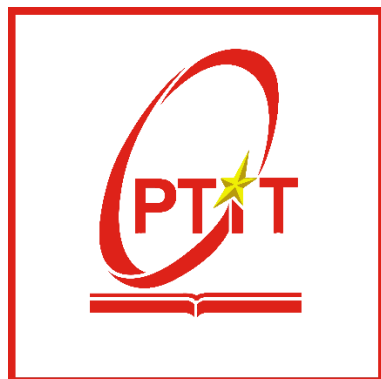


**BỘ THÔNG TIN VÀ TRUYỀN THÔNG  
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG  
CƠ SỞ TẠI THÀNH PHỐ HỒ CHÍ MINH**

=====Ω=====



**ĐỒ ÁN MÔN PHÁT TRIỂN CÁC HỆ THỐNG  
THÔNG MINH**

**Đề tài: Xây dựng hệ thống đề xuất sản phẩm  
cho website bán sách**

**Giảng viên hướng dẫn: ThS. Nguyễn Ngọc Duy**

**Sinh viên thực hiện:**

**Nguyễn Khang Hy - N21DCCN131 - D21CQCNPM02**

**HKII – Năm học 2024 - 2025**

## MỤC LỤC

<b>CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI.....</b>	<b>3</b>
1.1. Lý do chọn đề tài .....	3
1.2. Mục tiêu nghiên cứu .....	3
1.3. Đối tượng và phạm vi nghiên cứu .....	3
<b>CHƯƠNG 2: CƠ SỞ LÝ THUYẾT.....</b>	<b>4</b>
2.1. Hệ thống thông minh (Intelligent Systems) .....	4
2.2. Học máy.....	5
2.3. Xây dựng và triển khai hệ thống thông minh .....	6
2.4. Cơ sở dữ liệu.....	7
2.5. Hệ thống đề xuất.....	8
2.6. Xử lý ngôn ngữ tự nhiên.....	9
2.7. Thuật toán kNN .....	9
2.8. Công nghệ sử dụng.....	12
<b>CHƯƠNG 3: PHÂN TÍCH THIẾT KẾ HỆ THỐNG.....</b>	<b>16</b>
3.1. Công nghệ và tài nguyên sử dụng: .....	16
3.2. Đặc tả yêu cầu.....	16
3.3. Sơ đồ ca sử dụng mức tổng quát .....	17
3.4. Cơ sở dữ liệu.....	18
<b>CHƯƠNG 4: XÂY DỰNG HỆ THỐNG .....</b>	<b>30</b>
4.1. Hệ thống đề xuất sản phẩm thông minh .....	30
4.2. Giao diện của khách hàng.....	38
4.3. Giao diện của nhân viên .....	43
<b>CHƯƠNG 5: DEMO .....</b>	<b>46</b>
<b>CHƯƠNG 6: KẾT LUẬN.....</b>	<b>49</b>
6.1. Đánh giá.....	49
6.2. Kết luận.....	49

# CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI

## 1.1. Lý do chọn đề tài

Trong những năm trở lại đây, với sự phát triển mạnh mẽ của công nghệ thông tin, mọi ngành kinh doanh đều phát triển và giải quyết được rất nhiều khó khăn, việc kinh doanh từ đó cũng trở nên dễ dàng và thuận lợi hơn. Từ đó thương mại điện tử là một trong những xu hướng tất yếu của nền kinh tế hiện đại. Với sự hỗ trợ của công nghệ ngày càng phát triển, sự bùng nổ của các thiết bị kết nối internet, các doanh nghiệp ngày càng có cơ hội mở rộng kênh bán hàng, tiếp cận gần hơn với người tiêu dùng. Các website ra đời đã góp phần quan trọng to lớn trong việc thúc đẩy kinh doanh phát triển nhanh chóng.

Trong bối cảnh thị trường thương mại điện tử ngày càng phát triển, các website bán hàng trực tuyến không chỉ cần cung cấp sản phẩm đa dạng mà còn phải đáp ứng tốt nhu cầu cá nhân hóa trải nghiệm mua sắm của người dùng. Đặc biệt, trong lĩnh vực kinh doanh sách, việc xây dựng một hệ thống đề xuất sản phẩm hiệu quả là yếu tố quan trọng giúp cải thiện sự hài lòng của khách hàng, tăng doanh số bán hàng và tạo lợi thế cạnh tranh trên thị trường.

Lợi ích của hệ thống đề xuất sản phẩm không chỉ dừng lại ở việc gia tăng tỷ lệ chuyển đổi mua hàng mà còn tạo sự gắn kết lâu dài với khách hàng. Khi người dùng cảm thấy được "hiểu" và nhận được những gợi ý thực sự phù hợp, họ có xu hướng quay lại website thường xuyên hơn và dành nhiều thời gian hơn để khám phá sản phẩm.

Đề tài “Xây dựng hệ thống đề xuất sản phẩm cho website bán sách” tập trung nghiên cứu và phát triển một hệ thống đề xuất sản phẩm có khả năng gợi ý sản phẩm cho người dùng một cách thông minh. Hệ thống này áp dụng thuật toán học máy (machine learning) để xử lý yêu cầu của khách hàng và đề xuất sản phẩm phù hợp.

Xây dựng hệ thống đề xuất sản phẩm cho website bán sách không chỉ là một bài toán kỹ thuật mà còn là chìa khóa để nâng cao chất lượng dịch vụ, tăng trưởng doanh thu và tạo sự khác biệt trong môi trường cạnh tranh khốc liệt của thương mại điện tử hiện nay. Chính vì vậy, em đã chọn đề tài này nhằm mục đích nghiên cứu, ứng dụng công nghệ hiện đại vào thực tiễn cũng như nâng cao hiểu biết của bản thân về học máy và các công nghệ thông minh.

## 1.2. Mục tiêu nghiên cứu

Nghiên cứu và xây dựng, triển khai hệ thống đề xuất sản phẩm cho website bán sách.

## 1.3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu: Hệ thống đề xuất sản phẩm cho website bán sách.

Phạm vi nghiên cứu: Tìm hiểu lý thuyết về xây dựng phát triển hệ thống thông minh, hệ thống đề xuất, xử lý ngôn ngữ tự nhiên, thuật toán KNN, ứng dụng vào đề tài “Xây dựng hệ thống đề xuất sản phẩm cho website bán sách”.

## CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

### 2.1. Hệ thống thông minh (Intelligent Systems)

Khái niệm: Hệ thống thông minh là:

- Những hệ thống có công nghệ tiên tiến nhận thức và phản ứng với thế giới xung quanh. Nó kết nối người dùng với trí tuệ nhân tạo để đạt được mục tiêu có ý nghĩa.
- Hệ thống trong đó trí thông minh phát triển, cải thiện theo thời gian. Sự thông minh được cải thiện bằng cách xem cách người dùng tương tác với hệ thống.

Hệ thống thông minh ra đời dựa trên các nhu cầu:

- Yêu cầu chức năng:
  - Xử lý thông tin: Khả năng thu thập, lưu trữ và xử lý dữ liệu từ nhiều nguồn khác nhau.
  - Học hỏi: Khả năng phân tích dữ liệu và cải thiện hiệu suất thông qua học máy hoặc học sâu.
  - Tự động hóa: Hỗ trợ hoặc thay thế con người trong việc thực hiện các nhiệm vụ phức tạp.
- Yêu cầu hiệu suất:
  - Độ chính xác: Mức độ phù hợp với ứng dụng cụ thể, ví dụ: chẩn đoán y khoa đòi hỏi độ chính xác cao.
  - Tốc độ xử lý: Thời gian thực hiện nhiệm vụ cần nhanh để đáp ứng nhu cầu sử dụng thực tế.
  - Tính sẵn sàng: Hệ thống cần luôn hoạt động ổn định, ngay cả trong điều kiện áp lực cao.
- Yêu cầu an ninh và quyền riêng tư:
  - Bảo mật: Đảm bảo an toàn cho dữ liệu, chống lại các tấn công mạng hoặc truy cập trái phép.
  - Quyền riêng tư: Bảo vệ thông tin cá nhân của người dùng và tuân thủ các quy định pháp luật.
- Yêu cầu đạo đức và xã hội:
  - Đạo đức: Hệ thống phải được thiết kế để hành xử có đạo đức, tránh phân biệt đối xử hoặc gây hại.
  - Tương tác xã hội: Hỗ trợ sự tương tác với con người theo cách tự nhiên và thân thiện.
- Yêu cầu kỹ thuật:
  - Cơ sở hạ tầng: Đòi hỏi phần cứng và phần mềm hiện đại để xử lý dữ liệu lớn và phức tạp.
  - Khả năng tích hợp: Dễ dàng tích hợp với các hệ thống khác hoặc nâng cấp khi cần.

- Yêu cầu ngữ cảnh và ứng dụng cụ thể:
  - Mục tiêu cụ thể: Phục vụ cho một hoặc nhiều mục tiêu rõ ràng như: chẩn đoán bệnh, tối ưu hóa sản xuất, hỗ trợ học tập, v.v.
  - Tính thích nghi: Có khả năng điều chỉnh và hoạt động hiệu quả trong các môi trường khác nhau.

Ứng dụng của hệ thống thông minh:

- Các hệ thống khuyến nghị và cảnh báo: gợi ý sản phẩm, cảnh báo thiên tai,...
- Y tế: chẩn đoán bệnh, hỗ trợ điều trị,...
- Giáo dục: cá nhân hóa học tập, hỗ trợ giáo viên chấm điểm, soạn đề thi,...
- Tài chính: tư vấn đầu tư, phát hiện gian lận,...

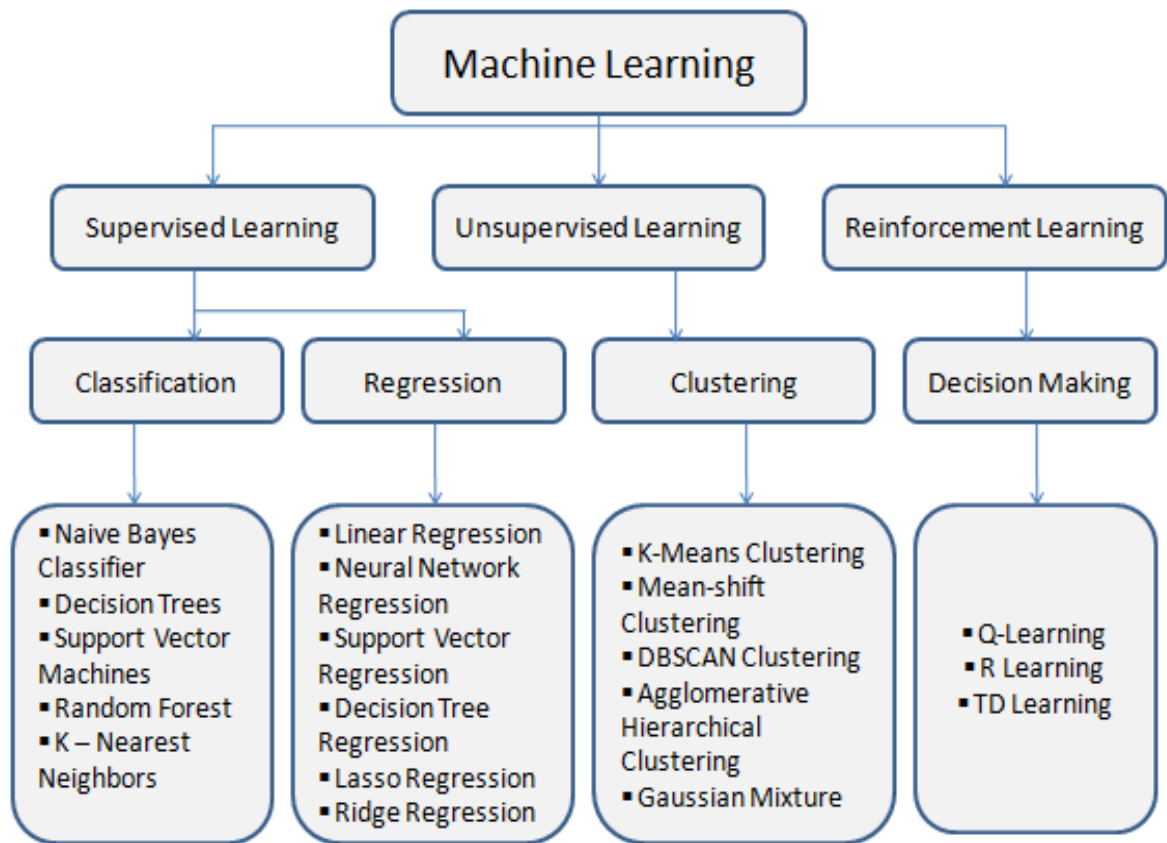
## 2.2. Học máy

Học máy (Machine Learning – ML) là khả năng của chương trình máy tính sử dụng kinh nghiệm, quan sát, hoặc dữ liệu trong quá khứ để cải thiện công việc của mình trong tương lai. Học máy khắc phục vấn đề không linh hoạt của suy diễn. Là một nhánh nghiên cứu rất quan trọng của trí tuệ nhân tạo, sử dụng kinh nghiệm thông qua quá trình học. Là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc nghiên cứu và xây dựng các kỹ thuật cho phép các hệ thống "học" tự động từ dữ liệu để giải quyết những vấn đề cụ thể. Các thuật toán học máy xây dựng một mô hình dựa trên dữ liệu mẫu, được gọi là dữ liệu huấn luyện, để đưa ra dự đoán hoặc quyết định mà không cần được lập trình chi tiết về việc đưa ra dự đoán hoặc quyết định này.

Học máy có hiện nay được áp dụng rộng rãi bao gồm máy truy tìm dữ liệu, chẩn đoán y khoa, phát hiện thẻ tín dụng giả, phân tích thị trường chứng khoán, phân loại các chuỗi DNA, nhận dạng tiếng nói và chữ viết, dịch tự động, chơi trò chơi và cử động robot (robot locomotion).

Học máy có hai dạng là: Học máy có giám sát và học máy không có giám sát.

- Học máy có giám sát (Supervised Learning) là một kỹ thuật của ngành học máy nhằm mục đích xây dựng một hàm  $f$  từ dữ tập dữ liệu huấn luyện (Training data). Dữ liệu huấn luyện bao gồm các cặp đối tượng đầu vào và đầu ra mong muốn. Đầu ra của hàm  $f$  có thể là một giá trị liên tục hoặc có thể là dự đoán một nhãn phân lớp cho một đối tượng đầu vào.
- Học máy không có giám sát (Unsupervised Learning) là nhóm thuật toán hay phương pháp kỹ thuật cho phép máy tự học hỏi và tìm ra một mô hình hay cấu trúc nào đó ẩn trong một bộ dữ liệu không được gắn nhãn trước.



Hình 2.1. Mô hình các thuật toán trong học máy

### 2.3. Xây dựng và triển khai hệ thống thông minh

- Quy trình xây dựng: Xây dựng hệ thống thông minh là một quá trình phức tạp và đòi hỏi sự cân nhắc kỹ lưỡng. Quy trình này thường bao gồm các bước sau:

- Thu thập và xử lý dữ liệu: Thu thập dữ liệu từ nhiều nguồn khác nhau, sau đó xử lý để đảm bảo chất lượng, độ tin cậy, và tính phù hợp cho hệ thống.
- Phân tích và thiết kế: Xác định mục tiêu, các yêu cầu chức năng, và kiến trúc hệ thống. Lựa chọn các thuật toán và mô hình phù hợp.
- Huấn luyện và kiểm tra: Sử dụng dữ liệu đã xử lý để huấn luyện hệ thống. Kiểm tra hiệu quả và tối ưu hóa các mô hình.
- Đánh giá và cải thiện: Đánh giá kết quả, thực hiện cải tiến dựa trên phản hồi từ các thử nghiệm thực tế hoặc môi trường giả lập.

- Triển khai hệ thống:

- Triển khai hệ thống thông minh là giai đoạn cuối cùng trong chu trình phát triển. Quá trình này bao gồm việc tích hợp hệ thống vào môi trường thực tế, đảm bảo khả năng tương thích và tối ưu hóa hoạt động. Ngoài ra, cần thiết lập cơ chế giám sát, bảo trì và cập nhật để đảm bảo hệ thống hoạt động hiệu quả lâu dài.

- Tổ chức hệ thống:

- Tổ chức hệ thống thông minh liên quan đến cách các thành phần của hệ thống hoạt động cùng nhau một cách tối ưu.
- Các yếu tố bao gồm: hệ thống lưu trữ dữ liệu, các mô-đun xử lý, giao diện người dùng, và các API kết nối với hệ thống khác.
- Cần đảm bảo rằng hệ thống có thể mở rộng (scalability), linh hoạt (flexibility), và đáng tin cậy (reliability).

- Vấn đề trade-offs trong xây dựng hệ thống thông minh:

- Trong quá trình xây dựng hệ thống thông minh, thường phải đối mặt với các vấn đề đánh đổi (trade-offs) như trade-off giữa độ chính xác và tốc độ. Đôi khi, bạn phải hy sinh độ chính xác để đạt tốc độ xử lý nhanh hơn hoặc ngược lại, tùy thuộc vào mục tiêu sử dụng.
- Trade-off giữa tiêu chuẩn đạo đức và hiệu suất: Cần cân nhắc giữa việc sử dụng dữ liệu có thể gây ra vấn đề về quyền riêng tư và đạo đức với việc đạt được hiệu suất tốt hơn.

- Đạo đức trong xây dựng hệ thống thông minh:

- Việc xây dựng hệ thống thông minh đòi hỏi tuân thủ nguyên tắc đạo đức. Điều này bao gồm việc đảm bảo rằng hệ thống không gây hại cho cá nhân hoặc xã hội, đảm bảo quyền riêng tư của người dùng, và tránh thiên vị (bias) không công bằng trong mô hình hoặc quyết định của hệ thống. Các vấn đề về đạo đức trong AI và hệ thống thông minh đang trở thành một chủ đề quan trọng trong cộng đồng nghiên cứu và công nghiệp.

## 2.4. Cơ sở dữ liệu

Cơ sở dữ liệu (database) là một tập hợp các dữ liệu rời rạc được tổ chức, lưu trữ và quản lý theo hệ thống để có thể dễ dàng truy xuất, chỉnh sửa và cập nhật. Chúng thường được sử dụng để lưu trữ thông tin trong các hệ thống phần mềm và ứng dụng, từ dữ liệu cá nhân cho đến thông tin doanh nghiệp quan trọng. Cơ sở dữ liệu giúp các tổ chức và cá nhân tổ chức dữ liệu một cách có cấu trúc, nhằm mục đích quản lý và xử lý dữ liệu hiệu quả hơn.

Ví dụ: Trong cuộc sống bạn có thể thấy rất nhiều cơ sở dữ liệu đó, nếu bạn vào thư viện thì ở trên đó là một cơ sở dữ liệu, phân loại sách theo chủ đề: xã hội, tự nhiên, khoa học ... đó cũng là một dạng cơ sở dữ liệu đơn giản.

Mục đích ta sử dụng CSDL là vì cho dễ quản lý và tìm kiếm, thống nhất việc lưu trữ thông tin, có thể tìm kiếm một cách nhanh chóng khi cần thông tin nào đó.

Cơ sở dữ liệu (CSDL) có nhiều đặc điểm quan trọng giúp quản lý, tổ chức và xử lý dữ liệu một cách hiệu quả. Những đặc điểm này đảm bảo rằng dữ liệu được lưu trữ, truy xuất và bảo mật một cách tối ưu, đồng thời hỗ trợ các ứng dụng và hệ thống trong việc quản lý thông tin. Dưới đây là các đặc điểm chính của cơ sở dữ liệu:

- Tính tổ chức và cấu trúc: CSDL tổ chức dữ liệu theo một cấu trúc có hệ thống, thường được thiết kế dưới dạng bảng, đối tượng hoặc mô hình dữ liệu khác. Cấu trúc này giúp dữ liệu được tổ chức một cách hợp lý, dễ dàng truy cập và quản lý.
- Khả năng truy xuất và tìm kiếm: CSDL cho phép người dùng truy xuất và tìm kiếm dữ liệu nhanh chóng và hiệu quả thông qua các truy vấn. Các hệ quản trị cơ sở dữ liệu (DBMS) cung cấp ngôn ngữ truy vấn, chẳng hạn như SQL, để thực hiện các thao tác này.
- Tính nhất quán và toàn vẹn: CSDL đảm bảo rằng dữ liệu luôn được duy trì một cách chính xác và nhất quán thông qua các ràng buộc toàn vẹn. Các quy tắc này bao gồm khóa chính, khóa ngoại và các hạn chế khác để bảo vệ dữ liệu khỏi lỗi và mất mát.
- Khả năng mở rộng: CSDL có thể mở rộng để xử lý khối lượng dữ liệu ngày càng tăng và đáp ứng nhu cầu của ứng dụng. Điều này có thể được thực hiện thông qua việc thêm phần cứng hoặc sử dụng các kỹ thuật phân tán và phân mảnh dữ liệu.
- Bảo mật và quyền truy cập: CSDL cung cấp các cơ chế bảo mật để bảo vệ dữ liệu khỏi các truy cập trái phép. Điều này bao gồm việc thiết lập quyền truy cập, mã hóa dữ liệu và theo dõi các hoạt động của người dùng.
- Khả năng sao lưu và phục hồi: CSDL hỗ trợ sao lưu dữ liệu định kỳ và phục hồi dữ liệu khi xảy ra sự cố. Các chức năng này giúp đảm bảo rằng dữ liệu có thể được khôi phục về trạng thái trước đó trong trường hợp mất mát hoặc hỏng hóc.
- Quản lý giao dịch: CSDL hỗ trợ quản lý giao dịch, đảm bảo rằng các thao tác trên dữ liệu được thực hiện một cách nguyên tử, nhất quán, tách biệt và bền vững. Điều này giúp duy trì tính toàn vẹn của dữ liệu ngay cả khi xảy ra lỗi hệ thống.
- Khả năng hỗ trợ đồng thời: CSDL có khả năng xử lý nhiều yêu cầu và giao dịch đồng thời từ nhiều người dùng mà không làm giảm hiệu suất hoặc làm mất dữ liệu. Điều này được thực hiện thông qua các cơ chế khóa và quản lý đồng thời.
- Tính tương thích: CSDL có thể tương thích với các ứng dụng và hệ thống khác thông qua các giao thức và giao diện chuẩn. Điều này cho phép tích hợp dữ liệu từ nhiều nguồn khác nhau và hỗ trợ các quy trình doanh nghiệp.

Những đặc điểm này giúp cơ sở dữ liệu đáp ứng nhu cầu quản lý dữ liệu trong các ứng dụng và hệ thống đa dạng, từ các hệ thống nhỏ lẻ đến các ứng dụng quy mô lớn và phức tạp.

## 2.5. Hệ thống đề xuất

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

Hệ thống đề xuất sử dụng các tri thức về sản phẩm, các tri thức của chuyên gia hay tri thức khai phá học được từ hành vi con người dùng để đưa ra các gợi ý về sản phẩm mà họ thích trong hàng ngàn hàng vạn sản phẩm có trong hệ thống. Các website thương mại điện tử, ví dụ như sách, phim, nhạc, báo... sử dụng hệ thống gợi ý để cung cấp các thông tin giúp cho người sử dụng quyết định sẽ lựa chọn sản phẩm nào. Các sản phẩm được gợi ý dựa trên số lượng sản phẩm đó đã được bán, dựa trên các thông tin cá nhân của



người sử dụng, dựa trên sự phân tích hành vi mua hàng trước đó của người sử dụng để đưa ra các dự đoán về hành vi mua hàng trong tương lai của chính khách hàng đó. Các dạng gợi ý bao gồm: gợi ý các sản phẩm tới người tiêu dùng, các thông tin sản phẩm mang tính cá nhân hóa, tổng kết các ý kiến cộng đồng, và cung cấp các chia sẻ, các phê bình, đánh giá mang tính cộng đồng liên quan tới yêu cầu, mục đích của người sử dụng đó.

## 2.6. Xử lý ngôn ngữ tự nhiên

Xử lý ngôn ngữ tự nhiên hay còn gọi là Natural Language Processing (NLP) là một lĩnh vực trong trí tuệ nhân tạo (AI) và khoa học máy tính. Chúng tập trung vào việc phát triển các phương pháp và công nghệ để máy tính có thể hiểu, phân tích và tạo ra ngôn ngữ tự nhiên của con người. Nói cách khác, NLP cho phép máy móc "hiểu" và "sử dụng" ngôn ngữ của chúng ta.

Cụ thể, Natural Language Processing có các nhiệm vụ như:

- Phân tích và hiểu văn bản: NLP giúp máy tính hiểu ý nghĩa của văn bản, bao gồm việc phân tích cú pháp (syntax), phân loại từ loại (POS tagging) và dự đoán ngữ cảnh (context prediction).
- Phân tích cảm xúc và ý kiến: NLP có thể phân tích cảm xúc và ý kiến trong văn bản, giúp đánh giá sự phản hồi của người dùng hoặc phản ứng của công chúng đối với các sự kiện hay sản phẩm.
- Dịch máy: Natural Language Processing giúp máy tính dịch từ ngôn ngữ này sang ngôn ngữ khác một cách tự động nhờ vào các mô hình dịch máy dựa trên học máy và deep learning.
- Tạo ra văn bản tự động: NLP có thể được sử dụng để tạo ra văn bản tự động, từ việc viết bài báo đến tạo ra nội dung cho website hoặc ứng dụng.
- Tương tác người - máy: NLP là cơ sở cho các ứng dụng tương tác người - máy. Có thể kể đến như chatbot, trợ lý ảo và hệ thống trả lời tự động.
- Phân tích và trích xuất thông tin: NLP giúp trích xuất thông tin từ văn bản, từ việc tìm kiếm thông tin đến phân tích dữ liệu trong các tài liệu lớn.

Xử lý ngôn ngữ tự nhiên Natural Language Processing hoạt động như sau:

- Thu thập dữ liệu: Dữ liệu văn bản hoặc giọng nói được thu thập từ nhiều nguồn khác nhau như sách, báo chí, mạng xã hội, email, ghi âm,....
- Xử lý dữ liệu: Dữ liệu được chia nhỏ thành các đơn vị nhỏ hơn (như từ, cụm từ) và được phân tích bằng các kỹ thuật học máy và trí tuệ nhân tạo.
- Phân tích dữ liệu: Hệ thống NLP sẽ trích xuất thông tin, xác định ý nghĩa và ngữ cảnh của dữ liệu.
- Tạo ra ngôn ngữ: Hệ thống NLP có thể sử dụng thông tin đã học được để tạo ra các văn bản mới, dịch ngôn ngữ hoặc trả lời câu hỏi một cách tự nhiên.

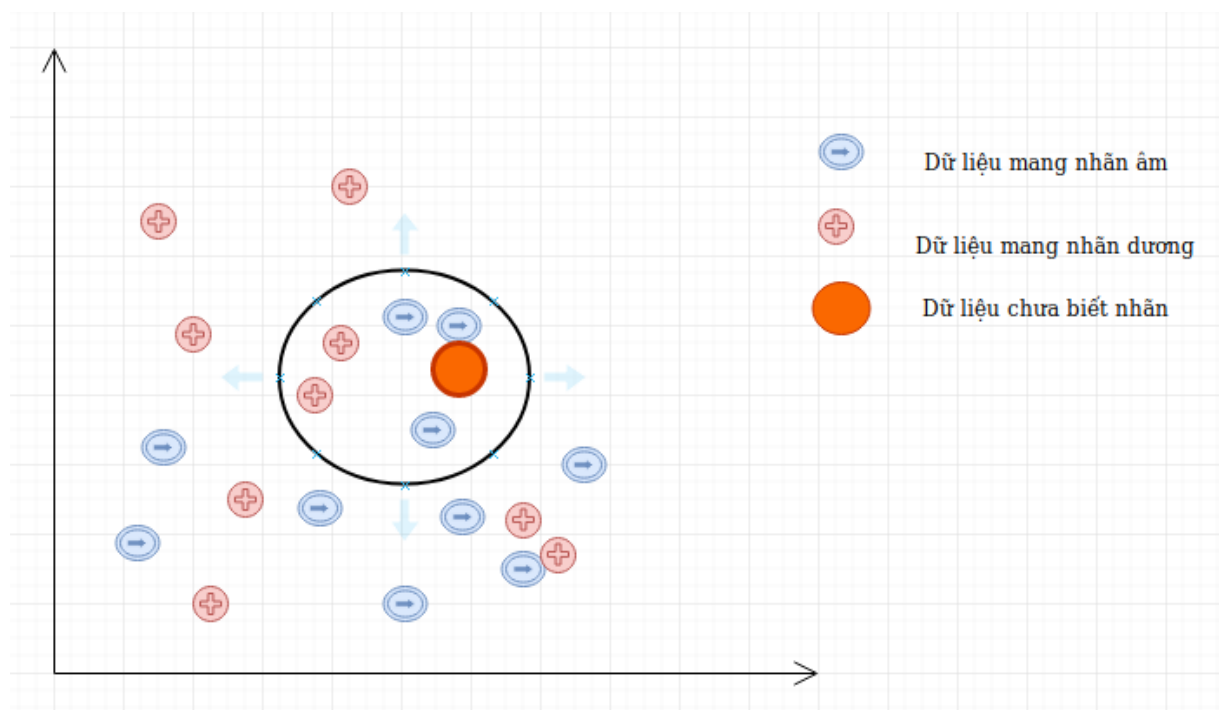
## 2.7. Thuật toán kNN

K-Nearest Neighbors (kNN) là một thuật toán học máy không giám sát (unsupervised) được sử dụng cho phân loại và dự đoán dựa trên dữ liệu. Thuật toán này hoạt động trên

nguyên tắc "người hàng xóm gần nhất" – nó dựa trên việc so sánh một điểm dữ liệu mới với K điểm dữ liệu gần nhất trong tập dữ liệu đã biết và dựa vào đó để xác định lớp hoặc giá trị dự đoán cho điểm dữ liệu mới. K là một tham số được thiết lập trước (ví dụ: K=3, K=5) và quyết định số lượng điểm gần nhất sẽ được sử dụng trong quá trình dự đoán.

Ví dụ:

- Giả sử ta có D là tập các dữ liệu đã được phân loại thành 2 nhãn (+) và (–) được biểu diễn trên trục tọa độ như hình vẽ và một điểm dữ liệu mới A chưa biết nhãn. Vậy làm cách nào để chúng ta có thể xác định được nhãn của A là (+) hay (–) ?
- Có thể thấy cách đơn giản nhất là so sánh tất cả các đặc điểm của dữ liệu A với tất cả tập dữ liệu học đã được gán nhãn và xem nó giống cái nào nhất, nếu dữ liệu (đặc điểm) của A giống với dữ liệu của điểm mang nhãn (+) thì điểm A mang nhãn (+), nếu dữ liệu A giống với dữ liệu nhãn (–) hơn thì nó mang nhãn (–), trông có vẻ rất đơn giản nhưng đó là những gì mà KNN làm.
- Trong trường hợp của KNN, thực tế nó không so sánh dữ liệu mới (không được phân lớp) với tất cả các dữ liệu khác, thực tế nó thực hiện một phép tính toán học để đo khoảng cách giữa dữ liệu mới với tất cả các điểm trong tập dữ liệu học D để thực hiện phân lớp. Phép tính khoảng cách giữa 2 điểm có thể là Euclidian, Manhattan, trọng số, Minkowski, ...



Hình 2.2. Mô hình ví dụ kNN

- Thuật toán kNN đối với việc tính toán khoảng cách giữa hai vector được thực hiện thông qua các bước sau:
- Xét hai đối tượng dữ liệu,  $r_i$  và  $r_j$ , mỗi đối tượng có  $n$  thuộc tính. Đối tượng  $r_i$  được biểu diễn bằng vector  $r_i = (x_{i1}, x_{i2}, \dots, x_{in})$ , trong khi đối tượng  $r_j$  được biểu diễn bằng vector  $r_j = (x_{j1}, x_{j2}, \dots, x_{jn})$ , trong đó  $n$  là số lượng thuộc tính của mỗi đối tượng.

- Để tính khoảng cách giữa hai đối tượng  $r_i$  và  $r_j$ , chúng ta có thể sử dụng các phương pháp tính toán khoảng cách phù hợp, như khoảng cách Euclidean, khoảng cách Manhattan hoặc khoảng cách Cosine.
- Công thức tính khoảng cách Euclidean:

$$d(r_i, r_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2}$$

- Công thức tính khoảng cách Manhattan:

$$d(r_i, r_j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{in} - x_{jn}|$$

Thuật toán của kNN có thể được mô tả như sau:

- Xác định giá trị của  $k$ , tức là số lượng láng giềng gần nhất sẽ được xem xét khi phân lớp.
- Tính toán khoảng cách giữa mẫu cần phân lớp (được gọi là điểm truy vấn) và tất cả các mẫu trong tập dữ liệu huấn luyện.
- Sắp xếp các khoảng cách theo thứ tự tăng dần và xác định  $k$  láng giềng gần nhất với điểm truy vấn.
- Thu thập các lớp của  $k$  láng giềng gần nhất đã xác định.
- Dựa vào phần lớn lớp trong tập hợp các láng giềng gần nhất để xác định lớp cho điểm truy vấn. Lớp này được coi là dự đoán cho điểm đó.

Một câu hỏi đặt ra đó là có phải cứ chọn  $K$  càng lớn thì càng tốt, thì câu trả lời đó là còn tùy thuộc vào dữ liệu đó như thế nào. Không phải lúc nào  $K$  càng lớn thì cho kết quả tốt và ngược lại. Việc lựa chọn tham số  $K$  của mô hình sẽ tiến hành thông qua thực nghiệm nhiều lần để chọn ra kết quả tốt nhất.

Ưu điểm của thuật toán kNN:

- Dễ triển khai: Thuật toán kNN có độ phức tạp không cao, dễ dàng thực hiện.
- Dễ thích nghi: Theo cách hoạt động của thuật toán kNN, nó lưu trữ tất cả dữ liệu trong bộ nhớ và do đó khi một ví dụ hoặc điểm dữ liệu mới được thêm vào, thuật toán sẽ tự điều chỉnh dựa trên ví dụ mới đó và ảnh hưởng đến các dự đoán trong tương lai.
- Ít siêu tham số: Các tham số duy nhất cần thiết trong quá trình huấn luyện thuật toán kNN là giá trị  $k$  và lựa chọn phương pháp khoảng cách mà chúng ta muốn sử dụng từ các độ đo đánh giá của chúng.

Nhược điểm của thuật toán kNN:

- Không thích ứng với quy mô lớn: Thuật toán kNN được coi là một thuật toán "lười" (lazy algorithm) vì nó đòi hỏi nhiều tài nguyên tính toán và lưu trữ dữ liệu. Điều này làm cho thuật toán trở nên tốn thời gian và tài nguyên.
- Sự khắc nghiệt của chiều dữ liệu (Curse of Dimensionality): Có một hiện tượng được gọi là hiện tượng đỉnh (peaking phenomenon), theo đó thuật toán kNN bị ảnh hưởng

bởi sự khắc nghiệt của chiều dữ liệu, ngụ ý rằng thuật toán gặp khó khăn trong việc phân loại các điểm dữ liệu một cách đúng đắn khi số chiều quá cao.

- Dễ bị overfitting: Vì thuật toán bị ảnh hưởng bởi hiện tượng đỉnh, nó dễ dàng gặp vấn đề quá khớp. Do đó, thường cần áp dụng các kỹ thuật lựa chọn đặc trưng và giảm chiều dữ liệu để giải quyết vấn đề này.

## **2.8. Công nghệ sử dụng**

### **2.8.1. Spring Boot**

Spring Boot là một framework Java mã nguồn mở dùng để phát triển các ứng dụng Spring đi kèm với các thư viện giúp việc phát triển và quản lý dự án dễ dàng hơn. Spring Boot là một phần mở rộng dựa trên nguyên tắc "cấu hình theo quy ước" (convention-over-configuration) dành cho nền tảng Spring Java, nhằm giảm thiểu các vấn đề liên quan đến cấu hình khi phát triển các ứng dụng Spring.

Spring Boot có thể được sử dụng để xây dựng các microservices, ứng dụng web và ứng dụng console. Các tính năng chính:

- Tích hợp sẵn web server: Tomcat, Jetty hoặc Undertow.
- Cung cấp sẵn các Project Object Models (POMs) cho build tool. Các build tool được hỗ trợ là Maven và Gradle.
- Cấu hình tự động cho ứng dụng Spring.
- Cung cấp các tính năng sẵn sàng cho môi trường production, như: thống kê số liệu, kiểm tra tình trạng hệ thống, và cấu hình ngoại vi.
- Không yêu cầu sinh mã tự động.
- Không yêu cầu cấu hình XML.
- Ngoài Java thì còn hỗ trợ Kotlin và Apache Groovy.

### **2.8.2. PostgreSQL**

PostgreSQL là một hệ thống quản trị cơ sở dữ liệu quan hệ và đối tượng (object-relational database management system) miễn phí và nguồn mở (RDBMS) tiên tiến nhất hiện nay. Khả năng mở rộng cao và tuân thủ các tiêu chuẩn kỹ thuật. Nó được thiết kế để xử lý một loạt các khối lượng công việc lớn, từ các máy tính cá nhân đến kho dữ liệu hoặc dịch vụ Web có nhiều người dùng đồng thời.

- PostgreSQL được phát triển bởi PostgreSQL Global Development Group, Phát hành lần đầu: 08/07/1996
- PostgreSQL linh động có thể chạy được trên nhiều nền tảng khác nhau như Mac OS X, Solaris và Windows.
- PostgreSQL là một phần mềm mã nguồn mở miễn phí bởi vậy PostgreSQL có thể được dùng, sửa đổi và phổ biến bởi bất kỳ ai cho bất kỳ mục đích nào.
- PostgreSQL có tính ổn định cao.
- PostgreSQL là hệ thống quản lý cơ sở dữ liệu đầu tiên triển khai tính năng kiểm soát đồng thời nhiều phiên bản (MVCC).

PostgreSQL tích hợp nhiều tính năng tuyệt vời giúp hỗ trợ nhà phát triển xây dựng app đáp ứng các chức năng phức tạp, truy vấn nhanh chóng và bảo mật duy trì tính toàn vẹn và độ tin cậy. Để đáng tin cậy hơn, Postgresql cung cấp các tùy chọn bảo mật, xác thực và khôi phục thảm họa khác nhau. PostgreSQL được chứng minh là có khả năng mở rộng cao cả về số lượng dữ liệu và số lượng người dùng có thể thao tác cùng lúc.

Một số các tính năng nổi bật:

- Kiểu dữ liệu
  - Nguyên hàm: Số nguyên, số, chuỗi, Boolean
  - Cấu trúc: Date/Time, Array, Phạm vi, UUID
  - Document: JSON/JSONB, XML, Key-value (Hstore)
  - Hình học: Điểm, Đường thẳng, Vòng tròn, Đa giác
  - Tùy chỉnh: Composite, Các kiểu tùy chỉnh
- Bảo mật
  - Xác thực: GSSAPI, SSPI, LDAP, SCRAM-SHA-256, Certificate và các hình thức khác
  - Hệ thống kiểm soát truy cập mạnh mẽ
  - Bảo mật cấp độ cột và hàng
- Khả năng mở rộng
  - Phương pháp lưu trữ
  - Ngôn ngữ thủ tục: PL / PGSQL, Perl, Python (và nhiều ngôn ngữ khác)
  - Trình wrapper dữ liệu ngoài: kết nối với các cơ sở dữ liệu hoặc luồng khác với giao diện SQL chuẩn
  - Và nhiều tiện ích mở rộng cung cấp chức năng bổ sung, bao gồm cả PostGIS
- Tìm kiếm văn bản
  - Hỗ trợ các bộ ký tự quốc tế, ví dụ: thông qua ICU collations
  - Tìm kiếm full-text

### 2.8.3. FastAPI

FastAPI là một framework web hiện đại, nhanh và hiệu suất cao dành cho việc xây dựng API bằng Python, dựa trên tiêu chuẩn type hints của Python.

Các tính năng chính:

- Nhanh: Hiệu suất rất cao, ngang tầm với NodeJS và Go (nhờ vào Starlette và Pydantic). Đây là một trong những framework Python nhanh nhất hiện nay.
- Lập trình nhanh hơn: Tăng tốc độ phát triển tính năng lên khoảng 200%-300%.
- Ít lỗi hơn: Giảm khoảng 40% lỗi do con người (nhà phát triển) gây ra.
- Trực quan: Hỗ trợ tốt cho code editor. Code completion được hỗ trợ ở mọi nơi. Giảm thời gian gỡ lỗi.
- Dễ sử dụng: Thiết kế thân thiện, dễ học và sử dụng. Giảm thời gian cần đọc tài liệu.
- Ngắn gọn: Giảm thiểu trùng lặp mã. Nhiều tính năng được cung cấp từ mỗi khai báo tham số. Ít lỗi hơn.

- Mạnh mẽ: Tạo code sẵn sàng cho môi trường production, tự động tạo documentation cho code.
- Dựa trên tiêu chuẩn: Xây dựng dựa trên (và hoàn toàn tương thích với) các tiêu chuẩn mở dành cho API: OpenAPI (Swagger) và JSON Schema.

#### 2.8.4. SpaCy

SpaCy là một thư viện phần mềm mã nguồn mở dành cho xử lý ngôn ngữ tự nhiên nâng cao, được viết bằng hai ngôn ngữ Python và Cython. Thư viện này được xuất bản với giấy phép MIT và các nhà phát triển chính là Matthew Honnibal và Ines Montani, cũng là những người sáng lập công ty phần mềm Explosion.

Không giống như NLTK, được sử dụng rộng rãi trong giảng dạy và nghiên cứu, spaCy tập trung vào việc cung cấp phần mềm để sử dụng trong sản xuất. spaCy cũng hỗ trợ các quy trình làm việc học sâu cho phép nối kết với các mô hình thống kê được huấn luyện bởi các thư viện máy học phổ biến như TensorFlow, PyTorch hay Apache MXNet thông qua thư viện học máy Thinc của riêng nó. Sử dụng Thinc làm chương trình phụ trợ (backend) của nó, spaCy làm nổi bật các mô hình mạng thần kinh tích chập cho các tác vụ gán nhãn từ loại (part-of-speech tagging), cây phân tích cú pháp, phân loại tài liệu và nhận dạng thực thể có tên. Các mô hình thống kê mạng thần kinh nhân tạo được tích hợp trước để thực hiện các tác vụ này sẵn có ở 17 ngôn ngữ, bao gồm tiếng Anh, Bồ Đào Nha, Tây Ban Nha, Nga, Trung Quốc, và cũng có một mô hình NER đa ngữ. Thêm nữa, spaCy cũng hỗ trợ token hóa cho hơn 65 ngôn ngữ, cho phép người dùng huấn luyện mô hình tùy chỉnh trên các tập dữ liệu của riêng mình.

#### 2.8.5. Scikit-learn

Scikit-learn là một thư viện học máy phổ biến và mạnh mẽ, hỗ trợ nhiều loại thuật toán cũng như các công cụ để trực quan hóa Machine Learning, tiền xử lý, điều chỉnh, lựa chọn và đánh giá mô hình. Dựa trên NumPy, SciPy và Matplotlib, Scikit-learn có một số thuật toán hiệu quả để phân loại, phân tích hồi quy và phân cụm. Bao gồm các vector machines, rain forests, gradient boosting, k-means, và DBSCAN.

Scikit-learn được biết đến với tính dễ phát triển tương đối nhờ các API được thiết kế nhất quán và hiệu quả, tài liệu mở rộng cho hầu hết các thuật toán và nhiều hướng dẫn online. API Scikit-learn đã trở thành tiêu chuẩn thực tế cho việc triển khai học máy nhờ tính dễ sử dụng, thiết kế chính chu và sự hỗ trợ đông đảo trong cộng đồng Scikit-learn.

Scikit-learn cung cấp các mô-đun để xây dựng, điều chỉnh và đánh giá mô hình ML:

- Preprocessing đề cập đến các công cụ Scikit-learn hữu ích trong việc trích xuất và chuẩn hóa features trong quá trình phân tích dữ liệu.
- Classification đề cập đến một bộ công cụ xác định danh mục liên kết với dữ liệu trong mô hình học máy. Ví dụ: những công cụ này được sử dụng để phân loại email hợp lệ hoặc thư rác. Về cơ bản, Classification xác định đối tượng thuộc về loại nào.
- Regression (Phân tích hồi quy) đề cập đến việc tạo ra một mô hình học máy để tìm hiểu mối quan hệ giữa dữ liệu đầu vào và đầu ra, chẳng hạn như hành vi hoặc giá

cổ phiếu. Hồi quy dự đoán một thuộc tính có giá trị liên tục được liên kết với một đối tượng.

- Các công cụ Clustering trong Scikit-learn tự động nhóm dữ liệu có đặc điểm tương tự thành các bộ (sets), chẳng hạn như dữ liệu khách hàng được sắp xếp theo bộ dựa trên vị trí thực tế.
- Dimensionality Reduction làm giảm số lượng biến ngẫu nhiên để phân tích. Ví dụ: để tăng hiệu quả trực quan hóa, dữ liệu bên ngoài có thể bị loại bỏ.
- Model selection đề cập đến các thuật toán và khả năng cung cấp các công cụ so sánh, xác thực và chọn tham số tối ưu để sử dụng trong các dự án học máy khoa học dữ liệu.
- Pipeline đề cập đến các tiện ích để xây dựng quy trình làm việc mô hình.
- Visualization cho học máy cho phép vẽ đồ thị và điều chỉnh trực quan nhanh chóng.

## CHƯƠNG 3: PHÂN TÍCH THIẾT KẾ HỆ THỐNG

### 3.1. Công nghệ và tài nguyên sử dụng:

- Web Application: Java Spring Boot + Thymeleaf
- Database: PostgreSQL
- AI: Python Machine Learning

### 3.2. Đặc tả yêu cầu

#### 3.2.1. Các tác nhân của hệ thống

- Khách hàng: Đăng ký, đăng nhập, đổi mật khẩu tài khoản cá nhân, chỉnh sửa thông tin cá nhân, tìm kiếm sách, xem chi tiết sách, thêm sách vào giỏ hàng, quản lý giỏ hàng, đặt hàng, quản lý đơn đã đặt.

- Nhân viên: Đăng nhập, đổi mật khẩu tài khoản cá nhân, quản lý tài khoản khách hàng, quản lý tài khoản nhân viên, chỉnh sửa thông tin cá nhân, quản lý sách, quản lý đơn đặt hàng của khách hàng.

#### 3.2.2. Các chức năng của hệ thống

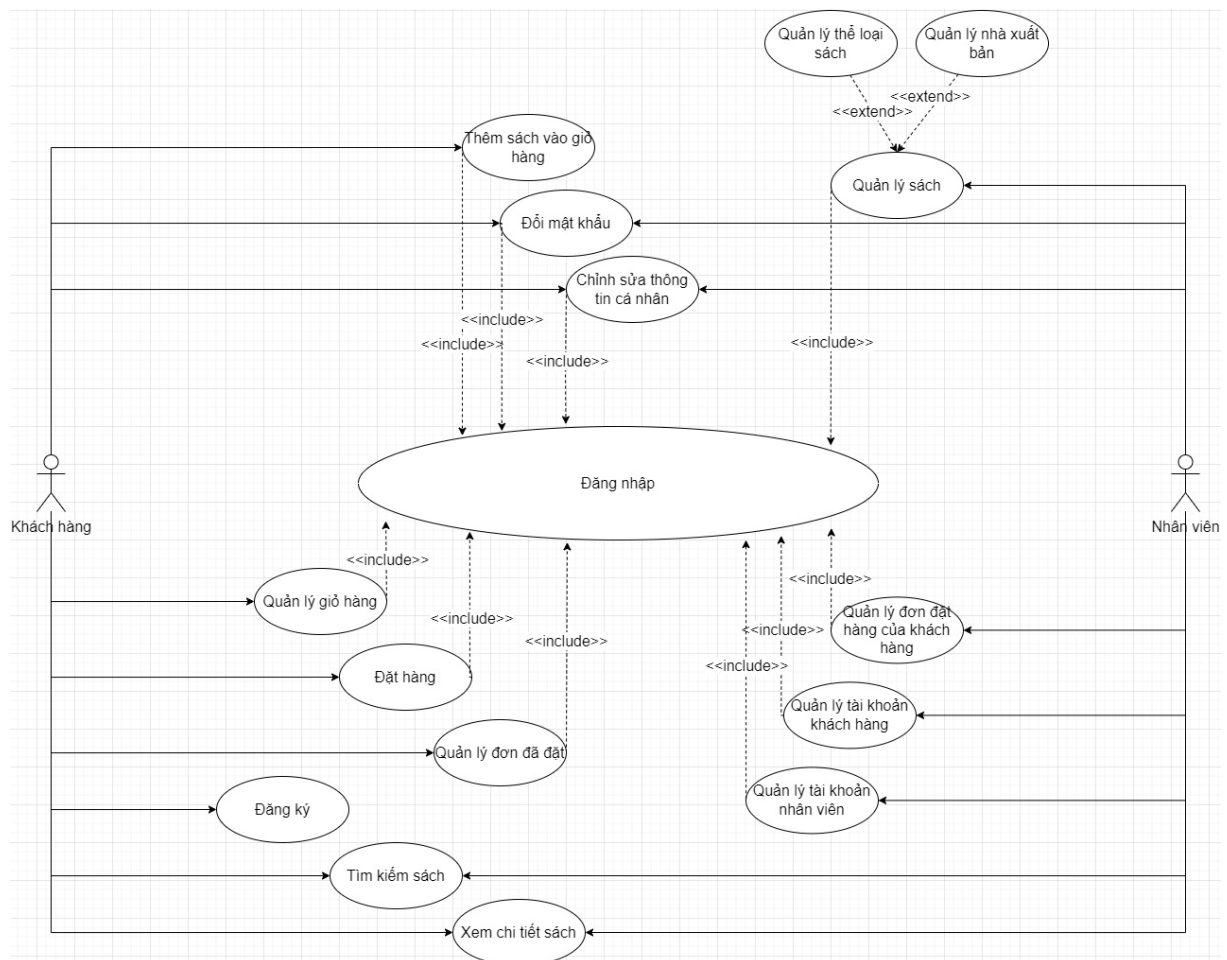
STT	Chức năng	Mô tả
1	Đăng ký	Khách hàng có thể tạo tài khoản để đăng nhập hệ thống
2	Đăng nhập	Khách hàng hoặc nhân viên có thể đăng nhập vào hệ thống
3	Đổi mật khẩu	Người dùng hệ thống có thể đổi mật khẩu tài khoản đăng nhập
4	Chỉnh sửa thông tin cá nhân	Người dùng hệ thống có thể đổi thông tin cá nhân liên quan đến tài khoản
5	Tìm kiếm sách	Khách hàng có thể tìm kiếm sách bằng các từ khoá liên quan đến tên sách, thể loại, tên tác giả
6	Xem chi tiết sách	Khách hàng có thể xem chi tiết của sách như tên sách, thể loại, tác giả, hình ảnh, giá cả,...
7	Thêm sách vào giỏ hàng	Khách hàng có thể thêm sách vào giỏ hàng cá nhân để lưu lại các sách cần mua sau.
8	Quản lý giỏ hàng	Khách hàng có thể chỉnh sửa giỏ hàng như thêm hoặc bớt số lượng sách, xoá sách khỏi giỏ hàng
9	Đặt hàng	Khách hàng có thể đặt mua các sách ở trong giỏ hàng sau khi xác nhận giá cả, thông tin giao hàng
10	Quản lý đơn đã đặt	Khách hàng có thể xem thông tin đơn hàng mình đã đặt, huỷ đơn hàng chưa giao
11	Quản lý tài khoản khách hàng	Nhân viên có thể đổi mật khẩu, thông tin đăng nhập của tài khoản khách hàng



12	Quản lý tài khoản nhân viên	Nhân viên có thể thêm tài khoản nhân viên, đổi mật khẩu, thông tin đăng nhập của tài khoản nhân viên
13	Quản lý sách	Nhân viên thêm và chỉnh sửa sách, quản lý các thông tin của sách như nhà xuất bản, thể loại, số lượng, giá cả,...
14	Quản lý đơn đặt hàng của khách hàng	Nhân viên có thể xem và quản lý các đơn đặt hàng mà khách hàng đã đặt

*Bảng 3.1. Các ca sử dụng hệ thống*

### 3.3. Sơ đồ ca sử dụng mức tổng quát



*Hình 3.1. Sơ đồ ca sử dụng tổng quát*

### 3.4. Cơ sở dữ liệu

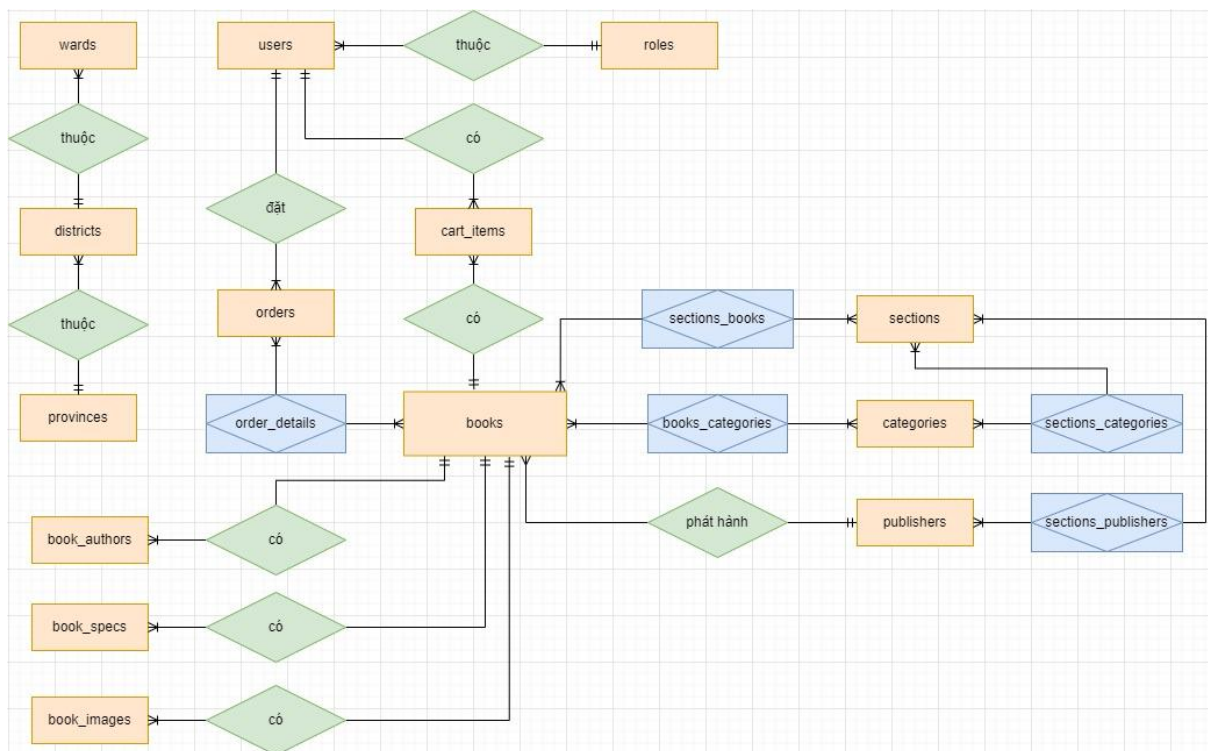
#### 3.4.1. Các thực thể

STT	Thực thể	Thuộc tính	Mô tả
1	provinces	code, name, name_en, full_name, full_name_en, code_name	Thông tin các tỉnh/thành phố
2	districts	code, name, name_en, full_name, full_name_en, code_name, province_code	Thông tin các quận/huyện
3	wards	code, name, name_en, full_name, full_name_en, code_name, district_code	Thông tin các phường/xã
4	roles	id, name	Vai trò của người dùng
5	users	id, first_name, last_name, email, password, role_id, image, enabled, created_at, verification_code	Thông tin người dùng
6	categories	id, name, slug	Thể loại sách
7	publishers	id, name, slug, logo	Thông tin nhà xuất bản
8	books	id, name, slug, main_image, short_description, full_description, age_limit, page_count, published_at, language, quantity, price, discount_percent, shipping_fee, publisher_id, enabled, created_at, updated_at	Thông tin sách

9	books_authors	id, book_id, author_name	Thông tin tác giả của sách
10	book_images	id, file_name, book_id	Ảnh của sách
11	book_specs	id, name, value, book_id	Thông số kỹ thuật của sách
12	cart_items	customer_id, book_id, quantity	Thông tin giỏ hàng của khách hàng
13	orders	id, customer_id, order_time, items_price, shipping_cost, total, status, full_name, phone_number, address_line, ward, district, province, deliver_time	Thông tin đơn hàng
14	order_details	order_id, book_id, quantity, unit_price	Chi tiết từng đơn hàng
15	sections	id, title, enabled, order, type	Thông tin các section của trang chủ
16	sections_books	id, book_id, section_id, order	Liên kết sách với các section
17	sections_categories	id, category_id, section_id, order	Liên kết danh mục với các section
18	sections_publishers	id, publisher_id, section_id, order	Liên kết nhà xuất bản với các section

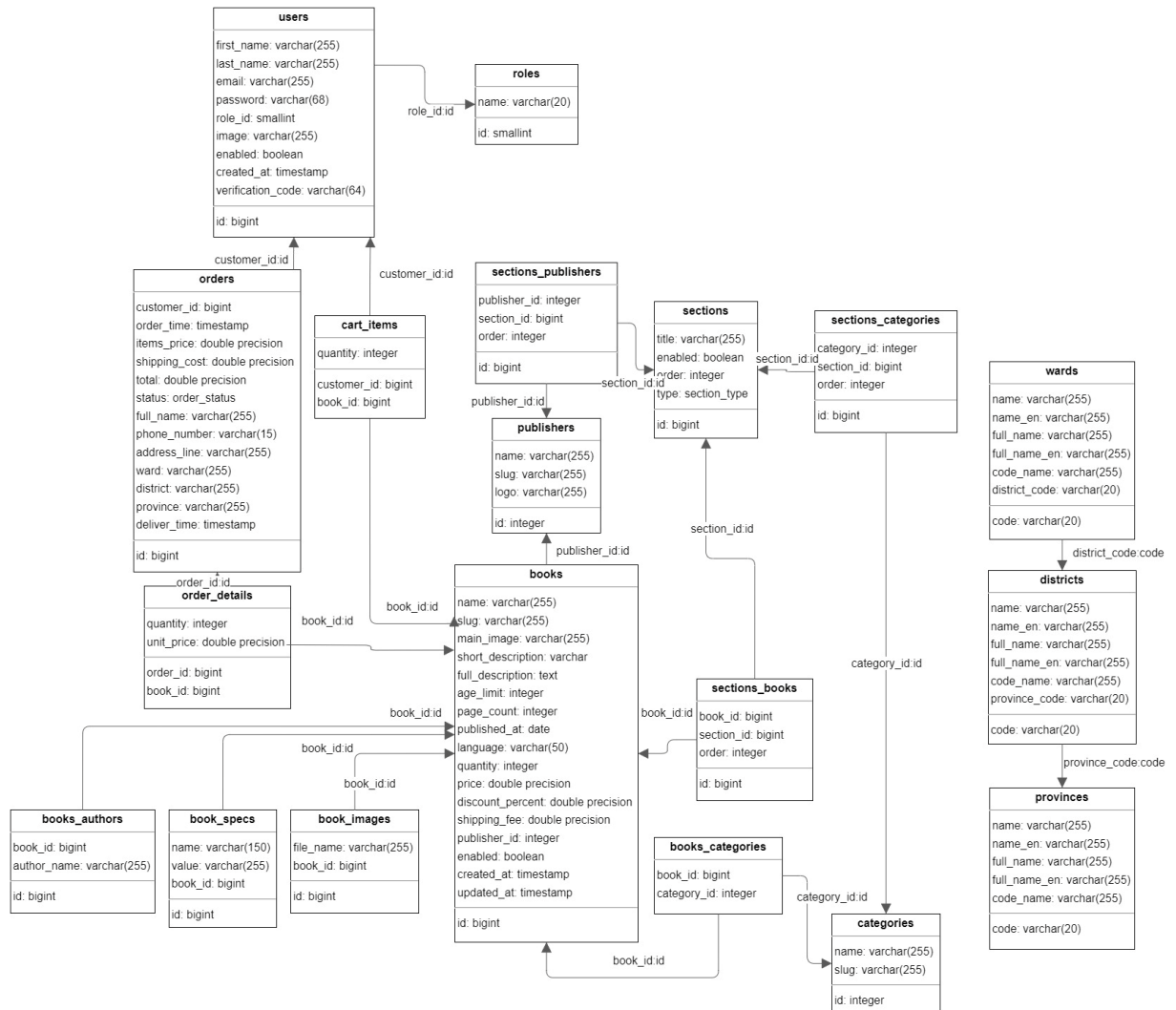
*Bảng 3.2. Các thực thể*

### 3.4.2. Mô hình ERD



Hình 3.2. Mô hình ERD

### 3.4.3. Mô hình Diagram



*Hình 3.3. Mô hình diagram*

### 3.4.4. Mô hình dữ liệu quan hệ

**Ghi chú:**

- Khoá chính: **khoa\_chinh**
- Khoá ngoại: `khoa_ngoai`
- Khoá tổ hợp: **khoa\_1, khoa\_2**

provinces (**code**, name, name en, full name, full name en, code name)

districts (code, name, name en, full name, full name en, code name, province code)

wards (**code**, name, name en, full name, full name en, code name, district code)roles (**id**, name)

users (**id**, first\_name, last\_name, email, password, **role\_id**, image, enabled, created\_at, verification code)

categories (**id**, name, slug)

publishers (**id**, name, slug, logo)

books (**id**, name, slug, main\_image, short\_description, full\_description, age\_limit, page\_count, published\_at, language, quantity, price, discount\_percent, shipping\_fee, publisher\_id, enabled, created\_at, updated\_at)

books\_authors (**id**, book\_id, author\_name)

books\_categories (**book\_id**, **category\_id**)

book\_images (**id**, file\_name, book\_id)

book\_specs (**id**, name, value, book\_id)

cart\_items (**customer\_id**, **book\_id**, quantity)

orders (**id**, customer\_id, order\_time, items\_price, shipping\_cost, total, status, full\_name, phone\_number, address\_line, ward, district, province, deliver\_time)

order\_details (**order\_id**, **book\_id**, quantity, unit\_price)

sections (**id**, title, enabled, order, type)

sections\_books (**id**, book\_id, section\_id, order)

sections\_categories (**id**, category\_id, section\_id, order)

sections\_publishers (**id**, publisher\_id, section\_id, order)

### 3.4.5. Từ điển dữ liệu

1. provinces (**code**, name, name\_en, full\_name, full\_name\_en, code\_name)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	code	VARCHAR	20	PK, Not Null	Mã tỉnh/thành phố
2	name	VARCHAR	255	Not Null	Tên tỉnh/thành phố
3	name_en	VARCHAR	255		Tên tiếng Anh của tỉnh/thành phố
4	full_name	VARCHAR	255	Not Null	Tên đầy đủ của tỉnh/thành phố
5	full_name_en	VARCHAR	255		Tên đầy đủ tiếng Anh
6	code_name	VARCHAR	255		Tên mã của tỉnh/thành phố

Bảng 3.3. Bảng provinces

2. districts (**code**, name, name\_en, full\_name, full\_name\_en, code\_name, province\_code)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	code	VARCHAR	20	PK, Not Null	Mã quận/huyện
2	name	VARCHAR	255	Not Null	Tên quận/huyện
3	name_en	VARCHAR	255		Tên tiếng Anh của quận/huyện
4	full_name	VARCHAR	255		Tên đầy đủ của quận/huyện
5	full_name_en	VARCHAR	255		Tên đầy đủ tiếng Anh
6	code_name	VARCHAR	255		Tên mã của quận/huyện
7	province_code	VARCHAR	20	FK, Not Null	Mã tỉnh/thành phố

*Bảng 3.4. Bảng districts*

3. wards (**code**, name, name\_en, full\_name, full\_name\_en, code\_name, district\_code)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	code	VARCHAR	20	PK, Not Null	Mã phường/xã
2	name	VARCHAR	255	Not Null	Tên phường/xã
3	name_en	VARCHAR	255		Tên tiếng Anh của phường/xã
4	full_name	VARCHAR	255		Tên đầy đủ của phường/xã
5	full_name_en	VARCHAR	255		Tên đầy đủ tiếng Anh
6	code_name	VARCHAR	255		Tên mã của phường/xã
7	district_code	VARCHAR	20	FK, Not Null	Mã quận/huyện

*Bảng 3.5. Bảng wards*

4. roles (**id**, name)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	SMALLINT		PK, Not Null	Mã vai trò
2	name	VARCHAR	20	Not Null	Tên vai trò

*Bảng 3.6. Bảng roles*

5. users (id, first\_name, last\_name, email, password, role\_id, image, enabled, created\_at, verification\_code)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	BIGINT		PK, Not Null	Mã người dùng
2	first_name	VARCHAR	255	Not Null	Tên người dùng
3	last_name	VARCHAR	255	Not Null	Họ của người dùng
4	email	VARCHAR	255	Unique, Not Null	Email người dùng
5	password	VARCHAR	68	Not Null	Mật khẩu
6	role_id	SMALLINT		FK, Not Null	Mã vai trò
7	image	VARCHAR	255		Hình ảnh đại diện của người dùng
8	enabled	BOOLEAN		Default False, Not Null	Trạng thái kích hoạt
9	created_at	TIMESTAMP		Default Now(), Not Null	Thời gian tạo
10	verification_code	VARCHAR	64		Mã xác minh

*Bảng 3.7. Bảng users*

6. categories (id, name, slug)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	INT		PK, Not Null	Mã thể loại sách
2	name	VARCHAR	255	Unique, Not Null	Tên thể loại
3	slug	VARCHAR	255	Unique, Not Null	Đường dẫn của danh mục

*Bảng 3.8. Bảng provinces*

7. publishers (id, name, slug, logo)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	INT		PK, Not Null	Mã nhà xuất bản
2	name	VARCHAR	255	Unique, Not Null	Tên nhà xuất bản
3	slug	VARCHAR	255	Unique, Not Null	Đường dẫn của nhà xuất bản
4	logo	VARCHAR	255		Logo của nhà xuất bản

*Bảng 3.9. Bảng publishers*



8. books (id, name, slug, main\_image, short\_description, full\_description, age\_limit, page\_count, published\_at, language, quantity, price, discount\_percent, shipping\_fee, publisher\_id, enabled, created\_at, updated\_at)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	BIGINT		PK, Not Null	Mã sách
2	name	VARCHAR	255	Unique, Not Null	Tên sách
3	slug	VARCHAR	255	Unique, Not Null	Đường dẫn của sách
4	main_image	VARCHAR	255		Hình ảnh chính của sách
5	short_description	VARCHAR			Mô tả ngắn của sách
6	full_description	TEXT			Mô tả đầy đủ của sách
7	age_limit	INT		Not Null	Độ tuổi giới hạn
8	page_count	INT		Not Null	Số trang
9	published_at	DATE		Not Null	Ngày xuất bản
10	language	VARCHAR	50	Not Null	Ngôn ngữ của sách
11	quantity	INT		Not Null	Số lượng
12	price	FLOAT		Not Null	Giá
13	discount_percent	FLOAT		Default 0, Not Null	Phần trăm giảm giá
14	shipping_fee	FLOAT		Not Null	Phí vận chuyển
15	publisher_id	INT		FK, Not Null	Mã nhà xuất bản
16	enabled	BOOLEAN		Default False, Not Null	Trạng thái kích hoạt
17	created_at	TIMESTAMP		Not Null	Thời gian tạo
18	updated_at	TIMESTAMP		Not Null	Thời gian cập nhật

*Bảng 3.10. Bảng books*

9. books\_authors (id, book\_id, author\_name)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	BIGINT		PK, Not Null	Mã tác giả sách
2	book_id	BIGINT		FK, Not Null	Mã sách
3	author_name	VARCHAR	255	Not Null	Tên tác giả

*Bảng 3.11. Bảng books\_authors*

10. books\_categories (book id, category id)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	book_id	BIGINT		FK, Not Null	Mã sách
2	category_id	INT		FK, Not Null	Mã thể loại

*Bảng 3.12. Bảng books\_categories*

11. book\_images (id, file\_name, book\_id)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	BIGINT		PK, Not Null	Mã hình ảnh sách
2	file_name	VARCHAR	255	Not Null	Tên tệp hình ảnh
3	book_id	BIGINT		FK, Not Null	Mã sách

*Bảng 3.13. Bảng book\_images*

12. book\_specs (id, name, value, book\_id)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	BIGINT		PK, Not Null	Mã thông số kỹ thuật sách
2	name	VARCHAR	150	Not Null	Tên thông số
3	value	VARCHAR	255	Not Null	Giá trị của thông số
4	book_id	BIGINT		FK, Not Null	Mã sách

*Bảng 3.14. Bảng book\_specs*

13. cart\_items (**customer\_id**, **book\_id**, quantity)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	customer_id	BIGINT		PK, FK, Not Null	Mã khách hàng
2	book_id	BIGINT		PK, FK, Not Null	Mã sách
3	quantity	INT		Not Null	Số lượng

Bảng 3.15. Bảng cart\_items

14. orders (**id**, customer\_id, order\_time, items\_price, shipping\_cost, total, status, full\_name, phone\_number, address\_line, ward, district, province, deliver\_time)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	BIGINT		PK, Not Null	Mã đơn hàng
2	customer_id	BIGINT		FK, Not Null	Mã khách hàng
3	order_time	TIMESTAMP		Not Null	Thời gian đặt hàng
4	items_price	FLOAT		Not Null	Tổng giá trị hàng hoá
5	shipping_cost	FLOAT		Not Null	Phí vận chuyển
6	total	FLOAT		Not Null	Tổng giá trị đơn hàng
7	status	ORDER_STATUS		Not Null	Trạng thái đơn hàng
8	full_name	VARCHAR	255	Not Null	Họ và tên người nhận
9	phone_number	VARCHAR	15	Not Null	Số điện thoại người nhận
10	address_line	VARCHAR	255	Not Null	Địa chỉ chi tiết
11	ward	VARCHAR	255	Not Null	Phường/xã của địa chỉ giao hàng
12	district	VARCHAR	255	Not Null	Quận/huyện của địa chỉ giao hàng
13	province	VARCHAR	255	Not Null	Tỉnh/thành phố của địa chỉ giao hàng
14	deliver_time	TIMESTAMP			Thời gian giao hàng (nếu có)

Bảng 3.16. Bảng orders

ORDER\_STATUS thuộc {'ORDERED', 'SHIPPING', 'DELIVERED', 'CANCELLED'}

15. order\_details (order id, book id, quantity, unit\_price)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	order_id	BIGINT		PK, FK, Not Null	Mã đơn hàng
2	book_id	BIGINT		PK, FK, Not Null	Mã sách
3	quantity	INT		Not Null	Số lượng sách
4	unit_price	FLOAT		Not Null	Giá của mỗi đơn vị sách

*Bảng 3.17. Bảng order\_details*

16. sections (id, title, enabled, order, type)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	BIGINT		PK, Not Null	Mã section
2	title	VARCHAR	255	Not Null	Tiêu đề section
3	enabled	BOOLEAN		Default False, Not Null	Trạng thái kích hoạt
4	order	INT		Not Null	Thứ tự sắp xếp
5	type	SECTION_TYPE		Not Null	Loại section

*Bảng 3.18. Bảng sections*

SECTION\_TYPE thuộc {'BOOK', 'CATEGORY', 'PUBLISHER'}

17. sections\_books (id, book\_id, section\_id, order)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	BIGINT		PK, Not Null	Mã
2	book_id	BIGINT		FK, Not Null	Mã sách
3	section_id	BIGINT		FK, Not Null	Mã section
4	order	INT		Not Null	Thứ tự của sách trong section

*Bảng 3.19. Bảng sections\_books*

18. sections\_categories (**id**, category\_id, section\_id, order)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	BIGINT		PK, Not Null	Mã
2	category_id	INT		FK, Not Null	Mã thể loại
3	section_id	BIGINT		FK, Not Null	Mã section
4	order	INT		Not Null	Thứ tự thể loại trong section

*Bảng 3.20. Bảng sections\_categories*

19. sections\_publishers (**id**, publisher\_id, section\_id, order)

STT	Thuộc tính	Kiểu dữ liệu	Độ dài	Ràng buộc	Ghi chú
1	id	BIGINT		PK, Not Null	Mã định danh quan hệ
2	publisher_id	INT		FK, Not Null	Mã định danh nhà xuất bản
3	section_id	BIGINT		FK, Not Null	Mã định danh section
4	order	INT		Not Null	Thứ tự sắp xếp trong section

*Bảng 3.21. Bảng sections\_publishers*

## CHƯƠNG 4: XÂY DỰNG HỆ THỐNG

### 4.1. Hệ thống đề xuất sản phẩm thông minh

#### 4.1.1. Thu thập và xây dựng cơ sở dữ liệu

**Dữ liệu mẫu:**

Book title	Publisher name	Category 1	Category 2	Category 3	Author 1	Author 2
Naruto - Tập 1	NXB Kim Đồng	Giã trường	Manga	missing	Masashi Kishimoto	Shin Towada
Naruto - Tập 12	NXB Kim Đồng	Giã trường	Manga	missing	Masashi Kishimoto	Shin Towada
Love Wager	NXB Trẻ	Lãng mạn	Nước ngoài	Tiểu thuyết	Lynn Painter	missing

**Chiến lược đề xuất:** Phân tích các thuộc tính của sách như tên tác giả, thể loại,...

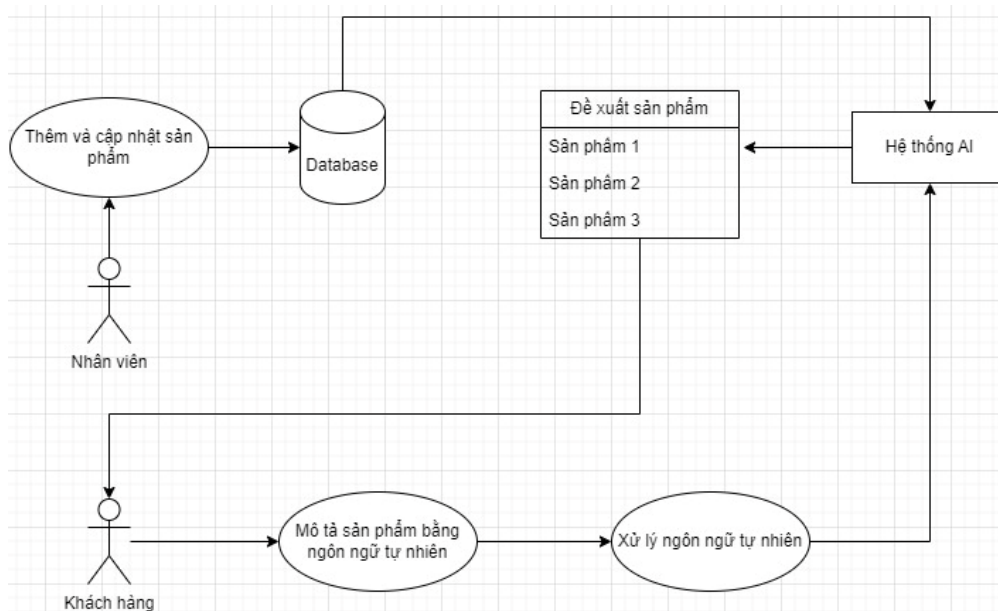
#### 4.1.2. Quy trình hoạt động

**Mô tả:** Hệ thống đề xuất sản phẩm thông minh là một ứng dụng nhân tạo (AI) sử dụng thuật toán kNN để đề xuất sản phẩm dựa vào mô tả sản phẩm mà người dùng nhập vào.

**Quy trình hoạt động:**

- Hệ thống cung cấp một giao diện cho phép người dùng tương tác trực tiếp bằng cách nhập từ khoá mô tả sản phẩm trên thanh tìm kiếm.
- Hệ thống sử dụng thuật toán kNN để huấn luyện mô hình bằng dữ liệu đã lưu từ database.
- Hệ thống sử dụng công nghệ xử lý ngôn ngữ tự nhiên để làm sạch mô tả của người dùng (xóa stop word, dấu câu,...)
- Hệ thống xử lý dữ liệu mô tả mà người dùng đã nhập, lấy kết quả đề xuất từ mô hình đã huấn luyện và trả về kết quả đề xuất liên quan đến mô tả mà người dùng đã nhập.

### Sơ đồ hoạt động:



Hình 4.1. Sơ đồ hoạt động hệ thống thông minh

### 4.1.3. Phát triển hệ thống đề xuất sản phẩm thông minh

#### Các thư viện sử dụng:

```
import os
import warnings

import joblib
import pandas as pd
import spacy
from dotenv import load_dotenv
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.neighbors import NearestNeighbors
from sqlalchemy import create_engine
```

Hình 4.2. Import thư viện

- **os**: Thư viện có sẵn của python, dùng để tương tác với hệ điều hành. Ta sẽ dùng os để lấy các biến môi trường (environment variable).
- **warnings**: Thư viện warnings được sử dụng để kiểm soát và quản lý các thông báo cảnh báo trong Python.
- **joblib**: Joblib là một thư viện python để chạy các tác vụ tính toán chuyên sâu song song. Nó cung cấp một tập hợp các hàm để thực hiện các hoạt động song song trên các tập dữ liệu lớn và để lưu trữ đệm kết quả của các hàm tính toán tốn kém. Ta sẽ dùng joblib để lưu các mô hình đã huấn luyện thành một file.
- **pandas**: Thư viện pandas là một thư viện mạnh mẽ để làm việc với dữ liệu dạng bảng, cung cấp các cấu trúc dữ liệu như DataFrame và Series.
- **spacy**: Thư viện spaCy là một thư viện xử lý ngôn ngữ tự nhiên (NLP). Ta sẽ dùng spaCy để xử lý dữ liệu huấn luyện và dữ liệu và xóa các stop word và dấu câu.

- **dotenv**: Thư viện dotenv dùng để đọc các biến môi trường từ file .env để python có thể sử dụng được.
- **sklearn**: Thư viện scikit-learn là thư viện mạnh mẽ dành cho các thuật toán học máy được viết trên ngôn ngữ Python. Thư viện này tích hợp rất nhiều thuật toán hiện đại và cổ điển, các công cụ xử lý các bài toán học máy, giúp cho việc phát triển các tính năng học máy trong python dễ dàng và nhanh chóng hơn.
- **sqlalchemy**: dùng để lấy dữ liệu huấn luyện từ cơ sở dữ liệu vào DataFrame của pandas.

**Bước 1:** Tạo class KnnModel chứa các phương thức dùng trong huấn luyện và lấy kết quả

```
class KnnModel: 2 usages  @ khanghy1000 *
    def __init__(self):  @ khanghy1000
        self.data = get_data()
        self.nlp = spacy.blank("vi")
        self.cv = CountVectorizer()
        self.model = NearestNeighbors(metric='cosine', algorithm='brute')

    def clean_text(self, text):...

    def train_model(self, model_path):...

    def load_model(self, model_path):...

    def recommend(self, text, k=5):...
```

*Hình 4.3. Class KnnModel*

- **data**: chứa dữ liệu huấn luyện.
- **nlp**: bộ xử lý ngôn ngữ tự nhiên.
- **cv**: CountVectorizer của thư viện scikit-learn, dùng để biến đổi văn bản thành các vectơ đặc trưng dựa trên số lượng từ xuất hiện trong văn bản.
- **model**: chứa mô hình huấn luyện. Ta dùng NearestNeighbors và cấu hình nó để sử dụng độ đo cosine để tìm kiếm lân cận.
- **clean\_text**: phương thức tiền xử lý văn bản, dùng để chuyển văn bản thành chữ thường, xóa dấu câu, xóa stop word.
- **train\_model**: phương pháp huấn luyện mô hình, sau đó lưu mô hình đã huấn luyện vào model\_path.
- **load\_model**: phương pháp load mô hình đã huấn luyện từ file.
- **recommend**: phương thức lấy kết quả đề xuất.



## Bước 2: Chuẩn bị dữ liệu huấn luyện

Hệ thống đề xuất sách sẽ được huấn luyện bằng dữ liệu sách từ cơ sở dữ liệu. Ta sẽ tạo 1 view ở cơ sở dữ liệu để lấy dữ liệu cần thiết để huấn luyện. Dữ liệu gồm các thuộc tính sau:

- id: id sách
- book\_title: tiêu đề sách
- publisher\_name: tên nhà xuất bản
- category\_1: thể loại sách 1
- category\_2: thể loại sách 2
- category\_3: thể loại sách 3
- author\_1: tên tác giả 1
- author\_2: tên tác giả 2

Trong cơ sở dữ liệu, một sách có thể có một hoặc nhiều thể loại, tác giả cũng tương tự. Trong dữ liệu huấn luyện, ta chỉ lấy 3 thể loại và 2 tác giả đầu tiên của sách.

```
1  ✓ CREATE OR REPLACE VIEW get_book_data AS
2  WITH RNs AS (SELECT b.id,
3                    b.name AS book_title,
4                    p.name AS publisher_name,
5                    c.name AS category_name,
6                    ba.author_name,
7                    ROW_NUMBER() OVER (PARTITION BY b.id, ba.author_name ORDER BY c.name) AS c_rn,
8                    ROW_NUMBER() OVER (PARTITION BY b.id, c.name ORDER BY ba.author_name) AS a_rn
9  FROM books b
10     LEFT JOIN public.books_categories bc 1<->0..n: ON b.id = bc.book_id
11     LEFT JOIN public.categories c 1..n<->1: ON c.id = bc.category_id
12     LEFT JOIN public.publishers p 1..n<->1: ON p.id = b.publisher_id
13     LEFT JOIN public.books_authors ba 1<->0..n: ON b.id = ba.book_id)
14  SELECT id,
15         book_title,
16         publisher_name,
17         MAX(CASE WHEN c_rn = 1 THEN category_name END) AS category_1,
18         MAX(CASE WHEN c_rn = 2 THEN category_name END) AS category_2,
19         MAX(CASE WHEN c_rn = 3 THEN category_name END) AS category_3,
20         MAX(CASE WHEN a_rn = 1 THEN author_name END) AS author_1,
21         MAX(CASE WHEN a_rn = 2 THEN author_name END) AS author_2
22  FROM RNs
23  GROUP BY id, book_title, publisher_name
24  ORDER BY id;
25  ✓ END;
26
```

Hình 4.4. View lấy dữ liệu

Sau đó, ta sẽ load bộ dữ liệu này vào DataFrame của thư viện pandas bằng python và thay thế các dữ liệu trống bằng “missing”.

```
def get_data(): 1 usage  khanghy1000
    POSTGRES_ADDRESS = os.getenv("POSTGRES_ADDRESS")
    POSTGRES_PORT = os.getenv("POSTGRES_PORT")
    POSTGRES_USERNAME = os.getenv("POSTGRES_USERNAME")
    POSTGRES_PASSWORD = os.getenv("POSTGRES_PASSWORD")
    POSTGRES_DBNAME = os.getenv("POSTGRES_DBNAME")

    postgres_str = ('postgresql://{username}:{password}@{ipaddress}:{port}/{dbname}'
                    .format(username=POSTGRES_USERNAME,
                            password=POSTGRES_PASSWORD,
                            ipaddress=POSTGRES_ADDRESS,
                            port=POSTGRES_PORT,
                            dbname=POSTGRES_DBNAME))

    cnx = create_engine(postgres_str)

    data = pd.read_sql_query( sql: "SELECT * FROM get_book_data", cnx)
    data.fillna( value: "missing", inplace=True)
    return data
```

Hình 4.5. Lấy dữ liệu vào DataFrame

```
list(data.columns)
✓ [10] < 10 ms

['id',
 'book_title',
 'publisher_name',
 'category_1',
 'category_2',
 'category_3',
 'author_1',
 'author_2']
```

Hình 4.6. Các column của DataFrame

### Bước 3: Tiền xử lý dữ liệu

Định nghĩa phương thức `clean_text` để làm sạch văn bản. Phương thức này sẽ chuyển văn bản thành chữ thường, chuyển văn bản thành các token, xoá các stop word và dấu câu, sau đó gộp các token lại.

```
def clean_text(self, text): 2 usages  khanghy1000 *
    doc = self.nlp(text.lower())
    tokens = [token.text for token in doc if not token.is_stop and not token.is_punct]
    return " ".join(tokens)
```

Hình 4.7. Phương thức `clean_text`

Gộp các cột thuộc tính cần dùng để huấn luyện vào cột “`clean_text`” của DataFrame. Sau đó ta dùng phương thức `clean_text` để làm sạch văn bản cột “`clean_text`”

```

self.data["clean_text"] = (
    self.data["book_title"] + " " +
    self.data["publisher_name"] + " " +
    self.data["category_1"] + " " +
    self.data["category_2"] + " " +
    self.data["category_3"] + " " +
    self.data["author_1"] + " " +
    self.data["author_2"]
)
self.nlp.Defaults.stop_words.add("sách")
self.data["clean_text"] = self.data["clean_text"].apply(self.clean_text)

```

*Hình 4.8. Làm sạch dữ liệu văn bản huấn luyện*

#### **Bước 4:** Triển khai mô hình

Dùng CountVectorizer để biến đổi văn bản của cột “clean\_text” của DataFrame thành vector đặc trưng. Sau đó fit mô hình NearestNeighbors với các vector đặc trưng. Sau khi fit xong, ta lưu mô hình vào file.

```

clean_text_vector = self.cv.fit_transform(self.data['clean_text']).toarray()
self.model.fit(clean_text_vector)

joblib.dump(self.model, model_path)

```

*Hình 4.9. Fit mô hình*

Định nghĩa phương thức load\_model dùng để load mô hình đã fit từ file.

```

def load_model(self, model_path): 1 usage  khanghy1000
    self.model = joblib.load(model_path)

```

*Hình 4.10. Phương thức load\_model*

#### **Bước 5:** Tìm kiếm sản phẩm dựa vào mô tả của người dùng

Định nghĩa phương thức recommend dùng để lấy kết quả đề xuất sách dựa vào mô tả. Phương thức sẽ trả về danh sách các sách, mỗi sách là một dict chứa id và tiêu đề sách.

```

def recommend(self, text, k=5): 1 usage  khanghy1000 *
    cleaned_text = self.cv.transform([self.clean_text(text)]).toarray()
    distances, indices = self.model.kneighbors(cleaned_text, n_neighbors=k)
    return self.data.iloc[indices[0]][["id", "book_title"]].to_dict(orient="records")

```

*Hình 4.11. Phương thức recommend*

```
user_description = "tôi muốn mua sách tiểu thuyết của nxb kim đồng"
recommend(user_description)
```

✓ [7] 17ms

```
[{'id': 111,
  'book_title': 'Tiểu thuyết Naruto - Itachi chân truyền - Âm dạ thiên'},
 {'id': 112,
  'book_title': 'Tiểu thuyết Naruto - Itachi chân truyền - Quang minh thiên'},
 {'id': 115, 'book_title': 'Gambit Hậu'},
 {'id': 87, 'book_title': 'Viên ngọc trai kì diệu'},
 {'id': 93, 'book_title': 'Tiểu Thuyết Naruto - Sasuke Chân Truyền'}]
```

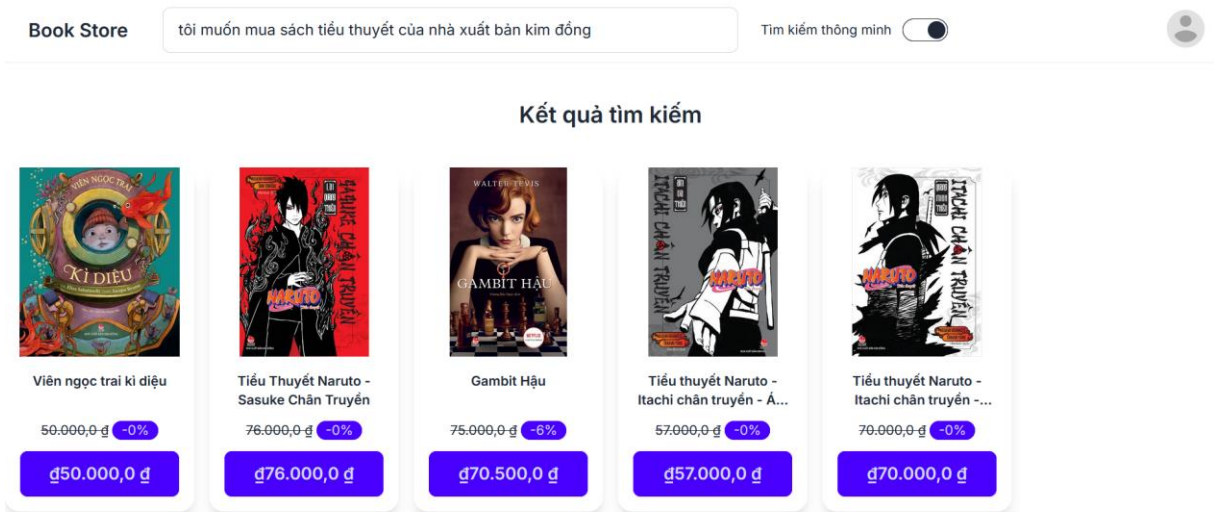
Hình 4.12. Test lấy kết quả đề xuất

### Bước 6: Hiển thị kết quả đề xuất lên giao diện website bán sách

Ta sẽ dùng FastAPI để website có thể lấy dữ liệu đề xuất sản phẩm thông qua Web API.

```
29 import warnings
28
27 from fastapi import FastAPI
26 from fastapi.middleware.cors import CORSMiddleware
25
24 from knn_model import KnnModel
23
22 warnings.filterwarnings("ignore")
21
20 knn_model = KnnModel()
19 knn_model.train_model("./models/knn.model")
18 knn_model.load_model("./models/knn.model")
17
16 /
15
14
13 @app.get("/recommend") new *
12 def rec(keyword: str, k: int = 5):
11     data = knn_model.recommend(keyword, k)
10     return data
9
8
```

Hình 4.13. Dùng FastAPI để trả kết quả qua API



Hình 4.14. Hiển thị kết quả lên website

#### 4.1.4. Cải thiện hệ thống đề xuất sản phẩm

- Thu thập thêm nhiều dữ liệu thực tế, mở rộng dữ liệu bằng cách thêm các thuộc tính mới cho sách.
- Kết hợp với các kỹ thuật đề xuất khác như lọc cộng tác (collaborative filtering) để cá nhân hoá kết quả đề xuất sản phẩm.

## 4.2. Giao diện của khách hàng

### 4.2.1. Giao diện đăng ký tài khoản

**Tạo tài khoản**

Họ đệm


Tên


Email

Mật khẩu

Nhập lại mật khẩu

Hình đại diện

 Chọn hình



**Tạo tài khoản**

Hình 4.15. Giao diện đăng ký

#### 4.2.2. Giao diện đăng nhập

## Đăng nhập

Email

Email

Mật khẩu

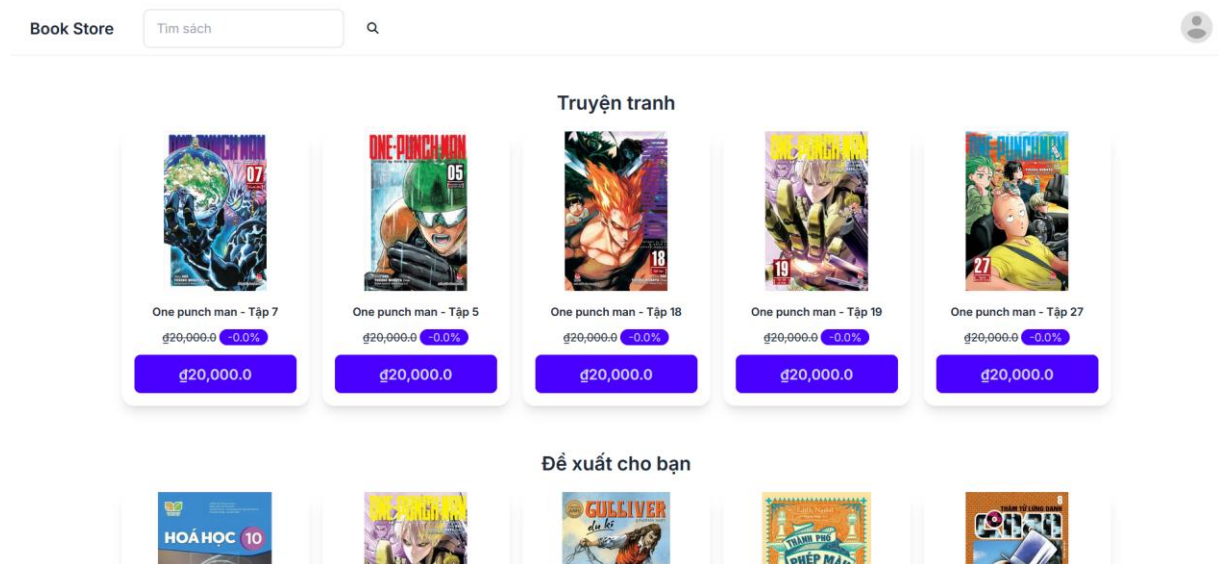
Mật khẩu

Ghi nhớ đăng nhập ☐

Đăng nhập

Hình 4.16. Giao diện đăng nhập

#### 4.2.3. Trang chủ



Hình 4.17. Giao diện trang chủ

### 4.2.4. Giao diện thông tin tài khoản

Book Store

Tim sách

Tim kiếm thông minh

Chỉnh sửa thông tin cá nhân

Mã tài khoản

3

Họ đệm

test

Tên

test

Email

a@b

Vai trò

Admin

Mật khẩu

(Để trống nếu không đổi mật khẩu)

Mật khẩu

Hình 4.18. Giao diện thông tin tài khoản

### 4.2.5. Giao diện chi tiết sách

Book Store

Tim sách

Sed elit dolor aliqua tempor lorem sit sit tempor lorem incididunt incididunt labore.

Elit sit lorem aliqua et et et eiusmod sed eiusmod dolor. Elit tempor ipsum elit eiusmod amet dolor magna. Aliqua elit magna labore eiusmod incididunt elit tempor. Amet consectetur tempor adipiscing do.

Thông tin sản phẩm

Nhà xuất bản

NXB Kim Đồng

Thể loại

Manga

Giả tưởng

Tác giả

Naruto - Tập 1

Dolor labore incididunt eiusmod magna labore. Dipiscing ipsum dolore amet eiusmod aliqua et.

Số hàng còn lại: 10

20,000.0đ

20,000.0đ

-0.0%

Số lượng

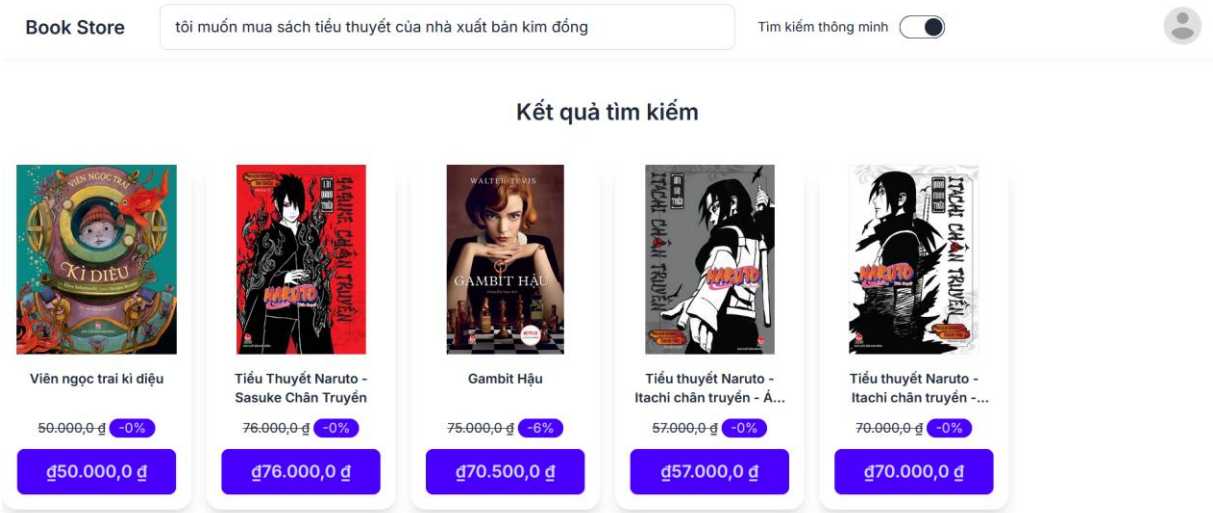
1

Thêm vào giỏ hàng

Hình 4.19. Giao diện chi tiết sách

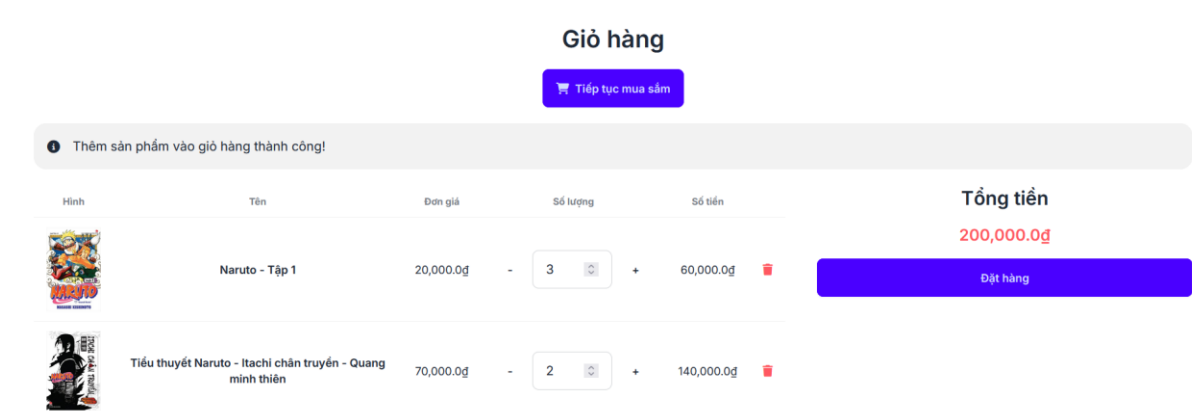


### 4.2.6. Giao diện tìm kiếm sách thông minh



Hình 4.20. Giao diện tìm kiếm sách

### 4.2.7. Giao diện giỏ hàng



Hình 4.21. Giao diện giỏ hàng

### 4.2.8. Giao diện xác nhận đơn hàng

Book Store

Tìm sách

Q

Xác nhận đơn hàng

Thông tin giao hàng

Sản phẩm

Họ và tên \*

Nguyễn Văn A

Số điện thoại \*

0123456789

Tỉnh \*

Tỉnh An Giang

Quận/huyện \*

Huyện Tri Tôn

Phường/xã \*

Xã Tân Tuyến

Số nhà, tên đường \*

78 Lê Văn Việt

Tên sản phẩm

Số lượng

Giá

Naruto - Tập 1

3

60,000.0đ

Tiểu thuyết Naruto - Itachi chân truyền - Quang minh thiên

2

140,000.0đ

Tổng tiền sản phẩm: 200,000.0đ

Phí vận chuyển: 21,000.0đ

Tổng tiền: **221,000.0đ**

Xác nhận đặt hàng

Hình 4.22. Giao diện xác nhận đơn hàng

### 4.2.9. Giao diện quản lý đơn đã đặt

Tìm đơn hàng

Q

Tất cả		Đã đặt		Đang giao		Đã giao		Đã hủy
Mã đơn	Ngày đặt	Tổng tiền	Trạng thái	Phường/xã	Quận/huyện	Tỉnh		
487	20-11-2024 23:31	221,000.0đ	Đã đặt	Xã Tân Tuyến	Huyện Tri Tôn	Tỉnh An Giang		
63	10-11-2024 12:53	17,000.0đ	Đã hủy	Xã Tân Thuận	Huyện Đầm Dơi	Tỉnh Cà Mau		
61	03-11-2024 12:53	23,000.0đ	Đã giao	Xã Tân Thuận	Huyện Đầm Dơi	Tỉnh Cà Mau		
59	30-10-2024 12:53	78,000.0đ	Đã đặt	Xã Thọ Sơn	Huyện Bù Đăng	Tỉnh Bình Phước		
62	26-10-2024 12:53	110,000.0đ	Đã đặt	Xã Đại Mạch	Huyện Đông Anh	Thành phố Hà Nội		
33	25-10-2024 15:24	94,000.0đ	Đã đặt	Xã Thới Xuân	Huyện Cờ Đỏ	Thành phố Cần Thơ		
32	19-10-2024 15:24	36,000.0đ	Đã hủy	Xã Mỹ Thạnh	Huyện Ba Tri	Tỉnh Bến Tre		
60	19-10-2024 12:53	94,000.0đ	Đã đặt	Xã Tân Lập	Huyện Bắc Tân Uyên	Tỉnh Bình Dương		

Hình 4.23. Giao diện quản lý đơn

### 4.2.10. Giao diện chi tiết đơn hàng

Chi tiết đơn hàng

Thông tin giao hàng

Sản phẩm

Mã đơn: 487

Ngày đặt: 20-11-2024 23:31

Ngày dự kiến giao: 27-11-2024

Trạng thái: Đã đặt

Mã khách hàng: 15

Tên người nhận: Nguyễn Văn A

Số điện thoại người nhận: 0123456789

Địa chỉ nhận hàng: 78 Lê Văn Việt, Xã Tân Tuyến, Huyện Tri Tôn, Tỉnh An Giang

Tên sản phẩm

Số lượng

Giá

Tiểu thuyết Naruto - Itachi chân truyền - Quang minh thiên

2

140,000.0đ

Naruto - Tập 1

3

60,000.0đ

Tổng tiền sản phẩm: 200,000.0đ

Phí vận chuyển: 21,000.0đ

Tổng tiền: **221,000.0đ**

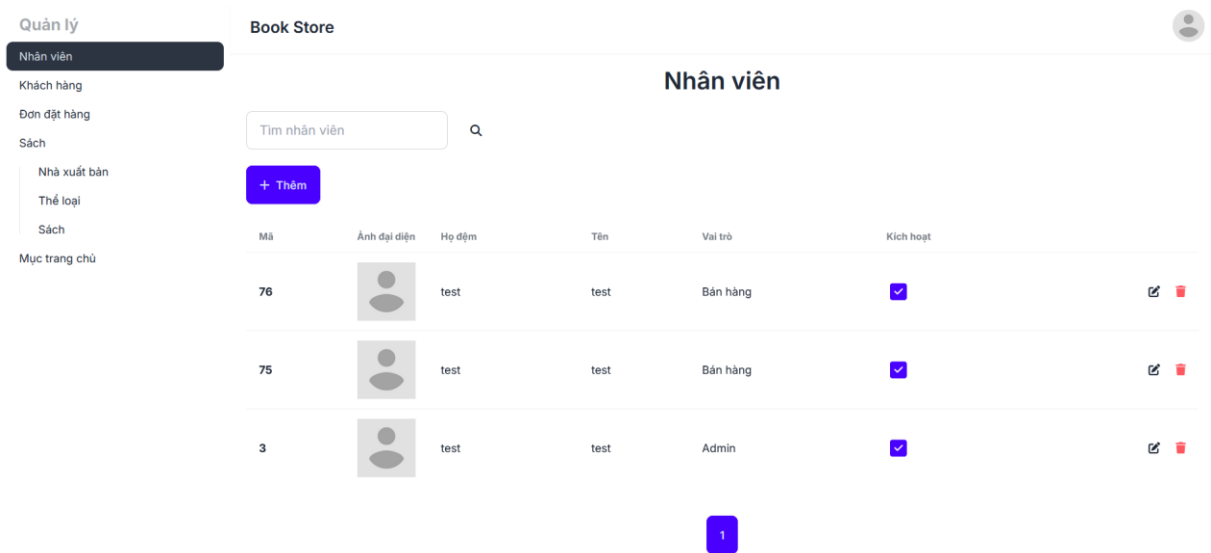
Hủy đơn hàng

Hình 4.24. Giao diện chi tiết đơn hàng

42

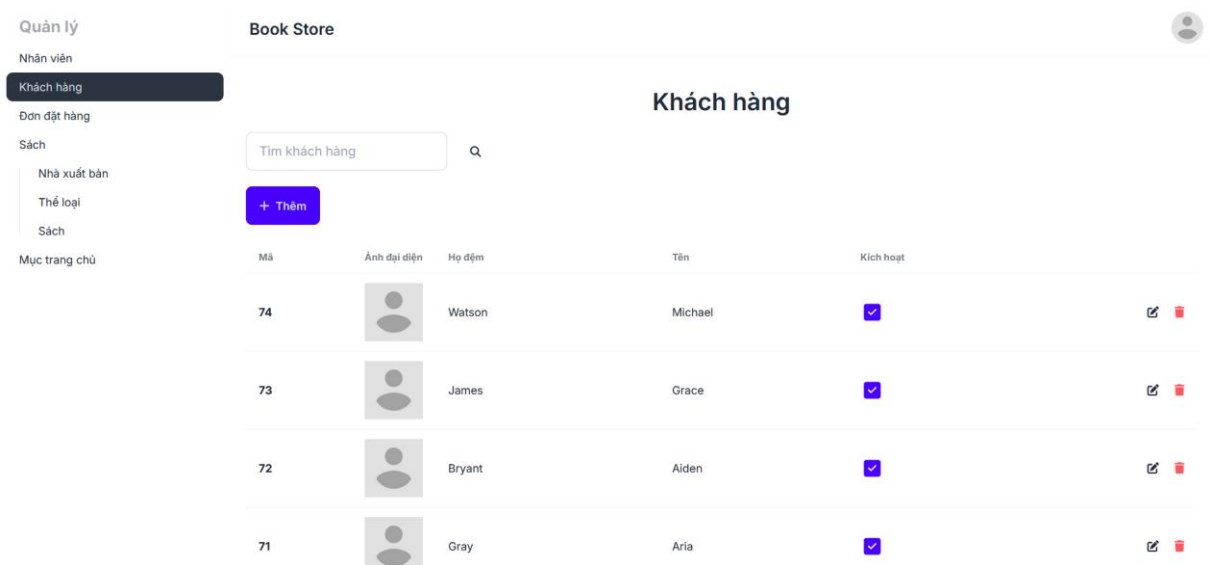
### 4.3. Giao diện của nhân viên

#### 4.3.1. Giao diện quản lý tài khoản nhân viên



Hình 4.25. Giao diện quản lý tài khoản nhân viên

#### 4.3.2. Giao diện quản lý tài khoản khách hàng



Hình 4.26. Giao diện quản lý tài khoản khách hàng

4.3.3. Giao diện quản lý đơn đặt hàng

Quản lý

Nhân viên

Khách hàng

Đơn đặt hàng

Sách

Nhà xuất bản

Thể loại

Sách

Mục trang chủ

Book Store

Đơn hàng

Tim đơn hàng

Q

Tất cả		Đã đặt	Đang giao	Đã giao	Đã hủy	
Mã đơn	Ngày đặt	Tổng tiền	Trạng thái	Phường/xã	Quận/huyện	Tỉnh
487	20-11-2024 23:31	221,000.0đ	Đã đặt	Xã Tân Tuyến	Huyện Trì Tôn	Tỉnh An Giang
470	16-11-2024 03:44	110,000.0đ	Đã đặt	00019	001	001
404	16-11-2024 03:39	59,000.0đ	Đã giao	00019	001	001
316	16-11-2024 03:34	30,000.0đ	Đã đặt	00019	001	001
174	15-11-2024 13:14	29,000.0đ	Đã hủy	Xã Ngũ Đoan	Huyện Kiến Thụy	Thành phố Hải Phòng
145	15-11-2024 13:08	59,000.0đ	Đang giao	Xã Thới Xuân	Huyện Cờ Đỏ	Thành phố Cần Thơ
136	15-11-2024 13:08	44,000.0đ	Đang giao	Xã Đại Mạch	Huyện Đông Anh	Thành phố Hà Nội
427	15-11-2024 03:42	94,000.0đ	Đã đặt	00019	001	001
406	15-11-2024 03:39	109,000.0đ	Đã giao	00019	001	001

Hình 4.27. Giao diện quản lý đơn đặt hàng

4.3.4. Giao diện quản lý sách

Quản lý

Nhân viên

Khách hàng

Đơn đặt hàng

Sách

Nhà xuất bản

Thể loại

Sách

Mục trang chủ

Book Store

+ Thêm

Mã	Hình	Tên sách	Nhà xuất bản	Niên yết	
92		Alice ở xứ sở thần tiên	Dân Trí	<input checked="" type="checkbox"/>	
98		Cái tết của mèo con	NXB Trẻ	<input checked="" type="checkbox"/>	
113		Đảo giấu vàng	NXB Trẻ	<input checked="" type="checkbox"/>	
107		Dragon Ball - Tập 19	NXB Kim Đồng	<input checked="" type="checkbox"/>	

Hình 4.28. Giao diện quản lý sách

4.3.5. Giao diện thêm sách

Quản lý

Nhân viên

Khách hàng

Đơn đặt hàng

Sách

Nhà xuất bản

Thể loại

Sách

Mục trang chủ

Book Store

Thêm sách mới

Tổng quát

Tác giả

Thể loại

Mô tả

Hình ảnh

Thông số

Tên sách \*

Tên sách

Nhà xuất bản \*

Dân Trí

Độ tuổi \*

Độ tuổi

Số trang \*

Số trang

Ngày xuất bản \*

mm / dd / yyyy

Hình 4.29. Giao diện thêm sách

4.3.6. Giao diện chỉnh sửa trang chủ

Quản lý

Nhân viên

Khách hàng

Đơn đặt hàng

Sách

Nhà xuất bản

Thể loại

Sách

Mục trang chủ

Book Store

Mục

+ Thêm mục

Mã	Tiêu đề	Loại	Thứ tự	Hiển thị	
2	Đề xuất cho bạn	Đề xuất	1	<input checked="" type="checkbox"/>	^ v ✎ 🗑
1	Truyện tranh	Sách	2	<input checked="" type="checkbox"/>	^ v ✎ 🗑
3	Thể loại sách hot	Loại	3	<input checked="" type="checkbox"/>	^ v ✎ 🗑

Hình 4.30. Giao diện chỉnh sửa trang chủ

## CHƯƠNG 5: DEMO

### Demo 1:

User input: Tôi muốn mua sách tiểu thuyết của nxb kim đồng

Output:

Book title	Publisher name	Category 1	Category 2	Category 3	Author 1	Author 2
Tiểu thuyết Naruto - Itachi chân truyền - Âm dạ thiên	NXB Kim Đồng	Giả tưởng	Nước ngoài	Tiểu thuyết	Masashi Kishimoto	missing
Tiểu thuyết Naruto - Itachi chân truyền - Quang minh thiên	NXB Kim Đồng	Giả tưởng	Nước ngoài	Tiểu thuyết	Masashi Kishimoto	missing
Gambit Hậu	NXB Kim Đồng	Nước ngoài	Tiểu thuyết	missing	Walter Tevis	missing
Viên ngọc trai kì diệu	NXB Kim Đồng	Nước ngoài	Tiểu thuyết	missing	Elisa Sabatinelli	missing
Tiểu Thuyết Naruto - Sasuke Chân Truyền	NXB Kim Đồng	Giả tưởng	Light novel	Nước ngoài	Masashi Kishimoto	Shin Towada

Bảng 5.1. Output demo 1

Book Store

Tôi muốn mua sách tiểu thuyết của nxb kim đồng

Tìm kiếm thông minh

Kết quả tìm kiếm

Viên ngọc trai kì diệu

50.000,0 đ -0%

₫50.000,0 đ

Tiểu Thuyết Naruto - Sasuke Chân Truyền

76.000,0 đ -0%

₫76.000,0 đ

Gambit Hậu

75.000,0 đ -6%

₫70.500,0 đ

Tiểu thuyết Naruto - Itachi chân truyền - Âm...

57.000,0 đ -0%

₫57.000,0 đ

Tiểu thuyết Naruto - Itachi chân truyền - ...

70.000,0 đ -0%

₫70.000,0 đ

Hình 5.1. Hiển thị output demo 1

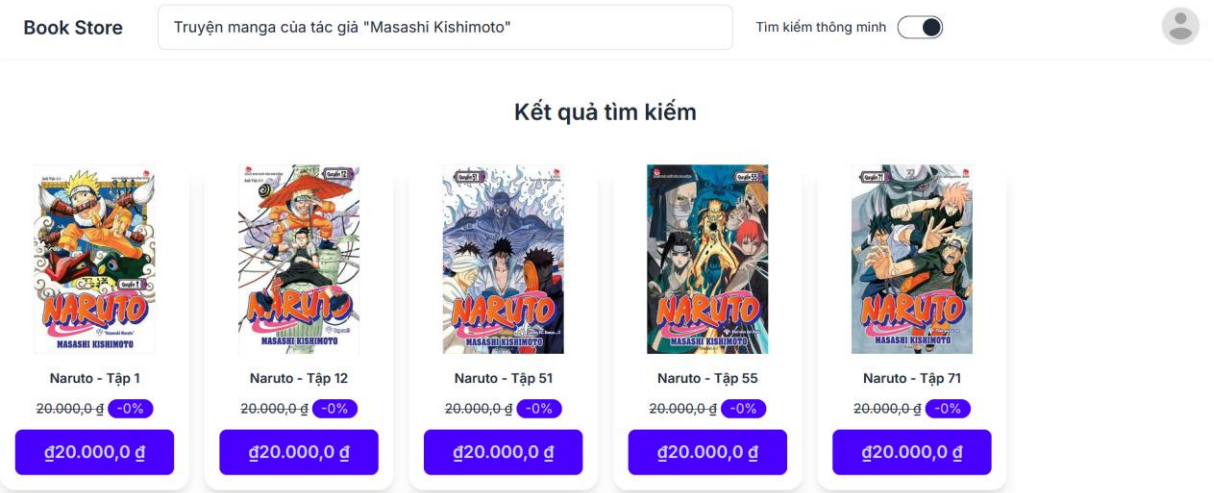
**Demo 2:**

User input: Truyện manga của tác giả "Masashi Kishimoto"

Output:

Book title	Publisher name	Category 1	Category 2	Category 3	Author 1	Author 2
Naruto - Tập 1	NXB Kim Đồng	Giả tưởng	Manga	missing	Masashi Kishimoto	missing
Naruto - Tập 12	NXB Kim Đồng	Giả tưởng	Manga	missing	Masashi Kishimoto	missing
Naruto - Tập 71	NXB Kim Đồng	Giả tưởng	Manga	missing	Masashi Kishimoto	missing
Naruto - Tập 55	NXB Kim Đồng	Giả tưởng	Manga	missing	Masashi Kishimoto	missing
Naruto - Tập 51	NXB Kim Đồng	Giả tưởng	Manga	missing	Masashi Kishimoto	missing

Bảng 5.2. Output demo 2



Hình 5.2. Hiển thị output demo 2

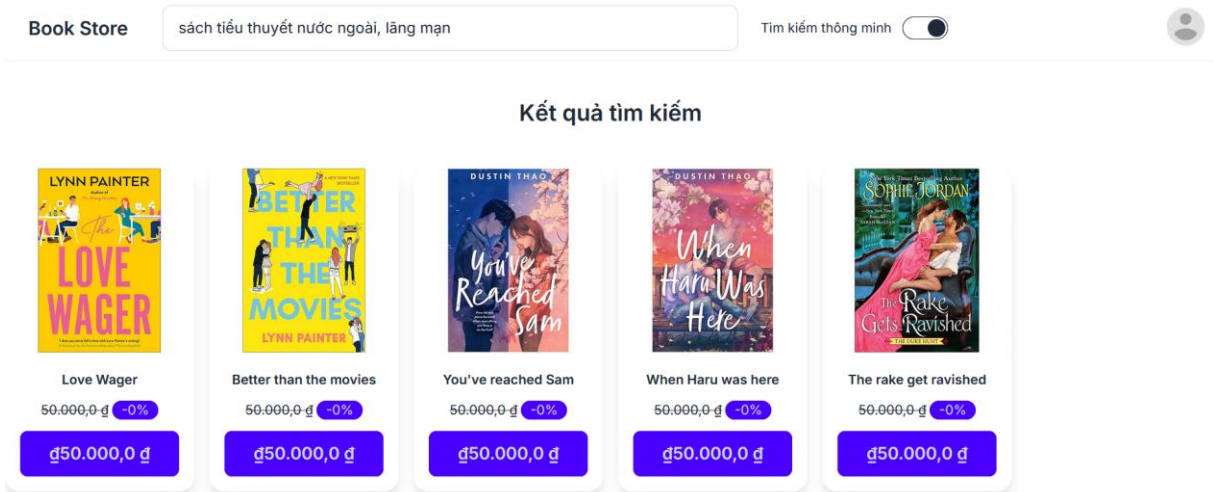
**Demo 3:**

User input: Sách tiểu thuyết nước ngoài, lãng mạn

Output:

Book title	Publisher name	Category 1	Category 2	Category 3	Author 1	Author 2
Love Wager	NXB Trẻ	Lãng mạn	Nước ngoài	Tiểu thuyết	Lynn Painter	missing
When Haru was here	NXB Trẻ	Lãng mạn	Nước ngoài	Tiểu thuyết	Dustin Thao	missing
Better than the movies	NXB Trẻ	Lãng mạn	Nước ngoài	Tiểu thuyết	Lynn Painter	missing
The rake get ravished	NXB Kim Đồng	Lãng mạn	Nước ngoài	Tiểu thuyết	Sophie Jordan	missing
You've reached Sam	NXB Tổng Hợp TPHCM	Lãng mạn	Nước ngoài	Tiểu thuyết	Dustin Thao	missing

Bảng 5.3. Output demo 3



Hình 5.3. Hiển thị output demo 3

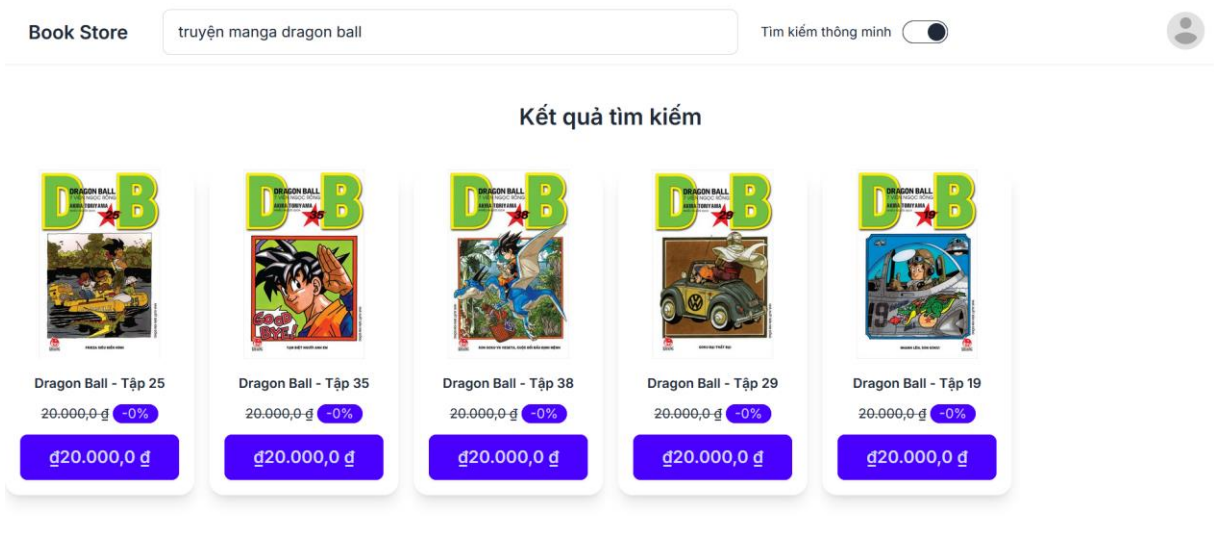
#### Demo 4:

User input: Truyện manga dragon ball

Output:

Book title	Publisher name	Category 1	Category 2	Category 3	Author 1	Author 2
Dragon Ball - Tập 29	NXB Kim Đồng	Giả tưởng	Manga	missing	missing	missing
Dragon Ball - Tập 25	NXB Kim Đồng	Giả tưởng	Manga	missing	missing	missing
Dragon Ball - Tập 38	NXB Kim Đồng	Giả tưởng	Manga	missing	missing	missing
Dragon Ball - Tập 19	NXB Kim Đồng	Giả tưởng	Manga	Nước ngoài	missing	missing
Dragon Ball - Tập 35	NXB Kim Đồng	Giả tưởng	Manga	missing	missing	missing

Bảng 5.4. Output demo 4



Hình 5.4. Hiển thị output demo 4



## CHƯƠNG 6: KẾT LUẬN

### 6.1. Đánh giá

- Giao diện: Giao diện hiển thị đầy đủ và chính xác.
- Phân quyền: đảm bảo đúng quyền mới được truy cập, đường dẫn chuyển tới trang muốn tới.
- Thêm / sửa sản phẩm: Sản phẩm được thêm với đúng thông tin đã nhập, có kiểm tra thông tin đã nhập, các sản phẩm được chỉnh sửa được lưu với đúng thông tin đã nhập, có hiển thị thông báo lỗi hoặc lưu thành công.
- Quản lý tài khoản người dùng: Nhân viên có thể chỉnh sửa thông tin tài khoản khách hàng và nhân viên, có thể thêm tài khoản nhân viên mới.
- Tìm kiếm sản phẩm: Chức năng tìm kiếm sản phẩm hoạt động tốt, có thể tìm kiếm theo tên sản phẩm, tên thể loại hoặc tên tác giả.
- Quy trình đặt hàng: Giỏ hàng khách hàng hoạt động tốt, có thể chỉnh sửa số lượng và xóa sản phẩm khỏi giỏ hàng. Chức năng đặt hàng hoạt động tốt, đơn hàng được tạo với đúng thông tin giao hàng đã nhập, đúng sản phẩm và số lượng, giá cả.
- Đề xuất sản phẩm: Giao diện cho người dùng nhập mô tả sản phẩm hoạt động tốt, website có thể lấy kết quả đề xuất và hiển thị đầy đủ cho người dùng, khi người dùng nhấn vào một sản phẩm trong kết quả được hiển thị thì sẽ được điều hướng sang trang chi tiết sản phẩm.

### 6.2. Kết luận

Đề tài “Xây dựng hệ thống đề xuất sản phẩm cho website bán sách” đã đạt được những mục tiêu đề ra, bao gồm tìm hiểu nội dung lý thuyết về xây dựng phát triển hệ thống thông minh, học máy, hệ thống đề xuất, thuật toán kNN, và nghiên cứu phát triển hệ thống website bán sách, cũng như hệ thống đề xuất sản phẩm dựa vào mô tả của người dùng. Thông qua việc ứng dụng công nghệ học máy, hệ thống đã phân tích mô tả sản phẩm của khách hàng nhằm cung cấp các gợi ý sản phẩm có liên quan, nâng cao trải nghiệm người dùng.

Về mặt lý thuyết, bài báo cáo đã giới thiệu về học máy, hệ thống thông minh, hệ thống đề xuất, xử lý ngôn ngữ tự nhiên và thuật toán K-Nearest Neighbors.

Về mặt dữ liệu, kết quả dự trên bộ dữ liệu huấn luyện cho ra kết quả tương đối chính xác. Tuy nhiên, để nâng cao khả năng dự đoán của thuật toán, ta nên tiếp tục thu thập thêm nhiều dữ liệu thực tế, mở rộng dữ liệu bằng các thuộc tính khác cũng như tối ưu thuật toán để có kết quả tốt hơn.

Về phần ứng dụng, website có giao diện đơn giản, thân thiện, người dùng dễ sử dụng. Website có khá đầy đủ chức năng của một trang web bán hàng, có phân quyền rõ ràng khách hàng và nhân viên.

Xây dựng hệ thống đề xuất sản phẩm cho website bán sách không chỉ là một bài toán kỹ thuật mà còn là chìa khóa để nâng cao chất lượng dịch vụ, tăng trưởng doanh thu và tạo sự khác biệt trong môi trường cạnh tranh khốc liệt của thương mại điện tử hiện nay. Chính vì vậy, nhóm đã chọn đề tài này nhằm mục đích nghiên cứu, ứng dụng công nghệ hiện đại vào thực tiễn cũng như nâng cao hiểu biết của bản thân về học máy và hệ thống đề xuất sản phẩm.

Trong quá trình thực hiện bài báo cáo, nhóm đã tập trung tìm hiểu và tham khảo các tài liệu và kiến thức liên quan. Tuy nhiên, với thời gian và trình độ có hạn nên không tránh khỏi những hạn chế và thiếu sót. Nhóm rất mong nhận được các nhận xét và góp ý của thầy để hoàn thiện hơn về kiến thức.