

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
KHOA CÔNG NGHỆ THÔNG TIN 1



Nhóm bài tập lớn : Nhóm 09
Tên bài tập lớn : Hệ thống mua hàng theo sở thích người dùng
Danh sách thành viên :
1. Đặng Quốc Khánh – B22DCCN444
2. Vũ Trọng Khôi – B22DCCN468
3. Phạm Trung Kiên – B22DCCN432

Bộ môn: Phát triển hệ thống thông minh
Giảng viên hướng dẫn: TS. Nguyễn Mạnh Hùng

HÀ NỘI - 2025

MỤC LỤC

MỤC LỤC	i
DANH MỤC TỪ VIẾT TẮT	iii
MỤC LỤC BẢNG BIỂU	iv
MỤC LỤC HÌNH ẢNH	v
MỞ ĐẦU	1
CHƯƠNG 1 : BÀI TOÁN GỢI Ý MUA HÀNG THEO SỞ THÍCH NGƯỜI DÙNG	2
1.1. Phát biểu bài toán	2
1.2. Các hướng tiếp cận.....	2
1.3. Đề xuất giải pháp.....	3
CHƯƠNG 2 : CƠ SỞ ĐỂ XÂY DỰNG ỨNG DỤNG GỢI Ý MUA HÀNG THEO SỞ THÍCH NGƯỜI DÙNG	5
2.1. Xây dựng bộ dữ liệu.....	5
2.2. Naive Bayes.....	6
2.2.1 Giới thiệu.....	6
2.2.2 Kịch bản và phương pháp đánh giá.....	6
2.2.3 Quá trình đánh giá.....	7
2.3. KNN (K – Nearest Neighbor)	8
2.3.1 Giới thiệu.....	8
2.3.2 Kịch bản và phương pháp đánh giá.....	9
2.3.3 Quá trình đánh giá.....	10
2.4. NCF (Neural Collaborative Filtering).....	11
2.4.1 Giới thiệu.....	11
2.4.2 Kịch bản và phương pháp đánh giá.....	11
2.4.3 Quá trình đánh giá.....	12
2.5. Kết quả đánh giá.....	13
CHƯƠNG 3 : THIẾT KẾ HỆ THỐNG TRÊN NỀN TẢNG WEBSITE ...	15

3.1. Thiết kế các lớp thực thể liên quan	15
3.2. Thiết kế cơ sở dữ liệu liên quan	16
3.3. Thiết kế hệ thống chức năng mua hàng	17
3.3.1 Thiết kế giao diện.....	17
3.3.2 Thiết kế biểu đồ lớp chi tiết	18
3.3.3 Thiết kế biểu đồ hoạt động và biểu đồ tuần tự.....	21
3.3.4 Kết quả chạy ứng dụng.....	27
KẾT LUẬN	30
TÀI LIỆU THAM KHẢO	31

DANH MỤC TỪ VIẾT TẮT

STT	Từ viết tắt	Ý nghĩa đầy đủ
1	KNN	K-Nearest Neighbors
2	NCF	Neural Collaborative Filtering
3	SVD	Singular Value Decomposition
4	MLP	Multi-Layer-Perceptron
5	CSDL	Cơ sở dữ liệu

MỤC LỤC BẢNG BIỂU

Bảng 2.1: Thống kê số mẫu tập huấn luyện và tập kiểm thử thuật toán Naive Bayes	7
Bảng 2.2: Tổng hợp các thông số đánh giá thuật toán Naive Bayes.....	8
Bảng 2.3: Thống kê số mẫu tập huấn luyện và tập kiểm thử thuật toán KNN	9
Bảng 2.4: Tổng hợp các thông số đánh giá thuật toán KNN	11
Bảng 2.5: Thống kê số mẫu tập huấn luyện và tập kiểm thử thuật toán NCF	12
Bảng 2.6: Tổng hợp các thông số đánh giá thuật toán NCF	13
Bảng 2.7: Bảng tổng hợp kết quả đánh giá các thuật toán.....	14

MỤC LỤC HÌNH ẢNH

Hình 2.1: Kết quả chạy lần 1 của thuật toán Naive Bayes.....	7
Hình 2.2: Kết quả chạy lần 2 của thuật toán Naive Bayes.....	7
Hình 2.3: Kết quả chạy lần 3 của thuật toán Naive Bayes.....	8
Hình 2.4: Kết quả chạy lần 4 của thuật toán Naive Bayes.....	8
Hình 2.5: Kết quả chạy lần 5 của thuật toán Naive Bayes.....	8
Hình 2.6: Kết quả chạy lần 1 của thuật toán KNN	10
Hình 2.7: Kết quả chạy lần 2 của thuật toán KNN	10
Hình 2.8: Kết quả chạy lần 3 của thuật toán KNN	10
Hình 2.9: Kết quả chạy lần 4 của thuật toán KNN	10
Hình 2.10: Kết quả chạy lần 5 của thuật toán KNN	10
Hình 2.11: Kết quả chạy lần 1 của thuật toán NCF	12
Hình 2.12: Kết quả chạy lần 2 của thuật toán NCF	12
Hình 2.13: Kết quả chạy lần 3 của thuật toán NCF	13
Hình 2.14: Kết quả chạy lần 4 của thuật toán NCF	13
Hình 2.15: Kết quả chạy lần 5 của thuật toán NCF	13
Hình 3.1: Biểu đồ các lớp thực thể liên quan.....	16
Hình 3.2: Biểu đồ cơ sở dữ liệu	17
Hình 3.3: Thiết kế giao diện trang đề xuất sản phẩm	17
Hình 3.4. Thiết kế giao diện chi tiết sản phẩm	18
Hình 3.5: Biểu đồ lớp chi tiết.....	21
Hình 3.6: Biểu đồ hoạt động	22
Hình 3.7. Biểu đồ tuần tự.....	24
Hình 3.8: Kết quả giao diện trang đề xuất	27
Hình 3.9: Kết quả hiển thị danh sách đề xuất	27
Hình 3.10: Kết quả hiển thị chi tiết sản phẩm.....	28
Hình 3.11: Kết quả hiển thị yêu cầu xác nhận	28
Hình 3.12: Kết quả hiển thị mua hàng thành công.....	29

MỞ ĐẦU

Trong môi trường thương mại điện tử hiện nay, số lượng sản phẩm và người dùng tăng lên nhanh chóng khiến việc tìm kiếm sản phẩm phù hợp trở nên khó khăn hơn. Người dùng thường phải mất nhiều thời gian duyệt qua hàng trăm, thậm chí hàng nghìn sản phẩm để chọn được mặt hàng đúng nhu cầu. Điều này gây ra trải nghiệm mua sắm kém hiệu quả cho khách hàng. Xuất phát từ thực tế đó, nhóm chọn đề tài “Hệ thống gợi ý mua hàng theo sở thích người dùng” nhằm xây dựng một hệ thống hỗ trợ người dùng trong quá trình mua sắm trực tuyến, giải quyết vấn đề thực tiễn nói trên.

Mục tiêu chính của đề tài là thiết kế, xây dựng một hệ thống có khả năng thu thập dữ liệu người dùng, bao gồm lịch sử mua sắm, hành vi tương tác giữa khách hàng và sản phẩm. Từ đó, hệ thống áp dụng các mô hình/thuật toán gợi ý đề xuất, gợi ý các sản phẩm phù hợp với sở thích của từng khách hàng.

Báo cáo đề tài được chia thành 3 chương chính với mỗi chương có nội dung cụ thể như sau:

- + Chương 1: Trình bày bài toán và các hướng tiếp cận liên quan. Đề xuất phương án giải quyết.
- + Chương 2: Trình bày việc lựa chọn các mô hình/thuật toán, quy trình đánh giá và kết quả đánh giá.
- + Chương 3: Thiết kế hệ thống theo các chức năng cá nhân.

CHƯƠNG 1 : BÀI TOÁN GỢI Ý MUA HÀNG THEO SỞ THÍCH NGƯỜI DÙNG

Chương 1 trình bày tổng quan về bối cảnh cũng như tầm quan trọng của bài toán gợi ý mua hàng theo sở thích của người dùng, đồng thời cung cấp một số hướng tiếp cận cũng như đề xuất phương pháp giải quyết được sử dụng trong đề tài.

1.1. Phát biểu bài toán

Trong bối cảnh ngành thương mại điện tử phát triển bùng nổ, các sàn mua sắm trực tuyến như Shopee, Lazada, Tiki, Amazon hay các website bán hàng riêng lẻ đang cung cấp hàng triệu sản phẩm thuộc nhiều danh mục khác nhau. Số lượng người dùng truy cập và thực hiện giao dịch ngày càng tăng, dẫn đến khối lượng dữ liệu phát sinh khổng lồ mỗi ngày.

Tuy nhiên, sự phong phú về sản phẩm lại mang đến thách thức lớn: người dùng thường gặp khó khăn trong việc tìm ra sản phẩm phù hợp với nhu cầu cá nhân giữa hàng nghìn lựa chọn. Quá trình tìm kiếm và lọc thông tin mất nhiều thời gian, khiến trải nghiệm mua sắm trở nên kém hiệu quả và giảm khả năng quay lại của khách hàng. Vấn đề này cũng mang đến nhiều hệ quả cho phía doanh nghiệp như: giảm hiệu suất tiếp cận khách hàng do đề xuất sản phẩm không đúng đối tượng; tỉ lệ chuyển đổi thấp do khách hàng không tìm được sản phẩm họ thực sự quan tâm; giảm mức độ gắn bó, trung thành của khách hàng.

Với vấn đề được trên, bài toán cấp thiết được đặt ra là làm thế nào để đề xuất được đến từng khách hàng những sản phẩm phù hợp nhất với nhu cầu, sở thích của họ. Đề tài “Hệ thống gợi ý mua hàng theo sở thích người dùng” được nhóm chọn để xây dựng với mục đích tạo ra hệ thống có khả năng thu thập dữ liệu người dùng, áp dụng các mô hình/thuật toán hiện đại nhằm phân tích và đề xuất những sản phẩm phù hợp với nhu cầu, sở thích của từng khách hàng cụ thể.

1.2. Các hướng tiếp cận

Đối với bài toán hệ đề xuất, theo T.Alexander [4], có 3 cách tiếp cận chính như sau:

- + Gợi ý dựa trên nội dung (Content-based recommendations): người dùng sẽ được gợi ý những mục (sản phẩm, bài viết, v.v.) tương tự với những

gì họ đã thích trong quá khứ. Phương pháp này gặp các vấn đề như: phụ thuộc nhiều vào dữ liệu mô tả của vật phẩm, không đa dạng với những sản phẩm chưa từng tương tác và không hoạt động hiệu quả với người dùng mới.

- + Gợi ý cộng tác (Collaborative recommendations): người dùng sẽ được gợi ý những mục mà những người có sở thích và thị hiếu tương tự đã thích trước đó. Phương pháp này gặp các vấn đề như: không hoạt động hiệu quả với người dùng mới, vật phẩm mới do thiếu dữ liệu tương tác hoặc khi dữ liệu bị thưa.
- + Phương pháp lai (Hybrid approaches): những phương pháp này kết hợp giữa gợi ý cộng tác và gợi ý dựa trên nội dung nhằm khắc phục nhược điểm của 2 phương pháp trên.

Với bài toán đề xuất sản phẩm dựa trên sở thích người dùng, báo cáo quan tâm đến tương tác giữa người dùng và sản phẩm, do đó các thuật toán sử dụng phương pháp gợi ý cộng tác sẽ được tìm hiểu và sử dụng trong hệ thống này.

1.3. Đề xuất giải pháp

Với bài toán đề xuất sản phẩm dựa trên sở thích người dùng, cụ thể là dựa vào lịch sử các lần chọn mua, chọn xem, đề tài quan tâm chủ yếu đến tương tác giữa người dùng và sản phẩm. Do đó, báo cáo lựa chọn sử dụng các mô hình/thuật toán sử dụng phương pháp gợi ý cộng tác (Collaborative recommendations) để áp dụng cho đề tài này.

Cụ thể, cách giải quyết được đề xuất gồm các bước chính như sau:

- + Thu thập và tiền xử lý dữ liệu: Dữ liệu là hành vi xem, mua của người dùng đối với sản phẩm và sẽ được chuẩn hóa cho phù hợp với mô hình học máy.
- + Áp dụng mô hình/thuật toán sử dụng phương pháp gợi ý cộng tác nhằm tìm ra mức độ tương thích giữa cặp người dùng-sản phẩm.
- + Xây dựng giao diện web/winform: Hiển thị danh sách gợi ý sản phẩm phù hợp với từng người dùng, cho phép người dùng tương tác với sản phẩm.

Kết chương

Chương 1 đã trình bày tổng quan về đề tài, bài toán đặt ra, các phương pháp tiếp cận phổ biến trong hệ thống gợi ý sản phẩm và cách giải quyết mà báo cáo lựa chọn để thực hiện. Đây là nền tảng lý luận quan trọng để nhóm tiến hành nghiên cứu chi tiết các mô hình và thuật toán gợi ý trong chương tiếp theo.

CHƯƠNG 2 : CƠ SỞ ĐỂ XÂY DỰNG ỨNG DỤNG GỢI Ý MUA HÀNG THEO SỞ THÍCH NGƯỜI DÙNG

Sau khi đã xác định bài toán cũng như phương pháp giải quyết của đề tại ở chương 1, chương 2 đi sâu vào việc lựa chọn, đánh giá các mô hình/thuật toán phù hợp với hệ thống gợi ý mua hàng.

Chương 2 sẽ giới thiệu và đánh giá 3 thuật toán phổ biến cho hệ thống gợi ý: Naive Bayes, K-Nearest Neighbors và Neural Collaborative Filtering. Để có cơ sở đánh giá khách quan, chương 2 sẽ trình bày phương pháp xây dựng bộ dữ liệu, quá trình đánh giá từng thuật toán và so sánh giữa 3 thuật toán với nhau.

2.1. Xây dựng bộ dữ liệu

Đề tài thu thập dữ liệu là tương tác của người dùng với sản phẩm của một trang web thương mại điện tử. Nguồn dữ liệu từ trang Kaggle:
[<https://www.kaggle.com/datasets/mkechinov/ecommerce-events-history-in-cosmetics-shop>]

- + Mỗi mẫu của dữ liệu đại diện cho tương tác của một khách hàng với một sản phẩm cụ thể với ba trường quan trọng: `user_id`, `product_id` và `event_type`. Trong đó `event_type` có các giá trị: `view`, `purchase`, `cart`, `remove_from_cart` tương đương với các loại tương tác: xem sản phẩm, mua sản phẩm, thêm sản phẩm vào giỏ hàng, xóa sản phẩm khỏi giỏ hàng.
- + Tập mẫu được gán nhãn theo trường hợp: Chỉ có 2 nhãn, mỗi mẫu 1 nhãn
- + Bộ dữ liệu được chia đều với 2 nhãn:
 - Tương thích cao (1) - tương ứng với hành động mua sản phẩm.
 - Tương thích thấp (0) - tương ứng với hành động xem sản phẩm, thêm giỏ hàng, xóa khỏi giỏ hàng.

Sau khi gán nhãn, báo cáo thực hiện các bước xử lý:

- + Loại bỏ các trường dữ liệu không cần thiết (`event_time`, `category_id`,...)
- + Loại bỏ các mẫu bị xung đột, thừa: cùng một `product` và một `user` nhưng có nhiều mẫu thì gộp lại thành một mẫu (nếu tồn tại mẫu nhãn 1 thì ưu tiên gộp thành mẫu nhãn 1).

Báo cáo thu được tập mẫu gồm 49676 mẫu, chia đều số lượng cho 2 loại nhãn. Mỗi mẫu có dạng <user_id, product_id, label (0/1)>

2.2. Naive Bayes

2.2.1 Giới thiệu

Naive Bayes là một bộ phân loại xác suất áp dụng định lý Bayes, với giả định rằng các đặc trưng là độc lập với nhau. Thuật toán này tính toán xác suất của việc một vật nào đó có liên quan dựa trên các xác suất có điều kiện của nhiều đặc trưng khác nhau, với giả định rằng mỗi đặc trưng đóng góp độc lập vào kết quả cuối cùng [1]. Với bài toán này, Naive Bayes sử dụng các đặc trưng để tính toán xác suất nhằm xác định mức độ phù hợp giữa người dùng và sản phẩm, từ đó phân loại để đưa ra sản phẩm được đề xuất hay không.

Naive Bayes được sử dụng rộng rãi với bài toán hệ đề xuất vì nó triển khai rất đơn giản, chạy nhanh và có khả năng hoạt động tốt ngay cả với tập dữ liệu nhỏ [2], [3]. Tuy nhiên, việc giả định các đặc trưng là độc lập là một nhược điểm lớn, vì trong thực tế, các dữ liệu thường có mối liên hệ với nhau [3]. Điều này trái ngược với giả định của Naive Bayes nên có thể làm cho độ chính xác và hiệu suất bị giảm.

2.2.2 Kịch bản và phương pháp đánh giá

a) Chiến lược chia dữ liệu

Để đánh giá thuật toán, báo cáo thực hiện huấn luyện với tập mẫu trong 5 lần. Mỗi lần huấn luyện, tập mẫu sẽ được chia ngẫu nhiên theo tỉ lệ 7/3 thành tập huấn luyện (~70%) và tập kiểm thử (~ 30%) cụ thể như sau:

Lần huấn luyện	Số mẫu tập huấn luyện	Số mẫu tập kiểm thử
1	35930	13746
2	35540	14136
3	35034	14642
4	35008	14668
5	35564	14112

Bảng 2.1: Thống kê số mẫu tập huấn luyện và tập kiểm thử thuật toán Naive Bayes

b) Phương pháp đánh giá

Phương pháp đánh giá được sử dụng là “Đánh giá ngẫu nhiên” với mỗi lần huấn luyện sẽ đo các tham số: Accuracy, Precision, Recall, F1 Score. Lặp lại các bước 5 lần và tính trung bình các tham số để đánh giá thuật toán.

2.2.3 Quá trình đánh giá

Để đánh giá, báo cáo thực hiện tính toán các thông số accuracy, precision, recall và F1-score

Kết quả chạy test:

```
Run 1
Train samples: 35930, Test samples: 13746
Accuracy: 65.74% | Precision: 76.46% | Recall: 45.01% | F1: 56.66%
```

Hình 2.1: Kết quả chạy lần 1 của thuật toán Naive Bayes

```
Run 2
Train samples: 35540, Test samples: 14136
Accuracy: 65.33% | Precision: 76.52% | Recall: 43.64% | F1: 55.58%
```

Hình 2.2: Kết quả chạy lần 2 của thuật toán Naive Bayes

```
Run 3
Train samples: 35034, Test samples: 14642
Accuracy: 65.14% | Precision: 77.32% | Recall: 43.68% | F1: 55.82%
```

Hình 2.3: Kết quả chạy lần 3 của thuật toán Naive Bayes

```
Run 4
Train samples: 35008, Test samples: 14668
Accuracy: 65.62% | Precision: 76.96% | Recall: 44.06% | F1: 56.04%
```

Hình 2.4: Kết quả chạy lần 4 của thuật toán Naive Bayes

```
Run 5
Train samples: 35564, Test samples: 14112
Accuracy: 65.55% | Precision: 76.40% | Recall: 44.04% | F1: 55.87%
```

Hình 2.5: Kết quả chạy lần 5 của thuật toán Naive Bayes

Bảng tổng hợp kết quả các lần chạy của thuật toán Naive Bayes:

Lần huấn luyện	Accuracy	Precision	Recall	F1
1	65.74%	76.46%	45.01%	56.66%
2	65.33%	76.52%	43.64%	55.58%
3	65.14%	77.32%	43.68%	55.82%
4	65.62%	76.96%	44.06%	56.04%
5	65.55%	76.40%	44.04%	55.87%
Average	65.48%	76.73%	44.09%	55.99%

Bảng 2.2: Tổng hợp các thông số đánh giá thuật toán Naive Bayes

2.3. KNN (K – Nearest Neighbor)

2.3.1 Giới thiệu

Thuật toán K-Nearest Neighbors (KNN) là một thuật toán đơn giản nhưng mạnh mẽ, được sử dụng cho các bài toán phân loại và hồi quy. Nguyên lý cốt lõi của KNN là xác định “k” điểm dữ liệu gần nhất với một điểm truy vấn, dựa trên một thước đo khoảng cách như khoảng cách Euclid. Trong hệ thống gợi ý, KNN

hoạt động bằng cách tìm những người dùng hoặc sản phẩm có mức độ tương đồng cao nhất với người dùng hoặc sản phẩm đang xét, sau đó đưa ra gợi ý dựa trên sở thích của các hàng xóm gần nhất này [1].

Thuật toán KNN có những ưu điểm như dễ triển khai, dễ hiểu, hiệu quả với các tập dữ liệu có kích thước nhỏ đến trung bình, không cần phải huấn luyện mô hình. Tuy nhiên, nó tốn kém chi phí khi tính toán với tập dữ liệu lớn, tốn bộ nhớ để lưu toàn bộ dữ liệu và hiệu suất có thể bị giảm khi gặp tình trạng dữ liệu thưa [5]. Bên cạnh đó, việc mở rộng gặp nhiều khó khăn vì khi mở rộng, bộ dữ liệu trở nên lớn hơn khiến cho chi phí tính toán trở nên tốn kém, thời gian tính toán lâu hơn rất nhiều.

2.3.2 Kịch bản và phương pháp đánh giá

a) Chiến lược chia dữ liệu

Để đánh giá thuật toán, báo cáo thực hiện huấn luyện với tập mẫu trong 5 lần. Mỗi lần huấn luyện, tập mẫu sẽ được chia ngẫu nhiên theo tỉ lệ 7/3 thành tập huấn luyện (~70%) và tập kiểm thử (~ 30%) cụ thể như sau:

Lần huấn luyện	Số mẫu tập huấn luyện	Số mẫu tập kiểm thử
1	35930	13746
2	35540	14136
3	35034	14642
4	35008	14668
5	35564	14112

Bảng 2.3: Thống kê số mẫu tập huấn luyện và tập kiểm thử thuật toán KNN

b) Phương pháp đánh giá

Phương pháp đánh giá được sử dụng là “Đánh giá ngẫu nhiên” với mỗi lần huấn luyện sẽ đo các tham số: Accuracy, Precision, Recall, F1 Score. Lặp lại các bước 5 lần và tính trung bình các tham số để đánh giá thuật toán.

2.3.3 Quá trình đánh giá

Để đánh giá, báo cáo thực hiện tính toán các thông số accuracy, precision, recall và F1-score

Kết quả chạy test:

```
Run 1
Train samples: 35930, Test samples: 13746
Accuracy: 73.21% | Precision: 74.66% | Recall: 69.87% | F1: 72.19%
```

Hình 2.6: Kết quả chạy lần 1 của thuật toán KNN

```
Run 2
Train samples: 35540, Test samples: 14136
Accuracy: 73.39% | Precision: 74.66% | Recall: 70.32% | F1: 72.43%
```

Hình 2.7: Kết quả chạy lần 2 của thuật toán KNN

```
Run 3
Train samples: 35034, Test samples: 14642
Accuracy: 73.54% | Precision: 75.88% | Recall: 69.66% | F1: 72.64%
```

Hình 2.8: Kết quả chạy lần 3 của thuật toán KNN

```
Run 4
Train samples: 35008, Test samples: 14668
Accuracy: 73.34% | Precision: 74.59% | Recall: 70.36% | F1: 72.41%
```

Hình 2.9: Kết quả chạy lần 4 của thuật toán KNN

```
Run 5
Train samples: 35564, Test samples: 14112
Accuracy: 73.19% | Precision: 74.35% | Recall: 70.02% | F1: 72.12%
```

Hình 2.10: Kết quả chạy lần 5 của thuật toán KNN

Bảng tổng hợp kết quả các lần chạy của thuật toán KNN

Lần huấn luyện	Accuracy	Precision	Recall	F1
1	73.21%	74.66%	69.87%	72.19%
2	73.39%	74.66%	70.32%	72.43%
3	73.54%	75.88%	69.66%	72.64%
4	73.34%	74.59%	70.36%	72.41%
5	73.19%	74.35%	70.02%	72.12%
Average	73.33%	74.83%	70.05%	72.36%

Bảng 2.4: Tổng hợp các thông số đánh giá thuật toán KNN

2.4. NCF (Neural Collaborative Filtering)

2.4.1 Giới thiệu

Neural Collaborative Filtering (NCF) là phiên bản cải tiến của các hệ thống đề xuất truyền thống, sử dụng học sâu (deep learning) để đưa ra các đề xuất chính xác hơn. Đây là phiên bản nâng cao của phương pháp Collaborative Filtering. Do các phương pháp truyền thống thường dựa trên những kỹ thuật đơn giản, nên dễ bỏ sót các mối quan hệ giữa người dùng và sản phẩm. Ngược lại NCF sử dụng mạng nơ-ron nhân tạo (neural network) để học và nắm bắt những mối quan hệ đó một cách hiệu quả hơn. Thay vì sử dụng các kỹ thuật phân rã ma trận cơ bản như Singular Value Decomposition - SVD, NCF dùng mạng perceptron nhiều lớp Multi-Layer-Perceptron MLP để mô hình hóa các mối quan hệ phi tuyến tính (non-linear) chi tiết hơn giữa người dùng và sản phẩm. Nhờ đó, mô hình có thể phát hiện các tương tác phức tạp, giúp tạo ra các gợi ý chính xác hơn và cá nhân hóa hơn cho từng người dùng [6].

2.4.2 Kịch bản và phương pháp đánh giá

a) Chiến lược chia dữ liệu

Để đánh giá thuật toán, báo cáo thực hiện huấn luyện với tập mẫu trong 5 lần. Mỗi lần huấn luyện, tập mẫu sẽ được chia ngẫu nhiên theo tỉ lệ 7/3 thành tập huấn luyện (~70%) và tập kiểm thử (~ 30%) cụ thể như sau:

Lần huấn luyện	Số mẫu tập huấn luyện	Số mẫu tập kiểm thử
1	35930	13746
2	35540	14136
3	35034	14642
4	35008	14668
5	35564	14112

Bảng 2.5: Thống kê số mẫu tập huấn luyện và tập kiểm thử thuật toán NCF

b) Phương pháp đánh giá

Phương pháp đánh giá được sử dụng là “Đánh giá ngẫu nhiên” với mỗi lần huấn luyện sẽ đo các tham số: Accuracy, Precision, Recall, F1 Score. Lặp lại các bước 5 lần và tính trung bình các tham số để đánh giá thuật toán.

2.4.3 Quá trình đánh giá

Để đánh giá, báo cáo thực hiện tính toán các thông số accuracy, precision, recall và F1-score

Kết quả chạy test:

```
Run 1
Train samples: 35930, Test samples: 13746
Accuracy: 75.35% | Precision: 75.32% | Recall: 75.08% | F1: 75.20%
```

Hình 2.11: Kết quả chạy lần 1 của thuật toán NCF

```
Run 2
Train samples: 35540, Test samples: 14136
Accuracy: 75.25% | Precision: 73.30% | Recall: 78.98% | F1: 76.03%
```

Hình 2.12: Kết quả chạy lần 2 của thuật toán NCF

```
Run 3
Train samples: 35034, Test samples: 14642
Accuracy: 75.27% | Precision: 74.93% | Recall: 76.57% | F1: 75.74%
```

Hình 2.13: Kết quả chạy lần 3 của thuật toán NCF

```
Run 4
Train samples: 35008, Test samples: 14668
Accuracy: 75.15% | Precision: 76.32% | Recall: 72.54% | F1: 74.38%
```

Hình 2.14: Kết quả chạy lần 4 của thuật toán NCF

```
Run 5
Train samples: 35564, Test samples: 14112
Accuracy: 75.05% | Precision: 74.28% | Recall: 75.90% | F1: 75.08%
```

Hình 2.15: Kết quả chạy lần 5 của thuật toán NCF

Bảng tổng hợp kết quả các lần chạy của thuật toán NCF:

Lần huấn luyện	Accuracy	Precision	Recall	F1
1	75.35%	75.32%	75.08%	75.20%
2	75.25%	73.30%	78.98%	76.03%
3	75.27%	74.93%	76.57%	75.74%
4	75.15%	76.32%	72.54%	74.38%
5	75.05%	74.28%	75.90%	75.08%
Average	75.21%	74.83%	75.81%	75.29%

Bảng 2.6: Tổng hợp các thông số đánh giá thuật toán NCF

2.5. Kết quả đánh giá

Từ các kết quả đánh giá ở trên, ta có bảng tổng hợp kết quả đánh giá 3 thuật toán Naive Bayes, KNN và NCF như sau:

Thuật toán	Accuracy	Precision	Recall	F1
Naive Bayes	65.48%	76.73%	44.09%	55.99%
KNN	73.33%	74.83%	70.05%	72.36%
NCF	75.21%	74.83%	75.81%	75.29%

Bảng 2.7: Bảng tổng hợp kết quả đánh giá các thuật toán

Kết quả đánh giá cho thấy mỗi thuật toán có những ưu nhược điểm riêng, cụ thể như sau:

- + Naive Bayes có ưu điểm là đơn giản, tốc độ huấn luyện nhanh, dễ triển khai nhưng độ chính xác còn hạn chế (Accuracy 65.48%, F1-score 55.99%)
- + K-Nearest Neighbors (KNN) cải thiện đáng kể độ chính xác (Accuracy 73.33%, F1-score 72.36%)
- + Neural Collaborative Filtering (NCF) đạt hiệu quả cao nhất trong ba thuật toán (Accuracy 75.21%, F1-score 75.29%)

Từ kết quả trên, có thể kết luận rằng mô hình NCF là lựa chọn phù hợp cho hệ thống gợi ý sản phẩm theo sở thích của người dùng

Kết chương

Chương 2 đã trình bày phương pháp xây dựng bộ dữ liệu, gán nhãn và thực hiện so sánh 3 thuật toán phổ biến trong hệ thống gợi ý. NCF đã được lựa chọn để áp dụng cho đề tài. Tiếp theo, chương 3 sẽ đi sâu vào phần thiết kế hệ thống ứng dụng.

CHƯƠNG 3 : THIẾT KẾ HỆ THỐNG TRÊN NỀN TẢNG WEBSITE

Chương 3 trình bày quá trình thiết kế hệ thống của đề tài Hệ thống gợi ý mua hàng theo sở thích trên nền tảng website.

3.1. Thiết kế các lớp thực thể liên quan

Bước 1: Thực hiện thêm các thuộc tính id cho các lớp thực thể được thiết kế: User, Product, Order, OrderDetail, Cart, CartDetail.

Bước 2: Bổ sung kiểu dữ liệu cho các thuộc tính theo ngôn ngữ lập trình Java

- User: id(int), username(String), password(String), fullName(String), phoneNumber(String), email(String).
- Product: id(int), name(String), brand(String), des(String), price(float), quantity(int)
- OrderDetail: id(int), quantity(int), unitPrice(float)
- Order: id(int), createTime(date), totalPrice(float), status(String)
- Cart: id(int), lastModifiedTime(date)
- CartDetail: id(int), quantity(int), createTime(date)

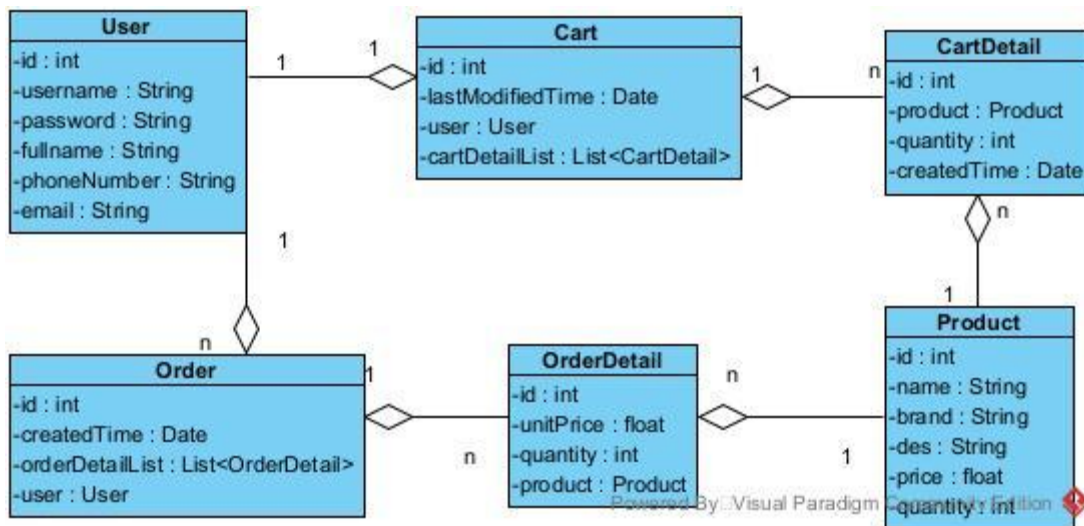
Bước 3: Chuyển quan hệ association thành aggregation/composition

- Order và Product quan hệ association ra OrderDetail → Chuyển thành Order chứa OrderDetail, OrderDetail chứa Product
- Cart và Product quan hệ association ra CartDetail → Chuyển thành Cart chứa CartDetail, CartDetail chứa Product

Bước 4: Bổ sung thuộc tính thành phần quan hệ aggregation/composition

- Order: user (User), orderDetailList (List<OrderDetail>)
- OrderDetail: product (Product)
- Cart: user (User), cartDetailList (List<CartDetail>)
- CartDetail: product (Product)

Sau khi thực hiện ta thu được Biểu đồ lớp thực thể liên quan:



Hình 3.1: Biểu đồ các lớp thực thể liên quan

3.2. Thiết kế cơ sở dữ liệu liên quan

Bước 1: Mỗi lớp thực thể đề xuất một bảng tương ứng → Ta có các bảng sau: tblUser, tblOrder, tblOrderDetail, tblProduct.

Bước 2: Với mỗi lớp, chuyển các thuộc tính không phải đối tượng thành thuộc tính của các bảng tương ứng:

- tblUser: id, username, password, fullName, email, phoneNumber.
- tblProduct: id, name, brand, des, price, quantity.
- tblOrderDetail: id, quantity, unitPrice.
- tblOrder: id, createdTime, totalPrice.
- tblCart: id, lastModifiedTime
- tblCartDetail: id, quantity, createdTime.

Bước 3: Chuyển quan hệ số lượng giữa lớp thực thể thành quan hệ số lượng giữa các bảng

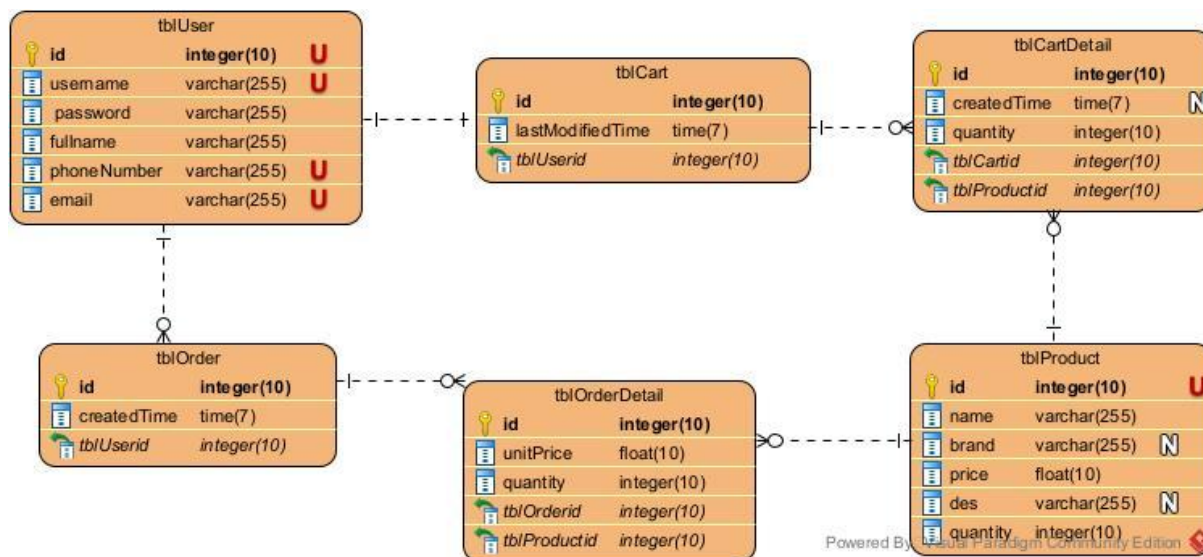
- tblUser 1 - n tblBill
- tblOrder 1 - n tblOrderDetail
- tblProduct 1 - n tblOrderDetail
- tblUser 1 - 1 tblCart
- tblCart 1 - n tblCartDetail
- tblProduct 1 - n tblCartDetail

Bước 4: Bổ sung thuộc tính khóa: khóa chính được thiết lập với thuộc tính id của tất cả bảng trên

Bước 5: Loại bỏ các thuộc tính dẫn xuất/trùng lặp

- Loại bỏ thuộc tính totalPrice của tblOrder (thuộc tính dẫn xuất).

Sau khi thực hiện các bước trên ta thu được biểu đồ cơ sở dữ liệu



Hình 3.2: Biểu đồ cơ sở dữ liệu

3.3. Thiết kế hệ thống chức năng mua hàng

3.3.1 Thiết kế giao diện

- Giao diện xem sản phẩm đề xuất

Trang đề xuất sản phẩm

Xin chào user abc

Tìm kiếm

Xem đề xuất

Danh sách đề xuất

STT	Tên sản phẩm	Giá VNĐ	Hành động
1	Điện thoại 101	3.000.000	Xem chi tiết
2	Điện thoại 102	4.000.000	Xem chi tiết
3	TV 101	8.000.000	Xem chi tiết

Hình 3.3: Thiết kế giao diện trang đề xuất sản phẩm

Giao diện này là nơi người dùng có thể xem danh sách các sản phẩm đề xuất tương ứng với sở thích, nhu cầu của cá nhân.

b. Giao diện chi tiết sản phẩm

Trang chi tiết sản phẩm

Xin chào user abc

Mã sản phẩm: 100001

Tên sản phẩm: Điện thoại 101

Thương hiệu: PTIT

Giá thành: 3.000.000 VNĐ

Số lượng còn: 15

Mô tả: Điện thoại xin bảo hành 12 tháng

Số lượng mua:

Mua hàng

Hình 3.4. Thiết kế giao diện chi tiết sản phẩm

Giao diện này là nơi người dùng xem thông tin chi tiết của một sản phẩm đồng thời có thể thực hiện thao tác mua hàng.

3.3.2 Thiết kế biểu đồ lớp chi tiết

Biểu đồ lớp chi tiết gồm các lớp như sau:

a. Các lớp giao diện

* *Giao diện xem đề xuất sản phẩm:*

Đề xuất lớp recommendPage.html

+ Thuộc tính tường minh:

- searchProduct: input
- tblProduct: table
- btnViewRecommend: button
- btnSearch: button

+ Thuộc tính ẩn:

- user: User
- recommendList: List<Product>

** Giao diện xem chi tiết sản phẩm:*

Đề xuất lớp productDetail.html

+ Thuộc tính tường minh:

- productIdTxt, productNameTxt, productPriceTxt, productBrandTxt, productDesTxtL: txt
- quantityPurchase: input
- btnPurchase: button
- btnAddToCart: button.

+ Thuộc tính ẩn:

- user: User
- product: Product

b. Lớp điều khiển

Đề xuất lớp ServerController thực hiện các nhiệm vụ:

+ Xử lý hiển thị trang xem đề xuất -> Đề xuất một phương thức:

Input: HttpSession (Lưu phiên đang nhập), Model (lưu dữ liệu chuyển sang giao diện)

Output: Tên trang html cần hiển thị.

=> Phương thức: showRecommendList(session: HttpSession, model: Model): String

+ Xử lý hiển thị trang xem chi tiết -> Đề xuất một phương thức:

Input: HttpSession, Model, productId

Output: Tên trang html cần hiển thị

=> Phương thức: showProductDetail(session: HttpSession, model: Model, productId: int): String

+ Xử lý mua hàng-> Đề xuất một phương thức:

Input: HttpSession, Model, quantity

Output: Tên trang html cần hiển thị.

=> Phương thức: buyProduct(session: HttpSession, model: Model, quantity: int): String

+ Xử lý gửi Log của người dùng về Server học máy -> Đề xuất một phương thức

Input: userId, productId, actionType

Output: không có

=> Phương thức: `sendLog(userId: int, productId: int, actionType: String):void.`

c. Các lớp giao tiếp cơ sở dữ liệu

* *ProductDAO*:

Giao tiếp với CSDL sản phẩm, thực hiện các nhiệm vụ:

+ Lấy danh sách sản phẩm theo danh sách id -> Đề xuất một phương thức

Input: danh sách id

Output: danh sách sản phẩm

=> Phương thức: `getProductsByIdList(idList: List<Integer>): List<Product>`

+ Lấy thông tin một sản phẩm theo id -> Đề xuất một phương thức

Input: id sản phẩm

Output: thông tin sản phẩm

=> Phương thức: `getById(id: int): Product`

* *OrderDAO*:

Giao tiếp với CSDL order, thực hiện nhiệm vụ:

+ Thêm mới Order đồng thời trừ số lượng của các Product được mua -> Đề xuất một phương thức

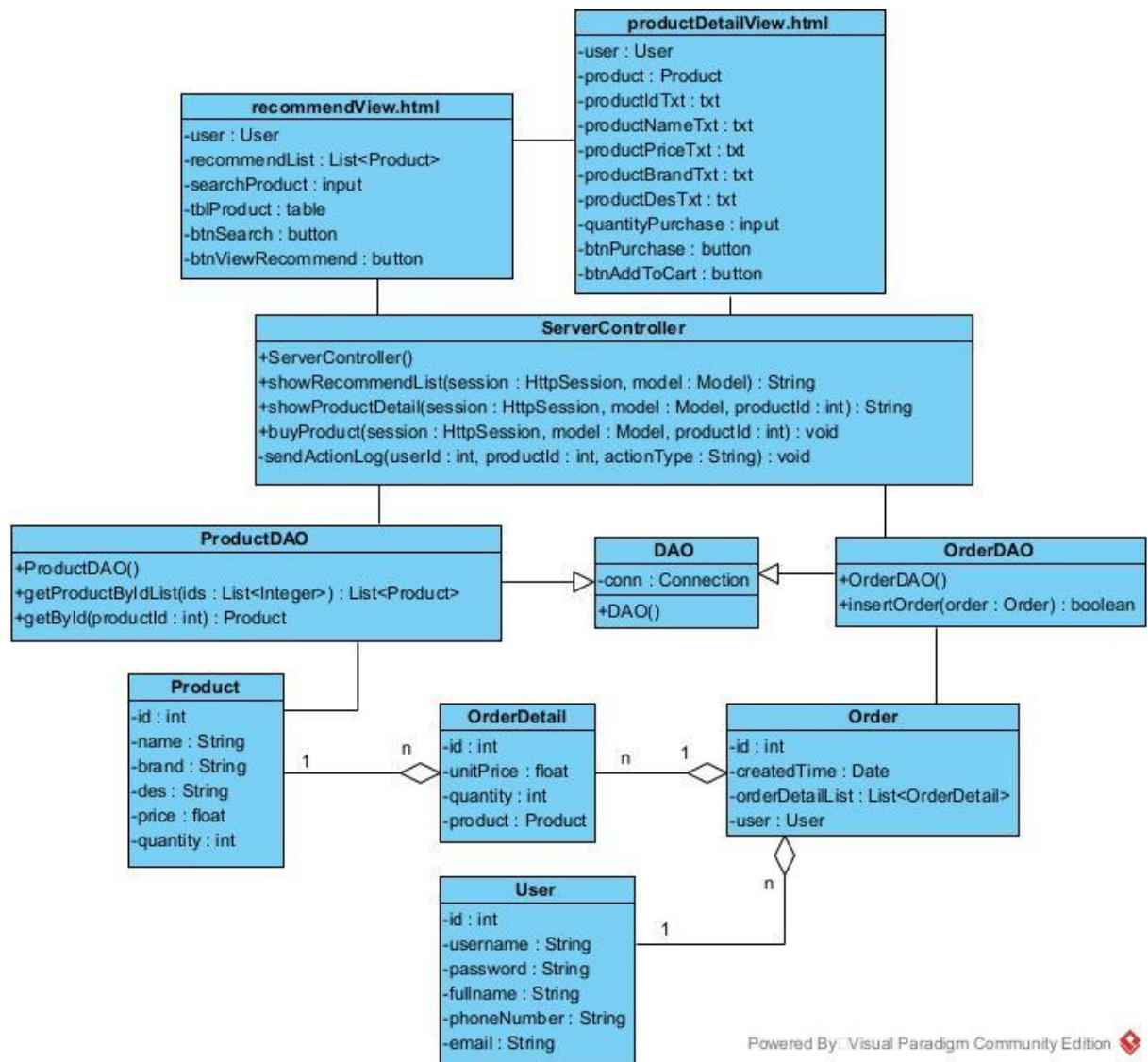
Input: Order

Output: boolean

=> Phương thức: `insertOrder(order: Order): boolean`

d. Các lớp thực thể

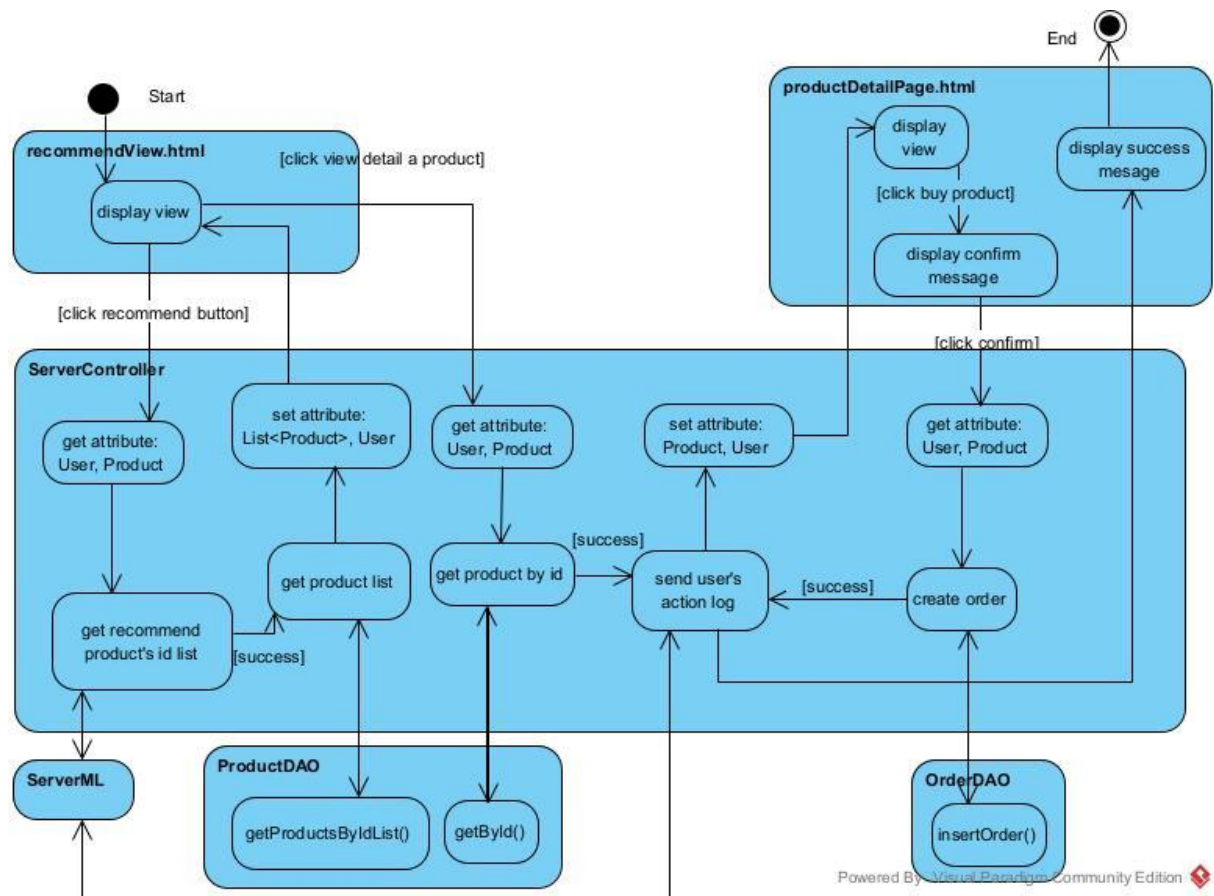
Các lớp thực thể tham gia bao gồm: User, Product, Order, OrderDetail



Hình 3.5: Biểu đồ lớp chi tiết

3.3.3 Thiết kế biểu đồ hoạt động và biểu đồ tuần tự

a. Biểu đồ hoạt động



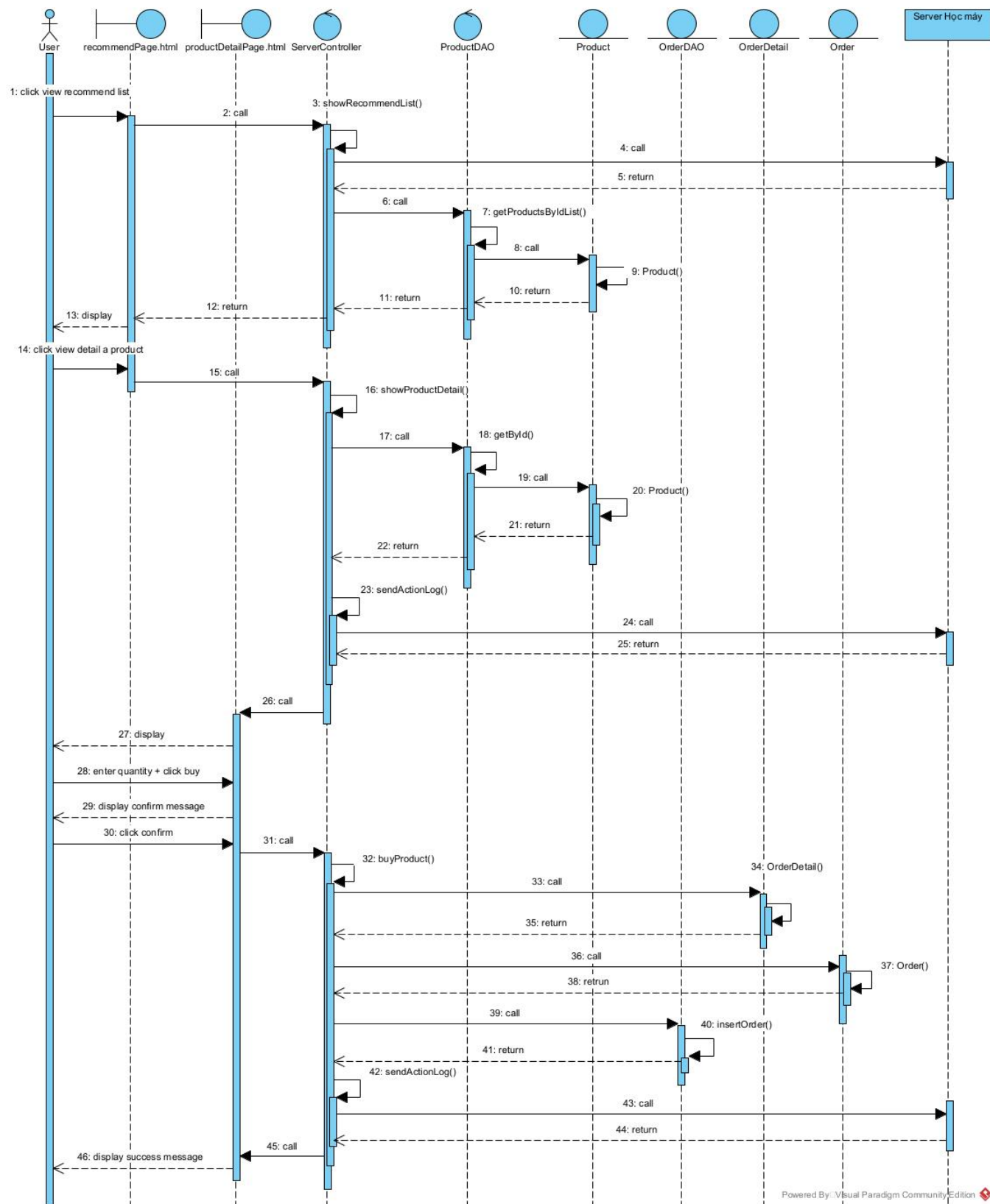
Hình 3.6: Biểu đồ hoạt động

Biểu đồ hoạt động được diễn giải như sau:

- + Lớp recommendView.html hiển thị giao diện cho người dùng sau khi đăng nhập
- + Người dùng click vào nút xem đề xuất
- + ServerController tiếp nhận yêu cầu và thực hiện chuỗi xử lý nghiệp vụ hiển thị danh sách sản phẩm gợi ý:
 - Nhận thông tin Người dùng từ session (phiên đăng nhập)
 - Gửi thông tin Người dùng về Server Học máy
 - Nhận danh sách mã sản phẩm từ Server Học máy
 - Gọi phương thức getProductByIdList() của lớp ProductDAO lấy danh sách sản phẩm gợi ý.
 - Lưu thông tin User, danh sách sản phẩm gợi ý vào model và trả về lớp recommendView.html
- + Lớp recommendView.html hiển thị kết quả cho người dùng

- + Người dùng click xem chi tiết một sản phẩm tại lớp recommendView.html
- + ServerController tiếp nhận yêu cầu và thực hiện chuỗi xử lý nghiệp vụ hiển thị chi tiết sản phẩm:
 - Nhận thông tin Người dùng từ session
 - Gọi phương thức getById() của ProductDAO để lấy thông tin sản phẩm.
 - Gửi log hành động “Xem sản phẩm” của người dùng về Server học máy
 - Lưu thông tin sản phẩm vào model và trả về lớp productDetailPage.html
- + Người dùng nhập số lượng và click mua sản phẩm
- + Lớp productDetailPage.html hiển thị yêu cầu xác nhận mua
- + Người dùng click xác nhận.
- + ServerController tiếp nhận yêu cầu và thực hiện chuỗi xử lý nghiệp vụ mua sản phẩm:
 - Nhận thông tin người dùng, sản phẩm từ session
 - Gọi hàm createOrder() của lớp OrderDAO để thêm Order.
 - Gửi log hành động “Xem sản phẩm” của người dùng về Server học máy
 - Trả về lớp productDetailPage.html
- + Lớp productDetailPage.html hiển thị thông báo mua hàng thành công cho người dùng.

b. Biểu đồ tuần tự



Hình 3.7. Biểu đồ tuần tự

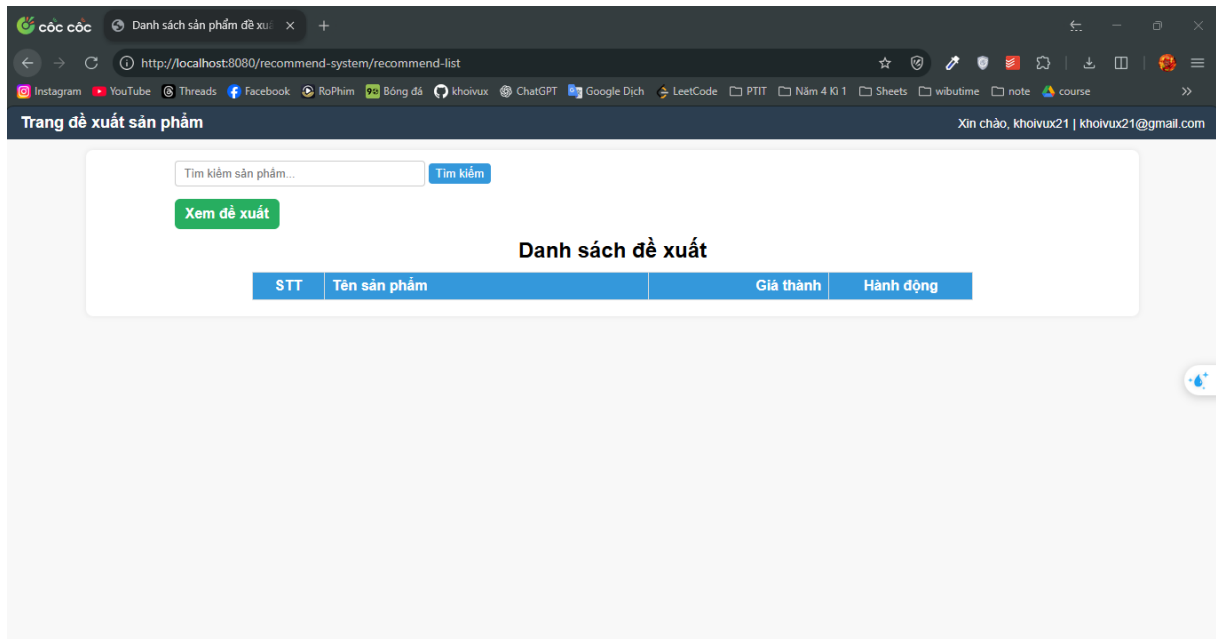
** Kịch bản biểu đồ tuần tự:*

1. Người dùng click nút xem danh sách đề xuất tại lớp recommendPage.html
2. Lớp recommendPage.html gọi đến lớp ServerController yêu cầu hiển thị danh sách đề xuất.
3. Hàm showRecommendList() của lớp ServerController được gọi
4. Hàm showRecommendList() gọi API của server học máy yêu cầu lấy danh sách sản phẩm đề xuất.
5. Server Học máy trả kết quả cho showRecommendList()
6. Hàm showRecommendList () gọi lớp ProductDAO yêu cầu lấy danh sách sản phẩm
7. Hàm getProductsByIdList() của lớp ProductDAO được gọi
8. Hàm getProductsByIdList() gọi Lớp Product để đóng gói thông tin
9. Lớp Product đóng gói thông tin thành đối tượng Product
10. Lớp Product trả về đối tượng cho hàm getProductsByIdList()
11. Hàm getProductsByIdList() trả kết quả về cho hàm showRecommendList()
12. Hàm showRecommendList() trả về lớp recommendPage.html
13. Lớp recommendPage.html hiển thị giao diện cho người dùng
14. Người dùng click xem chi tiết một sản phẩm
15. Lớp recommendPage.html gọi lớp ServerController yêu cầu hiển thị thông tin sản phẩm
16. Hàm showProductDetail() của lớp ServerController được gọi
17. Hàm showProductDetail() gọi lớp ProductDAO yêu cầu lấy thông tin sản phẩm
18. Hàm getById() của ProductDAO được gọi
19. Hàm getById() gọi lớp Product để đóng gói thông tin
20. Lớp Product đóng gói thông tin thành đối tượng Product
21. Lớp Product trả về đối tượng cho hàm getById ()
22. Hàm getById() trả về kết quả cho hàm showProductDetail()
23. Hàm showProductDetail() gọi hàm sendActionLog()
24. Hàm sendActionLog() gọi API của server học máy để gửi thông tin hành động của người dùng
25. Server học máy trả về kết quả
26. Hàm showProductDetail() gọi lớp productDetailPage.html
27. Lớp productDetailPage.html hiển thị giao diện cho người dùng.
28. Người dùng nhập số lượng và click nút Mua
29. Lớp productDetailPage.html hiển thị yêu cầu người dùng xác nhận

30. Người dùng click nút xác nhận
31. Lớp productDetailPage.html gọi lớp ServerControler
32. Hàm buyProduct() của lớp ServerControler được gọi
33. Hàm buyProduct() gọi lớp OrderDetail để đóng gói thông tin
34. Lớp OrderDetail đóng gói thông tin thành đối tượng OrderDetail
35. Lớp OrderDetail trả về đối tượng cho hàm buyProduct()
36. Hàm buyProduct() gọi lớp Order để đóng gói thông tin
37. Lớp Order đóng gói thông tin thành đối tượng Order
38. Lớp Order trả về đối tượng cho hàm buyProduct()
39. Hàm buyProduct gọi đến lớp OrderDAO
40. Hàm insertOrder() của lớp OrderDAO ddwoj gọi
41. Hàm insertOrder() trả về kết quả cho hàm buyProduct()
42. Hàm buyProduct () gọi hàm sendActionLog()
43. Hàm sendActionLog() gọi API của server học máy để gửi thông tin hành động của người dùng
44. Server học máy trả về kết quả
45. Hàm buyProduct() của lớp ServerController gọi lớp productDetail.html
46. Lớp productDetail.html hiển thị thông báo mua hàng thành công cho người dùng.

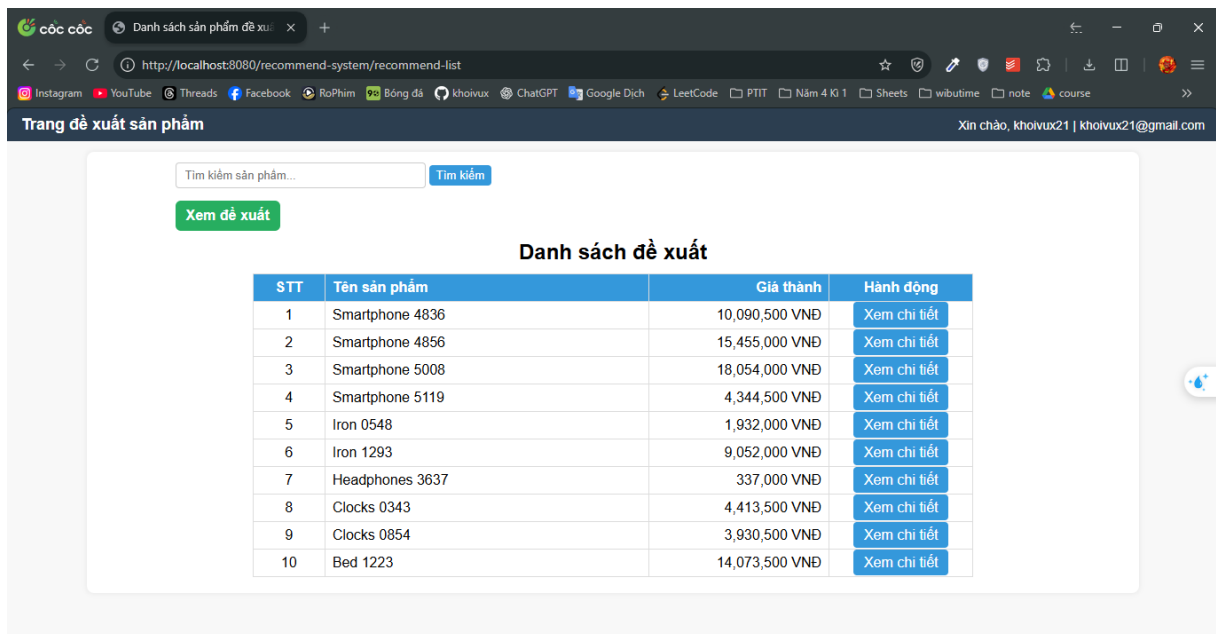
3.3.4 Kết quả chạy ứng dụng

Sau khi đăng nhập, người dùng sẽ truy cập vào trang đề xuất sản phẩm



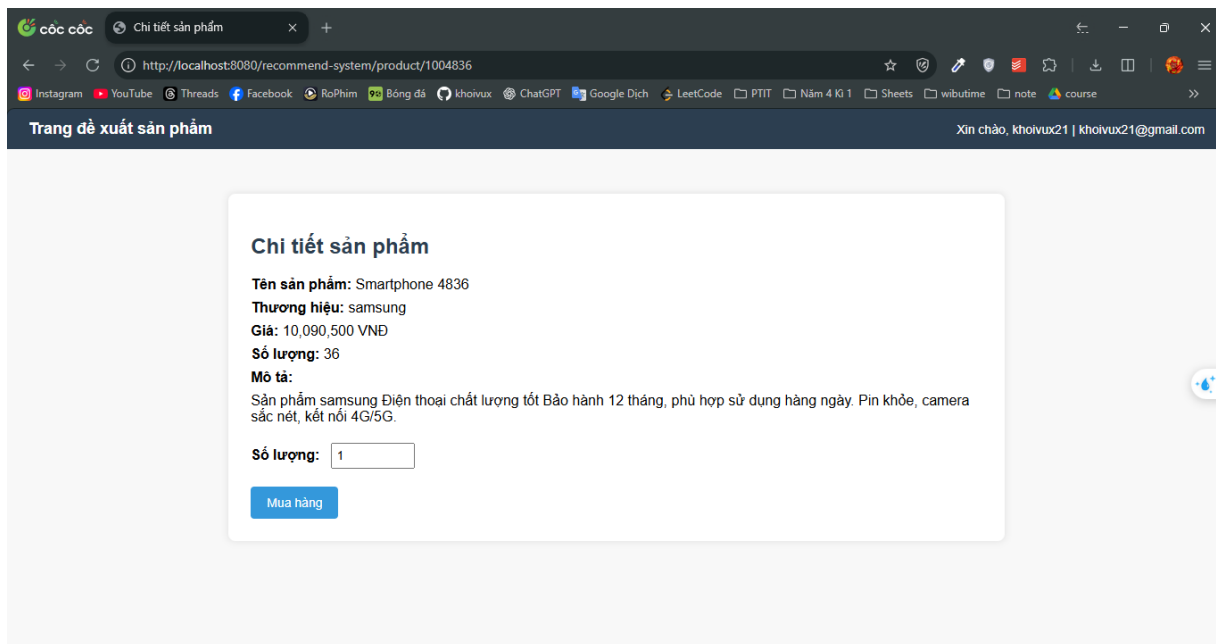
Hình 3.8: Kết quả giao diện trang đề xuất

Người dùng click vào nút xem đề xuất, giao diện sẽ hiển thị danh sách sản phẩm đề xuất



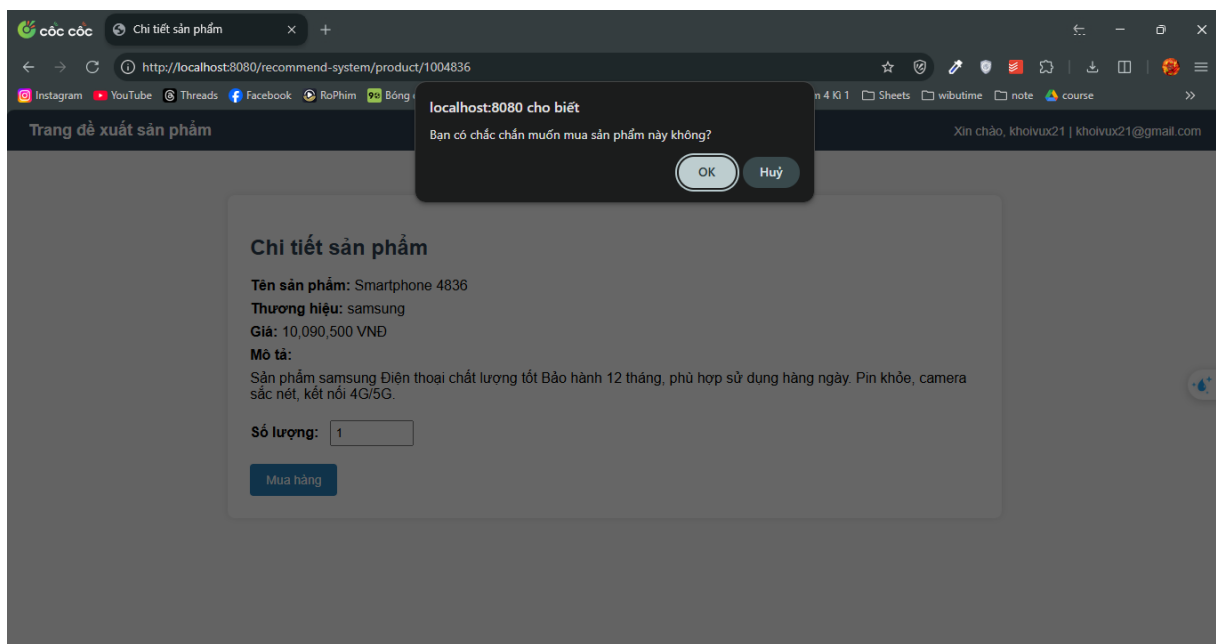
Hình 3.9: Kết quả hiển thị danh sách đề xuất

Người dùng click xem chi tiết sản phẩm



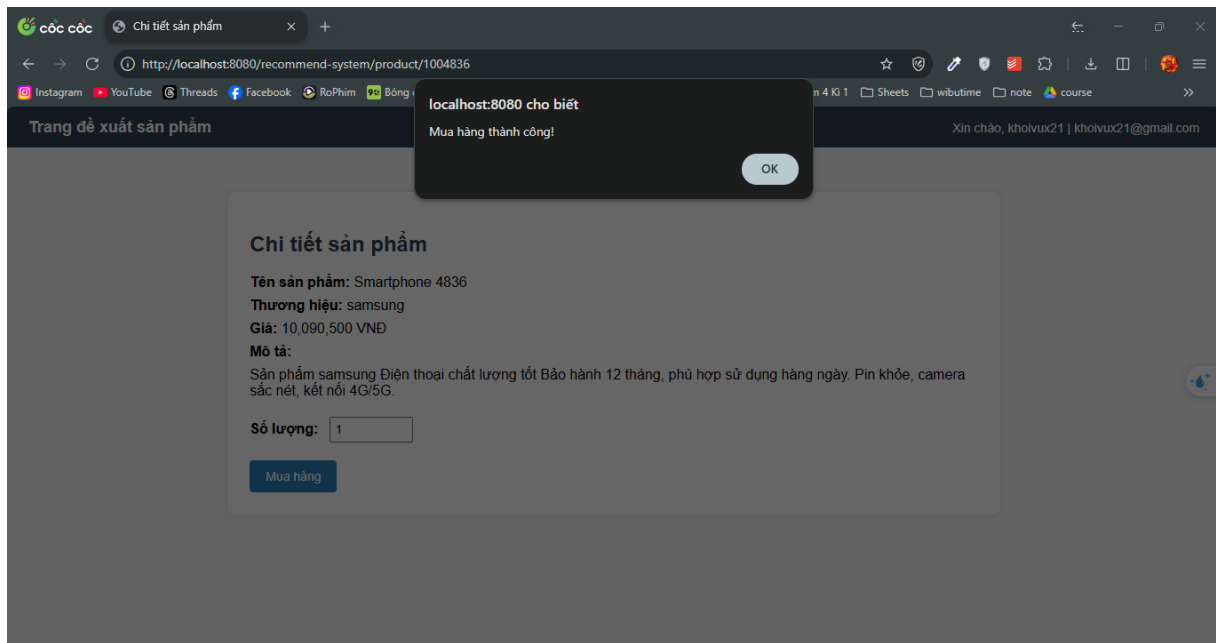
Hình 3.10: Kết quả hiển thị chi tiết sản phẩm

Người dùng click vào nút mua hàng, hiển thị giao diện xác nhận mua



Hình 3.11: Kết quả hiển thị yêu cầu xác nhận

Sau khi xác nhận mua hàng, giao diện hiển thị thông báo thành công



Hình 3.12: Kết quả hiển thị mua hàng thành công

Kết chương

Chương 3 đã trình bày quá trình thiết kế hệ thống của đề tài Hệ thống gợi ý mua hàng theo sở thích người dùng. Chương 3 cũng đã trình bày kết quả đạt được của hệ thống trên nền tảng website.

KẾT LUẬN

Bài báo cáo trên đã trình bày về quá trình nghiên cứu cũng như triển khai đề tài Hệ thống gợi ý mua hàng theo sở thích người dùng, áp dụng thuật toán NCF và triển khai trên nền tảng website. Các yêu cầu chính của bài toán như: thu thập dữ liệu hành vi người dùng, đánh giá, phân tích và đề xuất sản phẩm phù hợp với từng cá nhân đều đã được thực hiện và đáp ứng ở mức cơ bản về mặt kỹ thuật cũng như ứng dụng.

Bên cạnh kết quả đạt được, đề tài vẫn tồn tại một số hạn chế nhất định. Các chỉ số thực nghiệm đánh giá mô hình/ thuật toán chưa được cao. Giao diện trên nền tảng website chỉ nằm ở mức thử nghiệm, chưa thân thiện với người dùng. Bộ dữ liệu sản phẩm chưa được chi tiết, cụ thể, chỉ ở mức thử nghiệm. Hiện tại đề tài chỉ sử dụng thuật toán NCF với phương pháp gợi ý cộng tác, chưa đa dạng phương pháp gợi ý.

Với định hướng phát triển tương lai, đề tài sẽ phát triển hệ thống có giao diện thân thiện hơn với người dùng, có bộ dữ liệu sản phẩm tốt hơn, sát với thực tế cũng như có thể gợi ý được sản phẩm cho người dùng theo những phương pháp khác. Đề tài cũng hướng đến phát triển hệ thống ở nhiều nền tảng khác như mobile, winform,... cũng phát triển, cải thiện các tính năng như giỏ hàng, voucher,...

TÀI LIỆU THAM KHẢO

- [1] Medarametla, Suda & Thallapally, Glory. (2025). *Comparing K-Nearest Neighbors And Naive Bayes In Real-Time Recommendation Systems*. *Global Journal of Engineering Innovations and Interdisciplinary Research*, 5. <https://doi.org/10.33425/3066-1226.1075>
- [2] Muzaqi, Wildan & Rohman, M. & Reknadi, Danang. (2025). *Major Recommendation System for New Students at SMK Muhammadiyah 1 Lamongan with Naive Bayes Algorithm*. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 14, 357–364. <https://doi.org/10.32736/sisfokom.v14i3.2390>
- [3] Rrmoku, Korab & Selimi, Besnik & Ahmedi, Lule. (2022). *Application of Trust in Recommender Systems—Utilizing Naive Bayes Classifier*. *Computation*, 10(1), 6. <https://doi.org/10.3390/computation10010006>
- [4] Tuzhilin, Alexander. (2010). *Towards the Next Generation of Recommender Systems*. <https://doi.org/10.2991/icebi.2010.28>
- [5] GeeksforGeeks. (2025, October 24). *Recommender Systems using KNN*. *GeeksforGeeks*. <https://www.geeksforgeeks.org/machine-learning/recommender-systems-using-knn/>
- [6] GeeksforGeeks. (n.d.). *Neural Collaborative Filtering*. Retrieved October 26, 2025, from <https://www.geeksforgeeks.org/deep-learning/neural-collaborative-filtering>