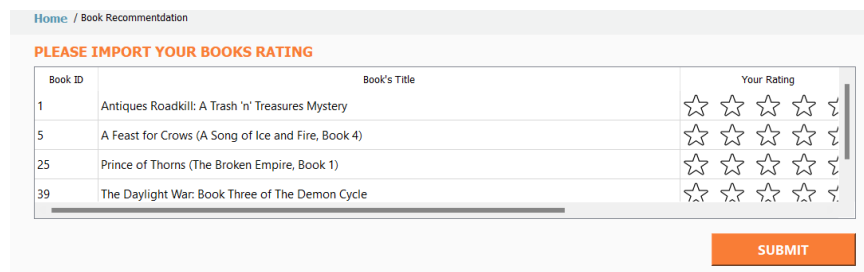
- **Chức năng:** Tại thanh tìm kiếm, người dùng có thể nhập vào tên của một quyển sách để hệ thống tự động gợi ý quyển sách đó cho họ. Hệ thống cũng hỗ trợ việc nhập chỉ một phần tên sách, và từ đó, dựa vào thuật toán tìm kiếm, hệ thống sẽ gợi ý các quyển sách liên quan đến từ khóa.
- **Triển khai:** Khi người dùng nhập từ khóa, hệ thống sẽ truy vấn vào cơ sở dữ liệu và hiển thị các quyển sách phù hợp với từ khóa đó.
- **Mô tả:**



- **Mục đích ứng dụng:** Chức năng này là giúp người dùng dễ dàng và nhanh chóng tìm kiếm sách theo tên hoặc từ khóa liên quan. Việc cung cấp gợi ý sách khi nhập từ khóa giúp tăng khả năng tìm kiếm chính xác và thuận tiện cho người dùng.

#### b. Đánh giá sách.

- **Chức năng:** Người dùng có thể chọn một hoặc nhiều quyển sách bất kỳ để đánh giá điểm rating (mức độ yêu thích) từ 1 đến 5 cho chúng.
- **Triển khai:** Hệ thống sẽ lưu giữ điểm đánh giá của người dùng cho từng quyển sách, đồng thời cập nhật cơ sở dữ liệu để dựa vào những đánh giá này khi đề xuất sách cho người dùng khác.
- **Mô tả:**



- **Mục đích thực hiện:** Chức năng này giúp thu thập phản hồi và đánh giá từ người dùng về các quyển sách có trong thư viện của trường. Điều này giúp cải thiện quá trình đề xuất sách và tăng tính chính xác của hệ thống gợi ý sách cho người dùng.

### c. Hiển thị danh sách khuyến nghị.

- **Chức năng:** Sau khi người dùng đã nhập đánh giá và nhấn nút **Submit**, hệ thống sẽ tự động hiển thị 10 quyển sách phù hợp nhất với người dùng.
- **Triển khai:** Hệ thống sẽ sử dụng thuật toán gợi ý dựa trên điểm đánh giá đã thu thập từ trước đó để đề xuất các quyển sách hấp dẫn cho người dùng.
- **Mô tả:**

PLEASE IMPORT YOUR BOOKS RATING

Book ID	Book's Title	Your Rating
5	A Feast for Crows (A Song of Ice and Fire, Book 4)	★★★★☆
25	Prince of Thorns (The Broken Empire, Book 1)	★★★★★
39	The Daylight War: Book Three of The Demon Cycle	★★★★★
6	God of War: The Official Novelization	★★★★★

**SUBMIT**

	Book ID	Book's Title
1	197	Shadowbred: The Twilight War, Book 1
2	68	A Game of Thrones: The Story Continues Books 1-5: A Game of Thrones, A Clash of Kings, A Storm of Swords, A Feast for Crows, A
3	115	Assassin's Apprentice (The Farseer Trilogy, Book 1)
4	87	Thirty-Five and a Half Conspiracies: Rose Gardner Mystery #8

- **Mục đích thực hiện:** Chức năng này nhằm đáp ứng nhu cầu người dùng trong việc khám phá sách mới có độ tương thích cao với mong muốn của họ. Hiển thị danh sách sách đề nghị giúp tăng tính tương tác và đáp ứng nhanh chóng những yêu cầu tìm kiếm của người dùng.

### d. Thông báo chờ.

- **Chức năng:** Khi người dùng nhấn nút **Submit**, hệ thống sẽ tự động hiển thị một hoạt ảnh thông báo cho người dùng biết chương trình đang được thực thi
- **Mô tả:**

Home / Book Recommendation

BOOK REVIEW SCORES

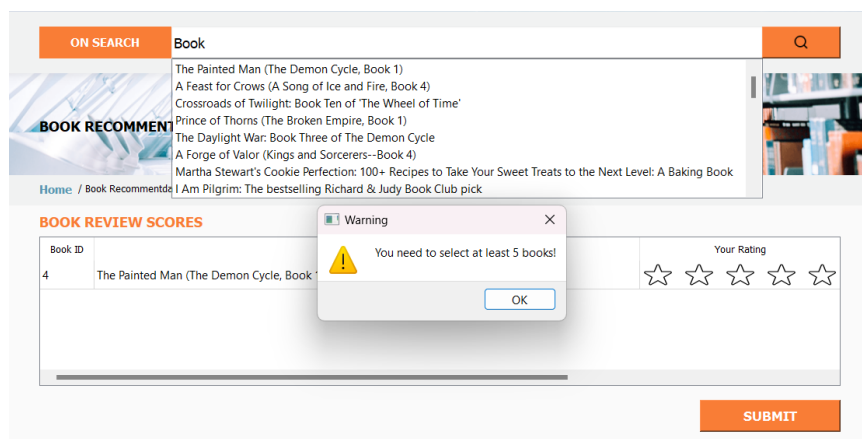
ID	Title	Author	Page Count	Genres	Language	Published Date
1	Antiques Roadkill: A Trash 'n' Treasures Mystery	Barbara Allan	288	Fiction , Mystery &amp; ...	English	7/1/2007
2	The Art of Super Mario Odyssey	Nintendo	368	Games &amp; Activities...	English	11/5/2019
3	Getting Away Is Deadly: An Ellie Avery Mystery	Sara Rosett	320	Fiction , Fantasy , Dark ...	English	3/1/2009
4	The Painted Man (The Demon Cycle, Book 1)	Peter V. Brett	544	Fiction , Media Tie-In	English	1/8/2009

**SUBMIT**

- **Mục đích thực hiện:** Chức năng thông báo chờ là thông báo cho người dùng biết rằng hệ thống đang xử lý thông tin và đưa ra đề xuất sách. Nhằm giúp cải thiện trải nghiệm người dùng.

e. **Thông báo người dùng chưa nhập đánh giá:**

- **Chức năng:** Khi người dùng nhấn nút **Submit** mà chưa nhập đánh giá cho bất kỳ quyển sách nào, hệ thống sẽ tự động thông báo cho người dùng biết rằng họ phải nhập đánh giá cho ít nhất một quyển sách để nhận được đề xuất từ hệ thống: **“You need to select at least 5 books!”**.
- **Mô tả:**



- **Mục đích thực hiện:** Thông báo chưa nhập đánh giá giúp nhắc nhở người dùng tham gia đánh giá sách, tăng tính chính xác và đáng tin cậy của quá trình đề xuất sách trong tương lai. Thông báo này cũng giúp người dùng hiểu rõ hơn về yêu cầu và quy trình thực hiện.

### 3. Kết quả giao diện

- **Giao diện chờ:**

SMART LIBRARY

ON SEARCH

Q

BOOK RECOMMENDATION

[Home](#) / [Book Recommendation](#)

PLEASE IMPORT YOUR BOOKS RATING

ID	Title	Author	Page Count	Genres	Language	Published Date
1	Antiques Roadkill: A Trash 'n' Treasures Mystery	Barbara Allan	288	Fiction , Mystery & amp; , ...	English	7/1/2007
2	The Art of Super Mario Odyssey	Nintendo	368	Games & amp; , Activities...	English	11/5/2019
3	Getting Away Is Deadly: An Ellie Avery Mystery	Sara Rosett	320	Fiction , Fantasy , Dark ...	English	3/1/2009
4	The Painted Man (The Demon Cycle, Book 1)	Peter V. Brett	544	Fiction , Media Tie-In	English	1/8/2009

SUBMIT

Hình 15. Mô phỏng giao diện chờ của hệ thống

- **Giao diện đánh giá:**

SMART LIBRARY

ON SEARCH

Book

Q

BOOK RECOMMENDATION

[Home](#) / [Book Recommendation](#)

PLEASE IMPORT YOUR BOOKS RATING

Book ID	Book's Title	Your Rating
4	The Painted Man (The Demon Cycle, Book 1)	★★★★★
5	A Feast for Crows (A Song of Ice and Fire, Book 4)	★★★★☆
15	Crossroads of Twilight: Book Ten of 'The Wheel of Time'	★★★★★
25	Prince of Thorns (The Broken Empire, Book 1)	★★☆☆☆

SUBMIT

Hình 16. Mô phỏng giao diện đánh giá sách của hệ thống

- **Giao diện kết quả:**

UEH

UNIVERSITY

SMART LIBRARY

ON SEARCH

Q

BOOK RECOMMENDATION

Home / Book Recommendation

PLEASE IMPORT YOUR BOOKS RATING

Book ID	Book's Title	Your Rating
5	A Feast for Crows (A Song of Ice and Fire, Book 4)	<input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
25	Prince of Thorns (The Broken Empire, Book 1)	<input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/>
39	The Daylight War: Book Three of The Demon Cycle	<input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/>
6	God of War: The Official Novelization	<input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/> <input checked="" type="radio"/>

SUBMIT

	Book ID	Book's Title
1	197	Shadowbred: The Twilight War, Book 1
2	68	A Game of Thrones: The Story Continues Books 1-5: A Game of Thrones, A Clash of Kings, A Storm of Swords, A Feast for Crows, A
3	115	Assassinâ€™s Apprentice (The Farseer Trilogy, Book 1)
4	87	Thirty-Five and a Half Conspiracies: Rose Gardner Mystery #8

Hình 17. Mô phỏng giao diện cho kết quả trả về của hệ thống



## PHẦN 7. KẾT LUẬN

### 1. Các kết quả đạt được

Bài nghiên cứu đã ứng dụng mô hình BLI<sub>GA</sub>, một phương pháp lọc tương tác đa lớp dùng giải thuật di truyền. BLI<sub>GA</sub> khởi tạo các danh sách khuyến nghị sách có độ tương quan và chưa được đọc bởi người dùng. Thông qua các phép chọn, lai tạo và đột biến, BLI<sub>GA</sub> cải thiện giá trị của hàm mục tiêu để tìm được danh sách khuyến nghị tốt nhất. BLI<sub>GA</sub> không tìm kiếm các vật phẩm tốt nhất mà tìm kiếm danh sách vật phẩm tốt nhất theo tiêu chí là tổng đánh giá dự đoán và tiêu chí phụ là độ tương quan ngữ nghĩa.

Nhóm tác đã sử dụng BLI<sub>GA</sub> với một số thay đổi để thí nghiệm trên các bộ dữ liệu mới so với bài viết gốc. Ngoài ra, nhóm tác giả đã so sánh hiệu quả của BLI<sub>GA</sub> qua các tổ hợp siêu tham số khác nhau.

Bằng cách đánh giá các chỉ số đo độ chính xác, nhóm tác giả nhận thấy phương pháp dự đoán bằng tương quan Pearson và Jaccard nhị phân đem lại độ chính xác cao trên tất cả các bộ dữ liệu sử dụng. Từ đó, cho thấy sự ổn định trong hiệu quả thể hiện của phương pháp trên.

Nhóm tác giả cũng đã thực hiện các thí nghiệm trên các tổ hợp siêu tham số khác nhau. BLI<sub>GA</sub> cho thấy giá trị của hàm mục tiêu được cải thiện qua từng thế hệ. Trong 48 tổ hợp siêu tham số được xét, 60% của 10 tổ hợp mang lại giá trị của hàm mục tiêu cao nhất có kích cỡ quần thể cao hơn. Tương tự, 60% có số lượng thế hệ cao hơn. Từ đó có thể kết luận, thông qua việc tăng kích cỡ quần thể và số lượng thế hệ hay tăng phạm vi tìm kiếm, giá trị tối ưu do BLI<sub>GA</sub> tìm được có thể được cải thiện.

Từ các thí nghiệm trên BLI<sub>GA</sub>, nhóm đã chọn ra tổ hợp siêu tham số phù hợp và ứng dụng cho dự án UEH Smart Digital Library. Thông qua phân tích giao diện có sẵn của UEH Smart Digital Library, nhóm đã điều chỉnh và thiết kế giao diện trên Windows với hai chức năng chính là đánh giá và khuyến nghị sách.

Để kiểm tra tính ứng dụng của hệ thống khuyến nghị, nhóm đã thu thập ý kiến từ 49 người dùng là các bạn sinh viên tại các trường Đại học ở Việt Nam. Nhóm đã thực hiện các thí nghiệm tương tự so với các bộ dữ liệu khác và cũng đem lại kết quả dự đoán với độ chính

xác cao, cho thấy việc ứng dụng hệ thống khuyến nghị cho UEH Smart Digital Library là hoàn toàn khả thi.

## **2. Những hạn chế và hướng phát triển**

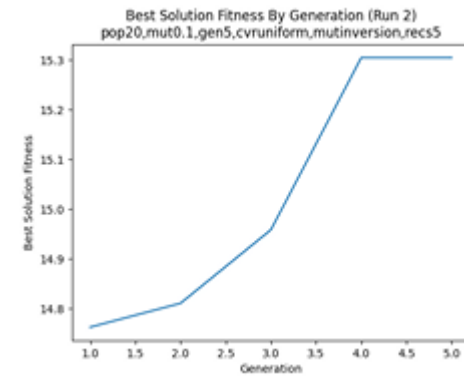
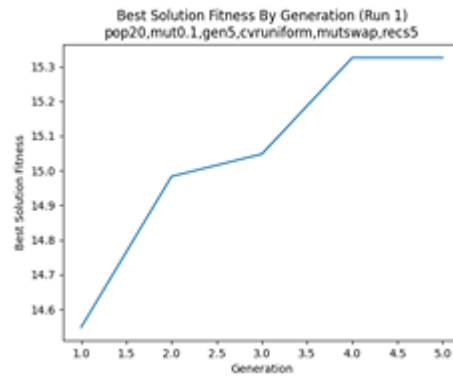
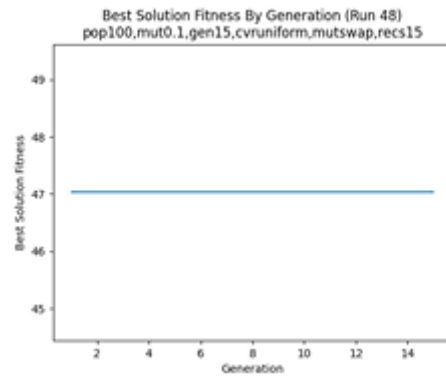
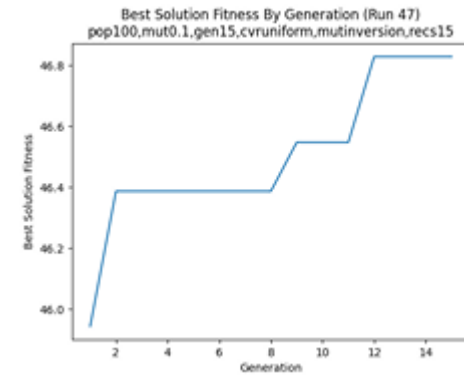
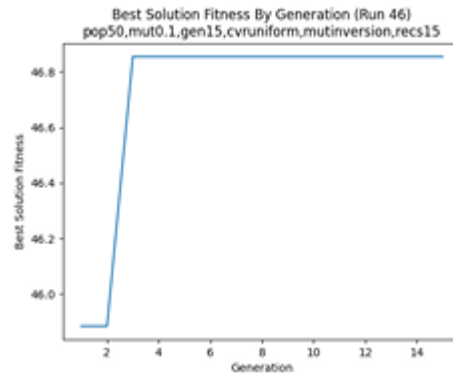
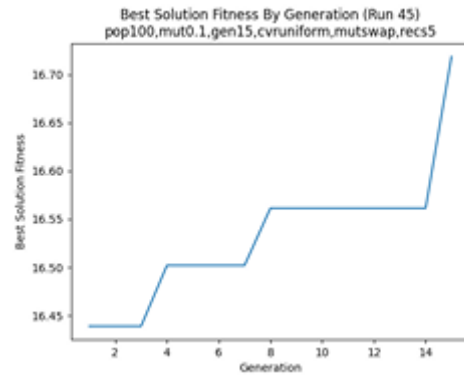
Ngoài các kết quả đạt được, nhóm nghiên cứu cũng đối diện với một số hạn chế quan trọng trong quá trình thực hiện báo cáo. Đầu tiên, nhóm chưa thử nghiệm cho các bộ dữ liệu lớn hơn để so sánh hiệu quả dự đoán.

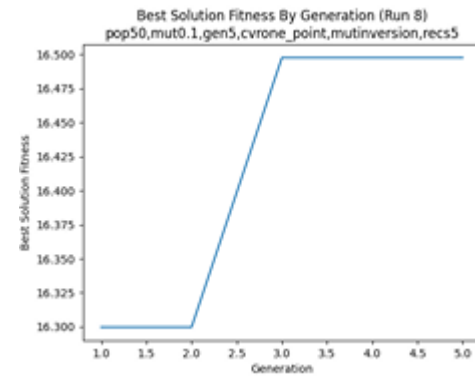
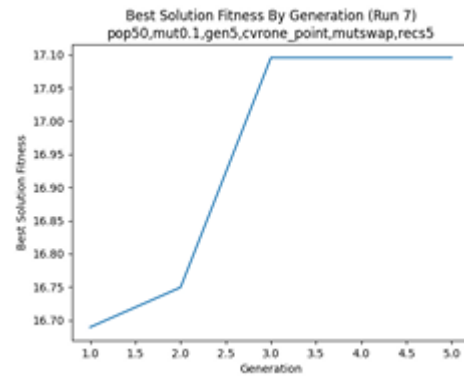
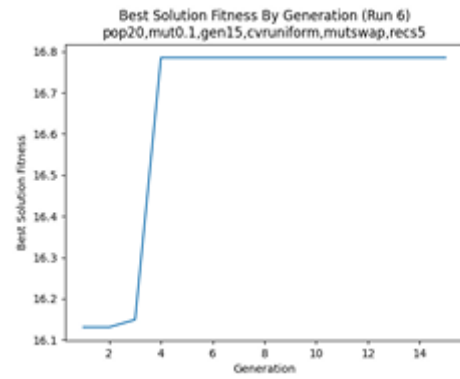
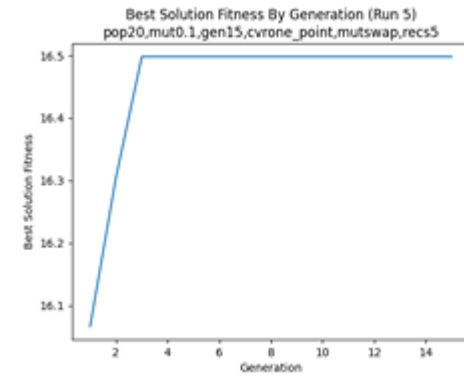
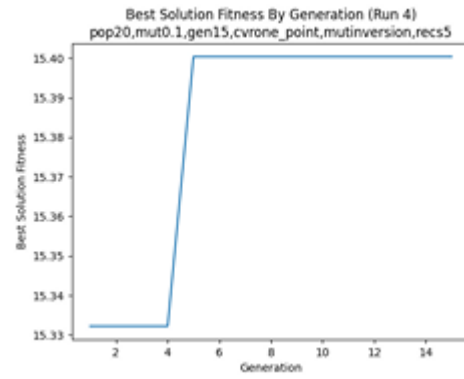
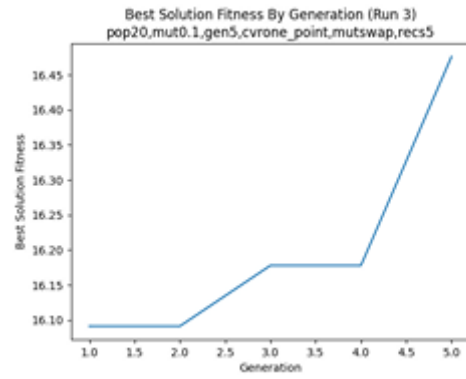
Hiệu quả của giải pháp đề xuất cũng chưa được so sánh với BLI<sub>GA</sub> gốc và các phương pháp trong các bài nghiên cứu khác. Điều này khiến nhóm khó có thể đánh giá được tính hiệu quả và xác định những điểm mạnh và yếu của các thay đổi của nhóm so với BLI<sub>GA</sub> gốc.

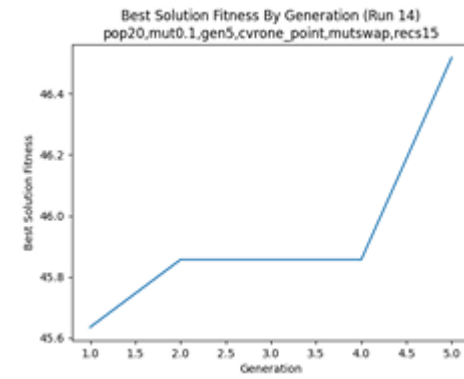
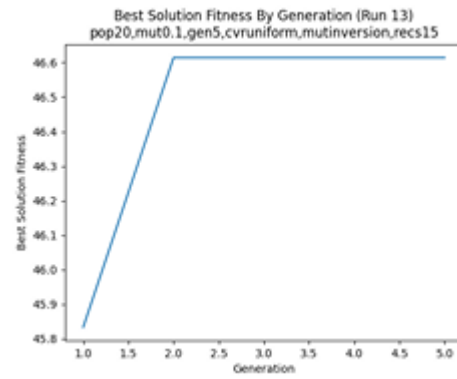
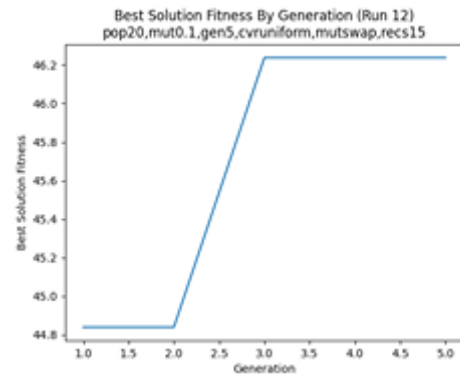
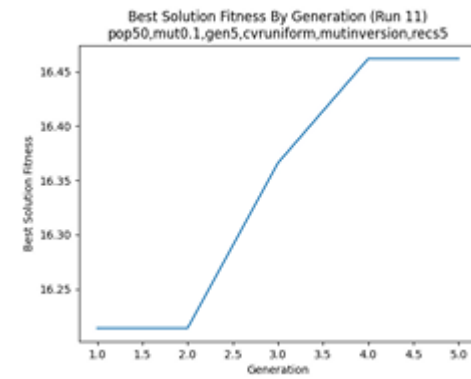
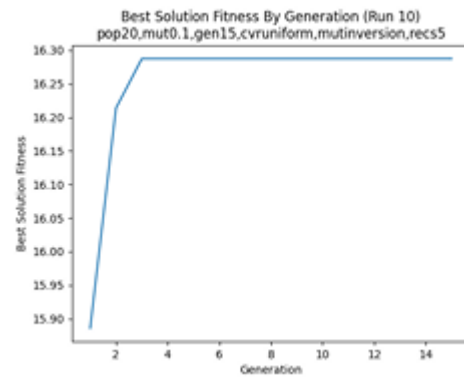
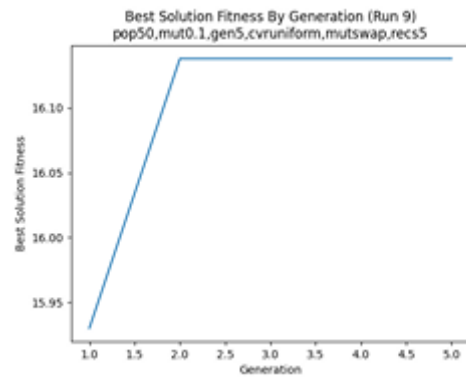
Thứ hai, ứng dụng của giải pháp chưa được thực hiện trên công nghệ giống với UEH Smart Digital Library, điều này có thể tạo ra những khó khăn trong quá trình tích hợp và sử dụng giải pháp mới vào hệ thống hiện tại. Điều này yêu cầu nhóm nghiên cứu phải cân nhắc và tối ưu hóa giao diện để đảm bảo tính tương thích và thuận tiện cho người dùng.

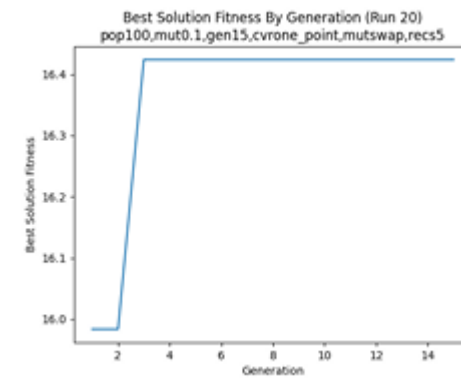
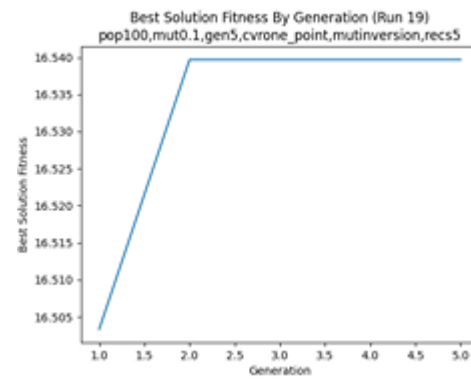
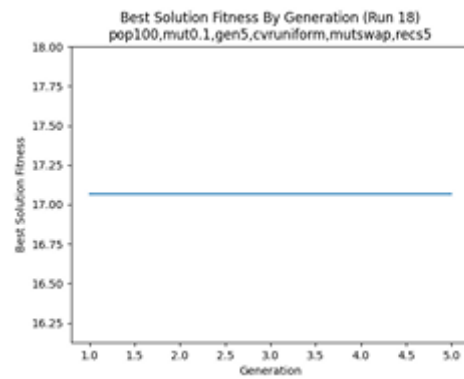
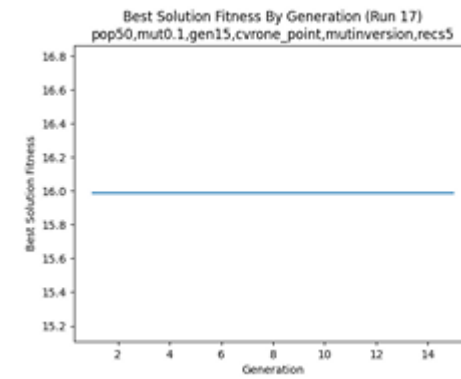
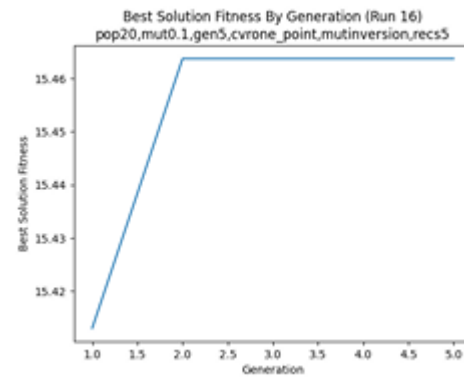
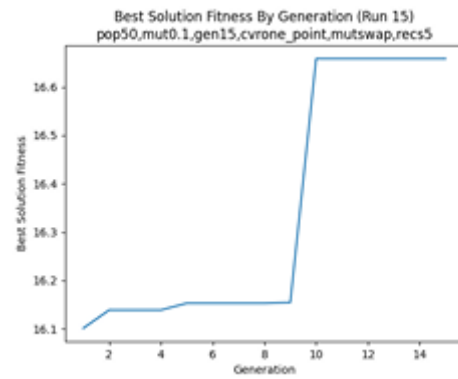
Cuối cùng, bộ dữ liệu sử dụng trong nghiên cứu chưa phải là bộ dữ liệu của UEH Smart Digital Library, do gặp khó khăn trong việc thu thập ý kiến của sinh viên về các đầu sách trong thư viện. Điều này có thể làm giảm tính đại diện và khả năng áp dụng của giải pháp mới trong bối cảnh thực tế của thư viện.

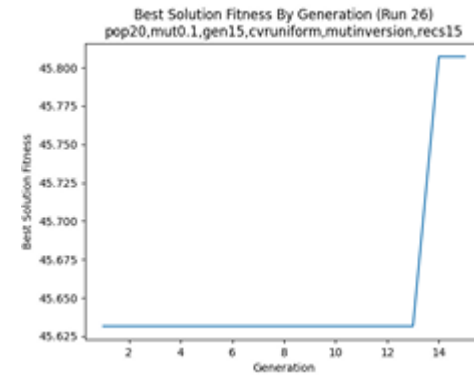
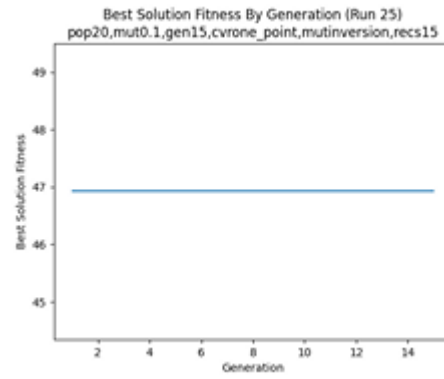
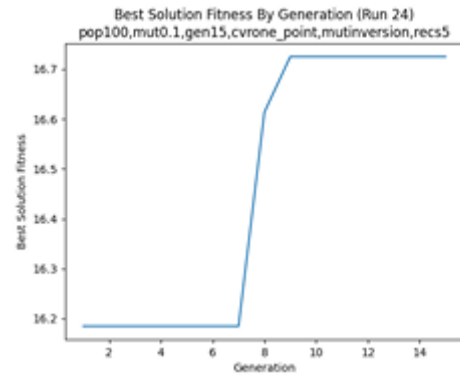
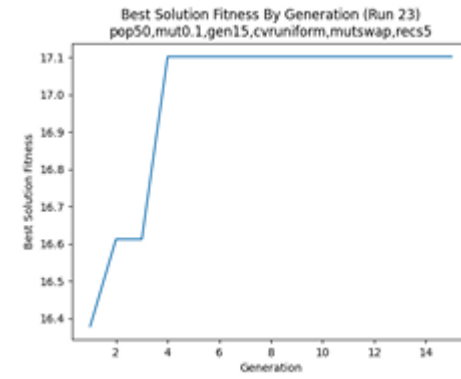
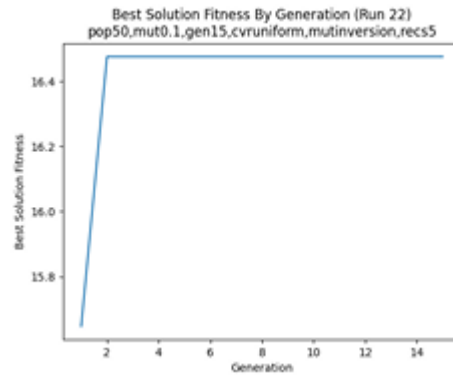
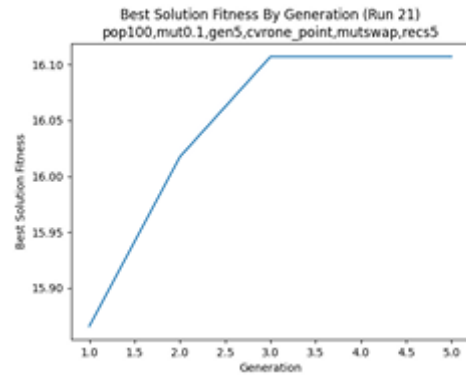
## PHỤ LỤC

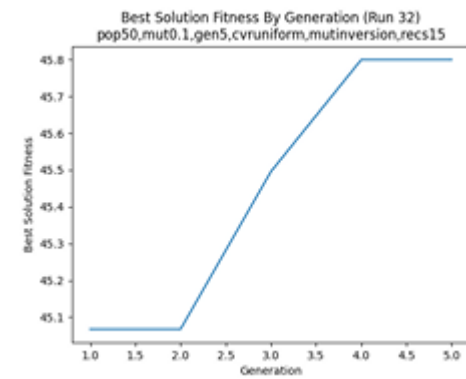
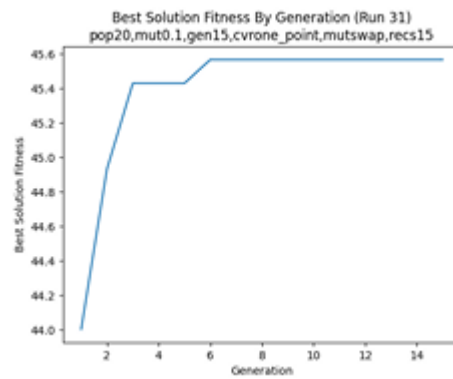
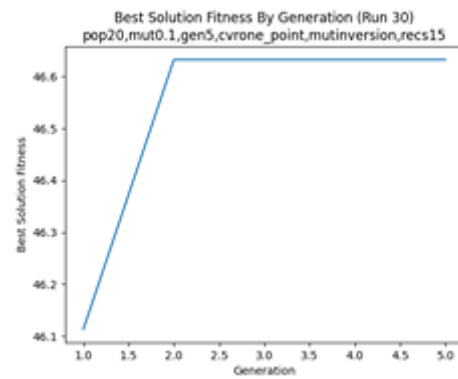
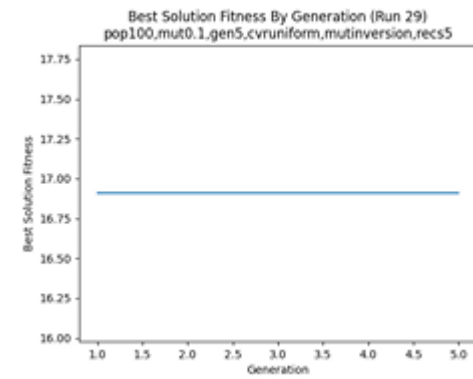
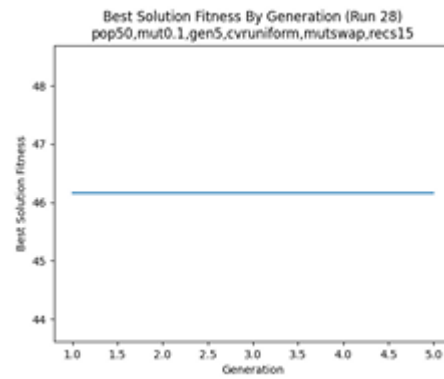
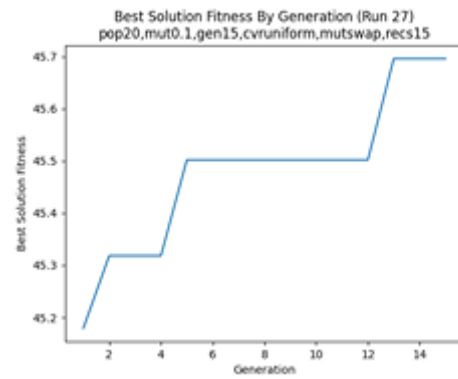




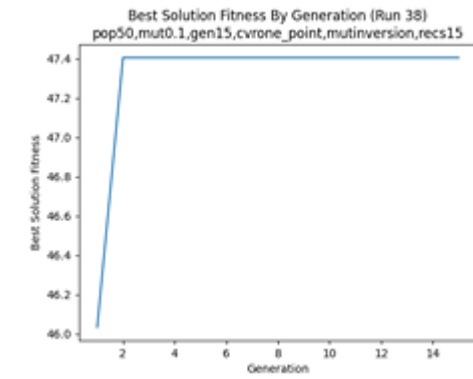
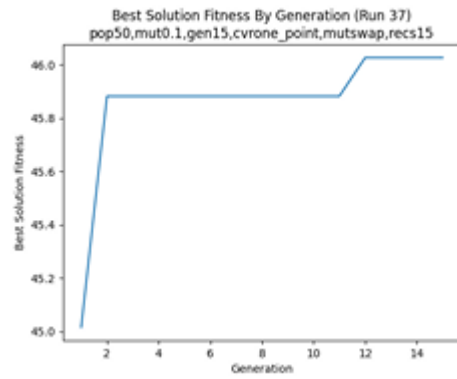
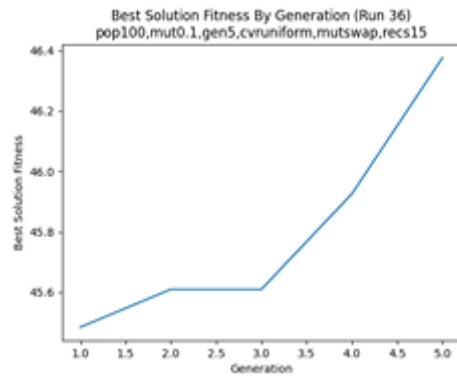
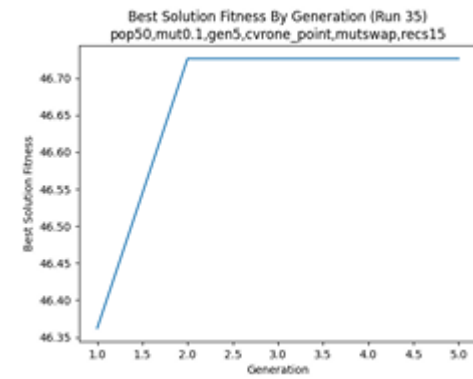
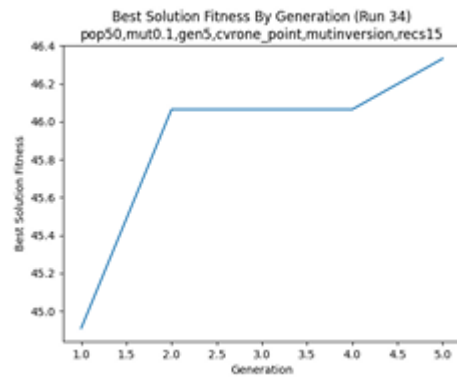
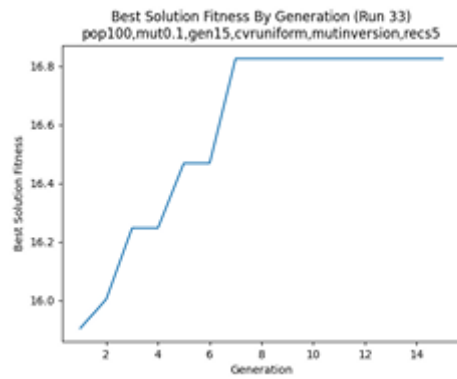


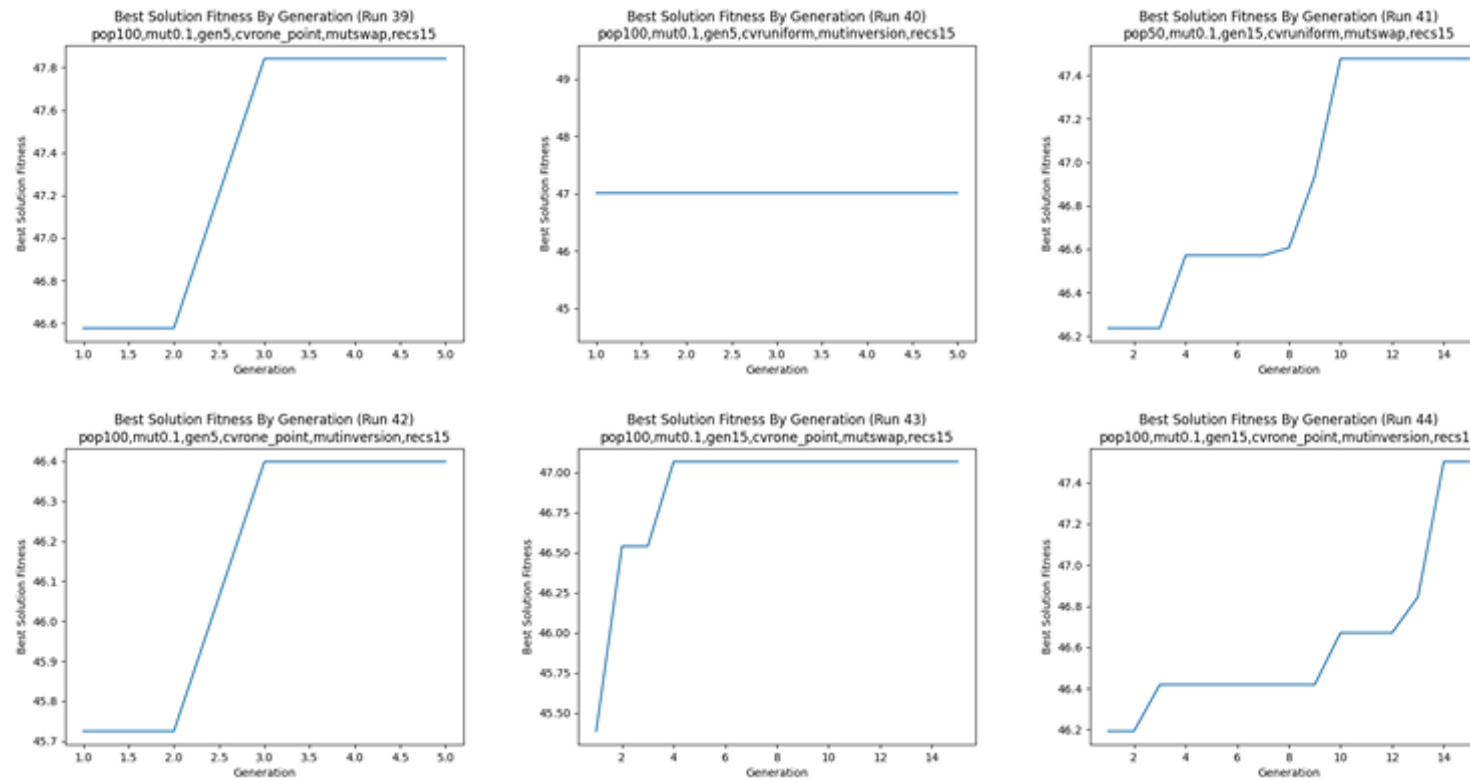












Hình 18. Giá trị hàm mục tiêu qua các thế hệ của các tổ hợp siêu tham số

Population Size	Mutation Rate	Number of Generations	Crossover Function	Mutation Function	No. of Recommendations	Best Average Solution	Best Solution Fitness	Best Solutions By Generations Fitness	Elapsed Time
-----------------	---------------	-----------------------	--------------------	-------------------	------------------------	-----------------------	-----------------------	---------------------------------------	--------------

---

20	0,1	5	uniform	swap	5	3,065159	15,3258	[14.548181187160209, 14.983263606916834, 15.04780456705009, 15.325797446809041, 15.325797446809041]	4,349995
20	0,1	5	uniform	inversion	5	3,060872	15,30436	[14.762397349618938, 14.81044079298676, 14.95818003384863, 15.30436122277594, 15.30436122277594]	5,276
20	0,1	5	one_point	swap	5	3,295193	16,47597	[16.091343445613894, 16.091343445613894, 16.17793677983803, 16.17793677983803, 16.47596540613234]	5,457001

[illegible]

[illegible]

20	0,1	15	uniform	swap	5	3,357005	16,78503	[16.130937731240092, 16.130937731240092, 16.14903295772764, 16.78502638583304, 16.78502638583304, 16.78502638583304, 16.78502638583304, 16.78502638583304, 16.78502638583304, 16.78502638583304, 16.78502638583304, 16.78502638583304, 16.78502638583304, 16.78502638583304, 16.78502638583304, 16.78502638583304]	9,231002
50	0,1	5	one_point	swap	5	3,419092	17,09546	[16.68947698402214, 16.74936828364006, 17.095461871803824, 17.095461871803824, 17.095461871803824]	13,97699



16.287406718156355]

50	0,1	5	uniform	inversion	5	3,292405	16,46202	[16.213984946743366, 16.213984946743366, 16.36596454203919, 16.462024529657544, 16.462024529657544]	17,00302
20	0,1	5	uniform	swap	15	3,082484	46,23726	[44.83957078355162, 44.83957078355162, 46.2372624634064, 46.2372624634064, 46.2372624634064]	18,309



20	0,1	5	uniform	inversion	15	3,123593	46,8539	[45.83372128154212, 46.61388663838886, 46.61388663838886, 46.61388663838886, 46.61388663838886]	19,36998
20	0,1	5	one_point	swap	15	3,101202	46,51803	[45.635383922170114, 45.85661215380753, 45.85661215380753, 45.85661215380753, 46.51802543014256]	19,41799
50	0,1	15	one_point	swap	5	3,331706	16,65853	[16.101341312683537, 16.13884161870752, 16.13884161870752, 16.13884161870752, 16.15290106180995, 16.15290106180995, 16.15290106180995, 16.15290106180995, 16.15442474367422, 16.658531891203296, 16.658531891203296, 16.658531891203296, 16.658531891203296, 16.658531891203296,	20,12105

16.658531891203296]

20	0,1	5	one_point	inversion	5	3,092746	15,46373	[15.413079865645592, 15.463730298772187, 15.463730298772187, 15.463730298772187, 15.463730298772187]	20,49604
----	-----	---	-----------	-----------	---	----------	----------	------------------------------------------------------------------------------------------------------------------	----------

50	0,1	15	one_point	inversion	5	3,197112	15,98556	[15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075, 15.985558372343075]	22,234
100	0,1	5	uniform	swap	5	3,412795	17,06398	[17.063976119080174, 17.063976119080174, 17.063976119080174, 17.063976119080174, 17.063976119080174]	26,13097

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]



20	0,1	15	uniform	inversion	15	3,08584	46,2876	[45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.63150303815563, 45.80718260091508, 45.80718260091508]	46,72299
----	-----	----	---------	-----------	----	---------	---------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------

20	0,1	15	uniform	swap	15	3,107736	46,61604	[45.179433074403654, 45.31815489236716, 45.31815489236716, 45.31815489236716, 45.501735609184905, 45.501735609184905, 45.501735609184905, 45.501735609184905, 45.501735609184905, 45.501735609184905, 45.501735609184905, 45.501735609184905, 45.501735609184905, 45.501735609184905, 45.69563416217045, 45.69563416217045, 45.69563416217045]	47,40301
50	0,1	5	uniform	swap	15	3,191034	47,86551	[46.15464738591462, 46.15464738591462, 46.15464738591462, 46.15464738591462, 46.15464738591462]	50,15419

[illegible]

45.56720422694524]

50	0,1	5	uniform	inversion	15	3,110022	46,65033	[45.06733642266402, 45.06733642266402, 45.495598042214205, 45.800002141705924, 45.800002141705924]	55,05406
----	-----	---	---------	-----------	----	----------	----------	----------------------------------------------------------------------------------------------------------------	----------

100	0,1	15	uniform	inversion	5	3,365009	16,82505	[15.905571606570602, 16.003948410508336, 16.247276728656484, 16.247276728656487, 16.468383604099753, 16.468383604099753, 16.825046267166712, 16.825046267166712, 16.825046267166712, 16.825046267166712, 16.825046267166712, 16.825046267166712, 16.825046267166712, 16.825046267166712, 16.825046267166712, 16.825046267166712]	55,76
50	0,1	5	one_point	inversion	15	3,094568	46,41853	[44.910264974756814, 46.064981103713414, 46.064981103713414, 46.064981103713414, 46.33265568366622]	60,22002

50	0,1	5	one_point	swap	15	3,115088	46,72632	[46.362085581625216, 46.72632244659397, 46.72632244659397, 46.72632244659397, 46.72632244659397]	60,26002
100	0,1	5	uniform	swap	15	3,121441	46,82161	[45.48446392045012, 45.60971207371502, 45.60971207371502, 45.92618165666816, 46.37674274976416]	77,63399
50	0,1	15	one_point	swap	15	3,109812	46,64718	[45.016311613876894, 45.882595569606124, 45.882595569606124, 45.882595569606124, 45.882595569606124, 45.882595569606124, 45.882595569606124, 45.882595569606124, 45.882595569606124, 45.882595569606124, 45.882595569606124, 46.02716467905802, 46.02716467905802, 46.02716467905802,	125,8049

46.02716467905802]

50	0,1	15	one_point	inversion	15	3,16042	47,4063	[46.03370892114694, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034, 47.40630069184034]	126,945
100	0,1	5	one_point	swap	15	3,189438	47,84157	[46.57654165419948, 46.57654165419948, 47.841567725826245, 47.841567725826245, 47.841567725826245]	145,423



[illegible]

[illegible]

100	0,1	15	one_point	inversion	15	3,166799	47,50198	[46.19240542334385, 46.19240542334385, 46.418485698670864, 46.418485698670864, 46.418485698670864, 46.418485698670864, 46.418485698670864, 46.418485698670864, 46.418485698670864, 46.67044190820925, 46.67044190820925, 46.67044190820925, 46.8444786034708, 47.50198125766699, 47.50198125766699]	194,2506
-----	-----	----	-----------	-----------	----	----------	----------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------

100	0,1	15	uniform	swap	5	3,34371	16,71855	[16.438932797669274, 16.438932797669274, 16.438932797669274, 16.50199431902146, 16.50199431902146, 16.50199431902146, 16.50199431902146, 16.561320243060795, 16.561320243060795, 16.561320243060795, 16.561320243060795, 16.561320243060795, 16.561320243060795, 16.561320243060795, 16.561320243060795, 16.718548149211987]	217,4646
-----	-----	----	---------	------	---	---------	----------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------

[illegible]

100	0,1	15	uniform	inversion	15	3,141836	47,12753	[45.94368187625238, 46.38685374112072, 46.38685374112072, 46.38685374112072, 46.38685374112072, 46.38685374112072, 46.38685374112072, 46.38685374112072, 46.54756049399956, 46.54756049399956, 46.54756049399956, 46.82893629818911, 46.82893629818911, 46.82893629818911, 46.82893629818911]	409,1342
-----	-----	----	---------	-----------	----	----------	----------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------

[illegible]

*Bảng 3. Thống kê kết quả cho các tổ hợp siêu tham số*

## TƯ LIỆU THAM KHẢO

1. Shardanand, U., Maes, P., 1995. Social information filtering: Algorithms for automating “word of mouth, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '95, ACM Press/Addison-Wesley Publishing Co., USA, pp. 210–217. doi:10.1145/223904.223931.
3. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J., 1994. Grouplens: An open architecture for collaborative filtering of netnews. In: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, CSCW '94, Association for Computing Machinery, New York, NY, USA, pp. 175–186. doi:10.1145/192844.192905.
4. [Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., 2010. Recommender Systems Handbook. Springer-Verlag, Berlin, Heidelberg.](#)
5. Sarwar, B., Karypis, G., Konstan, J., Riedl, J., 2001. Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web, WWW '01, Association for Computing Machinery, New York, NY, USA, pp. 285–295. doi:10.1145/ 371920.372071.
6. [Aggarwal, C.C., 2016. Recommender Systems: The Textbook. Springer Publishing Company Incorporated.](#)