

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH  
KHÓA CÔNG NGHỆ THÔNG**



**KHÓA LUẬN TỐT NGHIỆP**

**PHÂN TÍCH VÀ XÂY DỰNG HỆ THỐNG  
QUẢN LÝ CỬA HÀNG  
DỰA TRÊN GIẢI THUẬT FP-GROWTH**

**Giảng viên hướng dẫn : ThS. BÙI TIẾN ĐỨC**  
**Sinh viên thực hiện : NGUYỄN BÉ LAM**  
**MSSV : 2000001437**  
**Khóa : 2020**  
**Ngành/ chuyên ngành : TRÍ TUỆ NHÂN TẠO**

Tp HCM, ngày 5 tháng 1 năm 2024

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH  
KHÓA CÔNG NGHỆ THÔNG**



**KHÓA LUẬN TỐT NGHIỆP**

**PHÂN TÍCH VÀ XÂY DỰNG HỆ THỐNG  
QUẢN LÝ CỬA HÀNG  
DỰA TRÊN GIẢI THUẬT FP-GROWTH**

<b>Giảng viên hướng dẫn</b>	<b>: ThS. BÙI TIẾN ĐỨC</b>
<b>Sinh viên thực hiện</b>	<b>: NGUYỄN BÉ LAM</b>
<b>MSSV</b>	<b>: 2000001437</b>
<b>Khóa</b>	<b>: 2020</b>
<b>Ngành/ chuyên ngành</b>	<b>: TRÍ TUỆ NHÂN TẠO</b>

Tp HCM, ngày 5 tháng 1 năm 2024

## NHIỆM VỤ KHÓA LUẬN TỐT NGHIỆP

**HỌ VÀ TÊN :** NGUYỄN BÉ LAM ..... **MSSV:** 2000001437

**CHUYÊN NGÀNH :** TRÍ TUỆ NHÂN TẠO ..... **LỚP :** 20DTH1A

**Điện thoại:** 0825456076

**Email:** nlam020902@gmail.com

1. Tên đề tài: Phân tích và xây dựng hệ thống quản lý cửa hàng dựa trên FP-growth
2. Giảng viên hướng dẫn: ThS.Bùi Tiến Đức
3. Nhiệm vụ / nội dung đề tài (mô tả chi tiết nội dung, yêu cầu, phương pháp... ) :
  - Thu thập và tiền xử lý dữ liệu từ hóa đơn giao dịch bán lẻ.
  - Phân tích đặc điểm và xu hướng mua sắm từ dữ liệu.
  - Xây Dựng Mô Hình FP-growth
  - Sử dụng thuật toán FP-growth để khai thác mẫu tần suất từ dữ liệu mua sắm
  - Xây dựng mô hình dự đoán các mẫu mua sắm và tạo ra danh sách các sản phẩm liên quan.
  - Xây dựng giao diện người dùng bằng Django để tương tác với mô hình và hiển thị gợi ý sản phẩm.
  - Tích hợp mô hình FP-growth vào hệ thống sử dụng Django.
  - Viết báo cáo chi tiết về kết quả của đề tài và phân tích ý nghĩa của các mô hình gợi ý sản phẩm.
  - Sử dụng phương pháp nghiên cứu thực nghiệm, tiếp cận từng bước để xây dựng và đánh giá mô hình gợi ý sản phẩm. Sử dụng ngôn ngữ lập trình Python, Django Framework, và thư viện máy học.
4. Thời gian thực hiện từ: 29/10/2023 .....đến: 01/01/2024 .....

Nội dung và yêu cầu đồ án tốt nghiệp đã được thông qua Bộ môn.

*Ngày 05 tháng 01 năm 2024*

**TRƯỞNG BỘ MÔN**

*(Ký và ghi rõ họ tên)*

**GIÁO VIÊN HƯỚNG DẪN**

*(Ký và ghi rõ họ tên)*

## LỜI CẢM ƠN

Em xin bày tỏ lòng biết ơn chân thành đến tất cả những người đã đóng góp cho sự thành công của khóa luận này.

Trước hết, em xin gửi lời cảm ơn sâu sắc đến giáo viên hướng dẫn của em, Thầy Bùi Tiến Đức, với lòng tận tâm và kiến thức sâu sắc đã hỗ trợ em từ giai đoạn lập kế hoạch cho đến hoàn thiện khóa luận. Sự hướng dẫn và góp ý của ngài đã giúp em phát triển khả năng nghiên cứu và hiểu rõ hơn về lĩnh vực này.

Em cũng muốn bày tỏ lòng biết ơn đến Kaggle là một nền tảng trực tuyến, nơi em đã có cơ hội tiếp cận với dữ liệu và nguồn tài nguyên quan trọng, đồng thời nhận được sự hỗ trợ tích cực từ cộng đồng làm việc.

Không thể không nhắc đến sự đóng góp của bạn bè và gia đình, những người đã luôn ở bên cạnh và cung cấp hỗ trợ tinh thần. Sự khích lệ và hiểu biết từ họ đã giúp em vượt qua những thời kỳ khó khăn và tạo điều kiện thuận lợi cho quá trình nghiên cứu.

Cuối cùng, em muốn bày tỏ lòng biết ơn đến tất cả những người đã đọc và đánh giá dự án này. Sự quan tâm và phản hồi của mọi người là nguồn động viên quý báu.

Trân trọng

Nguyễn Bé Lam

## LỜI MỞ ĐẦU

Ngày càng tăng cường của thế giới kỹ thuật số đã mở ra những cơ hội đối với thương mại điện tử và phân tích dữ liệu. Việc hiểu rõ mẫu mua sắm của khách hàng và đề xuất sản phẩm một cách hiệu quả là một yếu tố quyết định đối với sự thành công của các doanh nghiệp trực tuyến. Trong bối cảnh này, sự kết hợp giữa thuật toán FP-growth và framework Django không chỉ mang lại sự tiện lợi trong xử lý dữ liệu mà còn mở ra những cánh cửa mới cho việc xây dựng các hệ thống gợi ý sản phẩm.

Khóa luận này tập trung vào việc phân tích mẫu mua sắm của khách hàng dựa trên dữ liệu hóa đơn giao dịch bán lẻ và triển khai mô hình gợi ý sản phẩm sử dụng thuật toán FP-growth và Django Framework. Chúng tôi đặt ra câu hỏi: "Làm thế nào chúng ta có thể hiểu rõ hơn về hành vi mua sắm của khách hàng và tối ưu hóa gợi ý sản phẩm để tăng cường trải nghiệm mua sắm trực tuyến?"

Trong phạm vi của khóa luận này, em sẽ trình bày quy trình phân tích dữ liệu, lựa chọn thuật toán FP-growth và triển khai mô hình gợi ý sản phẩm thông qua Django Framework. Bằng cách này, em hy vọng mang lại cái nhìn sâu sắc và thực tế lợi ích từ phân tích dữ liệu mua sắm và ứng dụng thuật toán khai thác mẫu FP-growth.

Khám phá sự liên kết giữa các sản phẩm, hiểu rõ hơn về nhu cầu của khách hàng, và đề xuất sản phẩm một cách linh hoạt. Em mong rằng công trình này sẽ là một đóng góp cho lĩnh vực nghiên cứu khoa học dữ liệu mà còn là một bước quan trọng trong việc ứng dụng nguyên lý khoa học dữ liệu vào thực tế thương mại.

Chân thành cảm ơn sự hỗ trợ và hướng dẫn từ Thầy Bùi Tiến Đức trong quá trình nghiên cứu và triển khai dự án này.

Em hy vọng rằng công trình này sẽ mang lại những thông điệp hữu ích và kích thích sự quan tâm trong lĩnh vực quan tâm.

## **NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

1. Hình thức (Bố cục, trình bày, lỗi, các mục, hình, bảng, công thức, phụ lục, ...) ....

.....

.....

.....

.....

.....

.....

.....

2. Nội dung (mục tiêu, phương pháp, kết quả, sao chép, các chương, tài liệu, ...).....

.....

.....

.....

.....

.....

3. Kết luận .....

.....

Tp HCM, ngày 5 tháng 1 năm 2024

**Giáo viên hướng dẫn**

(Ký tên, ghi rõ họ tên)

**NHẬN XÉT CỦA GIẢNG VIÊN PHẢN BIỆN**

1. Hình thức (Bố cục, trình bày, lỗi, các mục, hình, bảng, công thức, phụ lục, ) ...

.....

.....

.....

.....

.....

.....

.....

2. Nội dung (mục tiêu, phương pháp, kết quả, sao chép, các chương, tài liệu,..)

.....

.....

.....

.....

.....

.....

3. Kết luận.....

.....

Tp HCM, ngày 5 tháng 1 năm 2024

**Giáo viên phản biện**  
(Ký tên,ghi rõ họ tên)

# MỤC LỤC

LỜI CẢM ƠN .....	ii
LỜI MỞ ĐẦU .....	iv
NHẬN XÉT CỦA GIẢNG VIÊN PHẢN BIỆN .....	vi
MỤC LỤC .....	vii
DANH MỤC HÌNH .....	ix
DANH MỤC BẢNG .....	x
KÍ HIỆU CÁC CỤM TỪ VIẾT TẮT .....	xi
<b>Chương 1: GIỚI THIỆU CHUNG VỀ ĐỀ TÀI .....</b>	<b>1</b>
1.1 Lý do chọn đề tài .....	1
1.2 Mục tiêu nghiên cứu .....	1
1.3 Đối tượng nghiên cứu .....	1
1.4 Phạm vi nghiên cứu .....	2
1.5 Ý nghĩa của đề tài .....	2
1.6 Bố cục đề tài .....	2
<b>Chương 2: CƠ SỞ LÝ LUẬN VỀ VẤN ĐỀ NGHIÊN CỨU .....</b>	<b>4</b>
2.1 Lý thuyết liên quan .....	4
2.1.1 Django .....	4
2.1.2 Data Mining .....	5
2.1.3 Thuật toán FP-Growth .....	6
2.2 Nghiên cứu liên quan .....	10
<b>Chương 3: MÔ HÌNH LÝ THUYẾT .....</b>	<b>11</b>
3.1 Mô tả bài toán .....	11
3.2 Mô tả hệ thống mức cao .....	12
3.3 Mô tả công nghệ .....	13
3.4 Mô hình toán .....	14
<b>Chương 4: THỰC NGHIỆM .....</b>	<b>15</b>
4.1 Thiết kế giao diện .....	15
4.2 Mô tả database .....	21
4.3 Mô tả dữ liệu .....	22
4.4 Kết quả .....	24
4.4.1 Xử lý dữ liệu .....	24
4.4.2 Thuật toán .....	25
4.5 Phân tích và nhận xét .....	26
<b>Chương 5: KẾT LUẬN VÀ KIẾN NGHỊ .....</b>	<b>27</b>



<b>DANH MỤC TÀI LIỆU THAM KHẢO.....</b>	<b>29</b>
---	-----------

## DANH MỤC HÌNH

Hình 2. 1 Mô hình MVT của Django.....	4
Hình 2. 2 Ví dụ cây FP-Tree .....	9
Hình 3. 1 Sơ đồ hệ thống mức cao .....	12
Hình 4. 3 Phần slider 1 .....	16
Hình 4. 4 Phần slider 2 .....	16
Hình 4. 5 Giao diện trang chủ sản phẩm.....	16
Hình 4. 6 Giao diện trang chủ tình tức.....	17
Hình 4. 9 Giao diện danh sách tin tức .....	18
Hình 4. 10 Giao diện liên hệ .....	18
Hình 4. 11 Giao diện đăng nhập.....	19
Hình 4. 12 Giao diện đăng kí .....	19
Hình 4. 13 Giỏ hàng mini.....	20
Hình 4. 14 Giao diện giỏ hàng .....	20
Hình 4. 15 Giao diện trang thanh toán .....	21
Hình 4. 16 Cơ sở dữ liệu quan hệ.....	21
Hình 4. 17 Mô tả bộ dữ liệu .....	23
Hình 4. 19 Chuyển hóa dữ liệu .....	24
Hình 4. 20 Kết quả của thuật toán FP - GROWTH .....	25
Hình 4. 21 Các mẫu phổ biến.....	25
Hình 4. 22 Kết quả gợi ý sao khi dự đoán.....	26

## DANH MỤC BẢNG

Bảng 2.1 Ví dụ về cây FP-TREE .....	7
Bảng 2.2 Ví dụ về bảng Support item.....	8
Bảng 2.3 Ví dụ về bảng giao dịch item.....	8
Bảng 2.4 Ví dụ bảng Header .....	8
Bảng 2.5 Conditional pattern base .....	9
Bảng 2.6 Bảng kết quả .....	9

## KÍ HIỆU CÁC CỤM TỪ VIẾT TẮT

Chữ viết tắt	Ý nghĩa
FP-growth	Frequent Pattern growth
AI	artificial intelligence
Fp-tree	Frequent Pattern tree
ORM	Object-Relational Mapping

# GIỚI THIỆU CHUNG VỀ ĐỀ TÀI

## 1.1 Lý do chọn đề tài

Thương mại điện tử là một lĩnh vực mà hiện tại đang rất được ưu chuộng. Nó đem lại rất nhiều lợi ích và sự thuận tiện cho người tiêu dùng. Thương mại điện tử là tạo điều kiện thuận lợi cho sự mua bán, trao đổi hay giao nhận các sản phẩm, dịch vụ thông qua internet hay các thiết bị điện tử khác giữa người mua và người bán. Nó được phát minh từ những năm 70 và bởi vì sự hữu ích nó mang lại, cho đến bây giờ trở nên rất phổ biến và phát triển. Đây là một hình thức kinh doanh rất hiệu quả đối với doanh nghiệp hay cá nhân. Chúng ta có thể tận dụng từ những dữ liệu đã có trong thương mại điện tử để mang lại nhiều lợi ích hơn, nâng cao lợi nhuận và trải nghiệm người dùng.

Để đạt hiệu quả hơn trong lĩnh vực này, chúng ta có thể sử dụng từ nguồn dữ liệu khổng lồ của thương mại điện tử. Ứng dụng sự thông minh của trí tuệ nhân tạo để khai thác dữ liệu và tạo ra thêm nhiều dịch vụ giúp nâng cao sản phẩm cho khách hàng. Khai thác dữ liệu là máy tính có thể được sử dụng để hỗ trợ phân tích, khám phá ra những mối hệ tìm ẩn trong dữ liệu. Sau khi trải qua quá trình xử lý thì sẽ đưa ra những kiến thức có ích và có thể giúp giải quyết vấn đề và tăng lợi nhuận kinh doanh của doanh nghiệp.

Dựa vào những nhu cầu, xu hướng phát triển và lợi ích trên, dự án em thực hiện là Phân tích và xây dựng hệ thống quản lý cửa hàng dựa trên giải thuật FP-Growth.

## 1.2 Mục tiêu nghiên cứu

Mục tiêu nghiên cứu của đề tài “Phân tích và xây dựng hệ thống quản lý cửa hàng dựa trên giải thuật FP-Growth” là:

- ❖ Tìm hiểu, phân tích nhu cầu của cửa hàng như hóa đơn bán ra, ngày tháng.
- ❖ Xây dựng được cơ sở dữ liệu.
- ❖ Tìm hiểu và xây dựng Django để phát triển web
- ❖ Xây dựng giao diện người dùng, các chức năng quản lý sản phẩm, đơn hàng,...
- ❖ Tích hợp giải thuật fp-growth để đưa ra gợi ý sản phẩm.

## 1.3 Đối tượng nghiên cứu

Đối tượng nghiên cứu trong đề tài này là:

- ❖ Người quản trị cửa hàng mỹ phẩm: quản lý các hoạt động của cửa hàng
- ❖ Nhân viên: Theo dõi và quản lý các sản phẩm, đơn hàng
- ❖ Khách hàng: Tìm hiểu, phân tích nhu cầu khách hàng từ các đơn hàng
- ❖ Dữ liệu: Dữ liệu từ các cuộc giao dịch của khách hàng và cửa hàng
- ❖ Thuật toán: FP-growth được áp dụng để đưa ra những quyết định hỗ trợ gợi ý sản phẩm.

#### **1.4 Phạm vi nghiên cứu**

Phạm vi nghiên cứu của đề tài là:

- ❖ Thời gian diễn ra của các đơn hàng
- ❖ Dữ liệu về các đơn hàng để áp dụng cho giải thuật fp-growth
- ❖ Đối tượng sử dụng hệ thống hướng đến: người quản trị, nhân viên, khách hàng

#### **1.5 Ý nghĩa của đề tài**

Đề tài có thể giúp quản lý hệ thống cửa hàng dễ dàng hơn, cải thiện được doanh thu, tăng lợi ích cho người kinh doanh và đáp ứng được nhu cầu sử dụng của khách hàng.

Giúp mang lại giá trị thực tế cho cửa hàng, có thể ứng dụng cho ngành bán lẻ và có thể mở rộng cho các hệ thống cửa hàng.

Đề tài này cũng giúp em học hỏi được thêm nhiều thứ, từ quá trình phân tích, xây dựng và phát triển hệ thống. Học hỏi và tích lũy nhiều kinh nghiệm quý báu trong quá trình thực hiện.

#### **1.6 Bố cục đề tài**

Chương 1: Giới thiệu chung về đề tài

Lý do chọn đề tài

Mục tiêu nghiên cứu

Đối tượng nghiên cứu

Phạm vi nghiên cứu

Ý nghĩa của đề tài

Chương 2: Cơ sở lý thuyết

2.1 Lý thuyết liên quan

2.1.1 Django

2.1.2 Data Mining

### 2.1.3 Thuật toán FP-growth

## 2.2 Các nghiên cứu liên quan

Chương 3: Mô hình lý thuyết

Chương 4: Mô hình Thực nghiệm

Chương 5: Kết luận

## Chương 2: CƠ SỞ LÝ LUẬN VỀ VẤN ĐỀ NGHIÊN CỨU

### 2.1 Lý thuyết liên quan

#### 2.1.1 Django

Django là một framework mã nguồn mở, được phát triển dựa trên ngôn ngữ lập trình Python. Django phát hành năm 2005 và được tổ chức phi lợi nhuận hỗ trợ duy trì là Django Software Foundation. Nó hỗ trợ xây dựng các ứng dụng web nhanh chóng, với các tính năng như đăng nhập, kết nối cơ sở dữ liệu và CRUD ( create read update delete)

Django hoạt động theo mô hình MVT (model – views – template).

##### Ưu điểm:

Đơn giản, nhanh chóng, dễ học, dễ thực hiện.

Nhiều thư viện hỗ trợ.

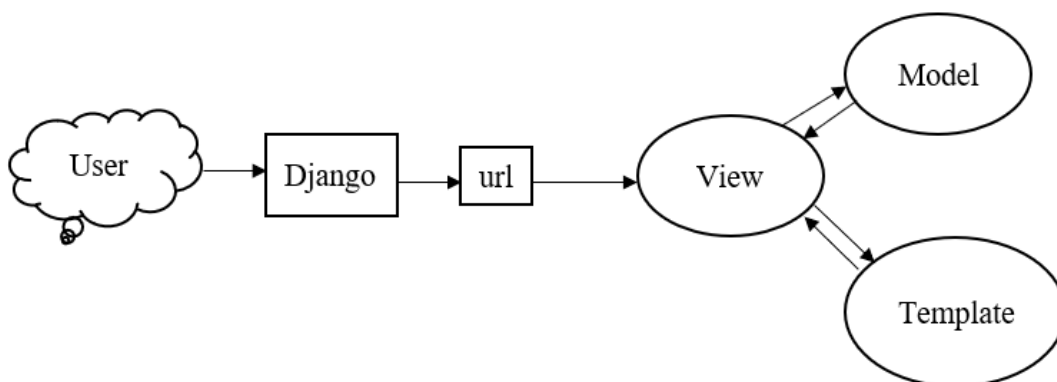
Bảo mật.

##### Nhược điểm:

Không phù hợp với các dự án nhỏ, không có khả năng thực hiện nhiều yêu cầu cùng lúc.

Khó khăn trong việc tìm ra lỗi.

##### Mô hình MVT của Django:



Hình 2. 1 Mô hình MVT của Django

Model: là đại diện cho dữ liệu. Nó được dùng để tương tác với cơ sở dữ liệu và thực hiện các truy vấn, cập nhật và xử lý. Mô hình được định nghĩa bằng cách sử dụng



các lớp Python, và Django sử dụng ORM (Object-Relational Mapping) để tạo ra cấu trúc cơ sở dữ liệu từ các mô hình. ORM cho phép tương tác với cơ sở dữ liệu mà không cần viết truy vấn SQL trực tiếp. Điều này giúp giảm thời gian và công sức khi làm việc với dữ liệu.

**View:** đại diện cho logic xử lý và giao tiếp với người dùng. Trong Django, view là các hàm Python, và chúng xử lý các yêu cầu HTTP từ người dùng. View trả về các phản hồi dựa trên yêu cầu của người dùng, thường là các mô hình dữ liệu được trả về bởi mô hình.

**Template:** Template đại diện cho phần giao diện người dùng của ứng dụng. Nó chứa mã HTML để hiển thị dữ liệu cho người dùng. Django sử dụng ngôn ngữ template riêng của mình, có thể kết hợp với HTML để tạo ra các trang web động.

#### **Mô tả cách hoạt động:**

Đầu tiên, ứng dụng web sẽ nhận yêu cầu từ người dùng, yêu cầu được truyền xuống Django, sau đó sử dụng cơ chế định tuyến url để xác định views tương ứng với yêu cầu. View nhận yêu cầu và xử lý nó, có thể tương tác với Model để xử lý dữ liệu. Sau khi xử lý xong yêu cầu View sẽ lựa chọn Template phù hợp để hiển thị. View sẽ truyền dữ liệu đã xử lý từ mô hình qua cho Template, sau khi xử lý trong dữ liệu để hiển thị thì template sẽ gửi về cho View và view sẽ trả phản hồi về trình duyệt người dùng hiển thị ra trang web cuối cùng.

### **2.1.2 Data Mining**

Data Mining (khai phá dữ liệu) là quá trình tìm kiếm, khai thác thông tin, các quy luật của bộ dữ liệu, là một lĩnh vực của khoa học máy tính. Mục tiêu của nó là phân tích, trích xuất thông tin từ bộ dữ liệu và biến đổi đơn giản để sử dụng sau này.

Data mining giúp chúng ta hiểu được bộ dữ liệu hơn và tìm được các thông tin có hữu ích, thường được sử dụng để giải quyết các vấn đề phức tạp như phân loại, phân tích xu hướng,..

Một số thuật toán để khai thác dữ liệu: Apriori, FP-Growth, K-means clustering, decision trees, ... Mỗi thuật toán sẽ có mục tiêu và ứng dụng riêng.

#### **Ứng Dụng:**

- ❖ Sử dụng để phân tích hành vi khách hàng, gợi ý mua sắm. Giúp các chiến dịch quảng cáo và xác định khách hàng tiềm năng.

- ❖ Giúp dự đoán xu hướng thị trường, quản lý rủi ro tài chính và dự đoán doanh thu.
- ❖ Nó có thể hỗ trợ tìm ra quy luật ẩn trong lĩnh vực mua hàng.
- ❖ Hỗ trợ quyết định, phân loại, ...

#### **Một số công cụ sử dụng trong data mining:**

R: ngôn ngữ lập trình hỗ trợ tính toán thống kê mạnh mẽ. Cung cấp nhiều gói thư viện giúp thực hiện data mining, phân tích dữ liệu.

Python: Là ngôn ngữ lập trình phổ biến và có nhiều thư viện hỗ trợ mạnh mẽ cho data mining trong phân tích và mô hình dự đoán.

Weka: là phần mềm chuyên dụng cho khai thác dữ liệu và machine learning. Nó cung cấp nhiều công cụ và thuật toán cho phân tích dữ liệu, có giao diện dễ sử dụng.

RapidMiner: là công cụ khai phá dữ liệu nổi bật hiện nay được viết bằng java, cung cấp nhiều thư viện cho data mining, machine learning và quản lý dữ liệu.

Knime (Konstanz Information Miner): là một công cụ giúp khai phá, phân tích dữ liệu. Cung cấp nhiều thuật toán và tính năng cho xây dựng mô hình dự đoán và phát hiện quy luật.

#### **Quy trình thực hiện:**

Thu thập dữ liệu: Thu thập dữ liệu theo mục tiêu của dự án.

Tiền xử lý dữ liệu: Áp dụng các phương pháp và kỹ thuật để xử lý sao cho dữ liệu được thống nhất, phù hợp để khai phá.

Khám phá dữ liệu: Sử dụng các phương pháp trực quan để hiểu cấu trúc và tính chất của dữ liệu.

Xây dựng và đánh giá mô hình: Áp dụng mô hình để khai phá thì bước này thực hiện đánh giá và lựa chọn ra mô hình tối ưu nhất.

Triển khai và khai thác tri thức: triển khai vào thực tế để hỗ trợ quyết định hay dự đoán

### **2.1.3 Thuật toán FP-Growth**

FP-Growth (Frequent Pattern growth) là thuật toán trong khai thác dữ liệu (data mining). Nó được sử dụng để tìm các mẫu kết hợp trong bộ dữ liệu.

Thuật toán FP-growth (Frequent Pattern growth) là một thuật toán khai phá dữ liệu, được sử dụng trong mảng khai phá dữ liệu và phân tích dữ liệu. Thuật toán này được thiết kế để tìm ra các mẫu thường xuyên trong tập dữ liệu một cách hiệu quả,

đặc biệt là trong việc khám phá các mẫu thường xuyên trong dữ liệu ghi lại bằng các tập hợp phổ biến (frequent itemsets) và cây FP-tree.

Dưới đây là nói sơ lược về cách FP-growth hoạt động:

❖ *Xây dựng bảng hỗ trợ (Support Table):*

Đầu tiên, tính toán hỗ trợ của từng mục (item) trong tập dữ liệu. Hỗ trợ của một mục là số lần mục đó xuất hiện trong tập dữ liệu.

Loại bỏ những mục có hỗ trợ thấp hơn một ngưỡng được đặt trước (min\_support).

❖ *Sắp xếp các mục theo thứ tự hỗ trợ giảm dần:*

Sắp xếp các mục còn lại theo thứ tự giảm dần của hỗ trợ.

❖ *Xây dựng cây FP-tree:*

Quét lại tập dữ liệu và xây dựng cây FP-tree. Cây này sẽ có mỗi nút đại diện cho một mục và mỗi cạnh đại diện cho một sự kết hợp giữa các mục trong một giao dịch.

Dựa trên FP-tree đã tạo, xây dựng các conditional pattern base (cơ sở mẫu có điều kiện) cho mỗi mục.

Tạo các conditional FP-tree cho mỗi conditional pattern base.

❖ *Ghép kết quả từ các cây con thành kết quả chung:*

Kết hợp kết quả từ các cây con để tạo ra các mẫu thường xuyên toàn cục.

Thuật toán FP-growth có hiệu suất tốt so với một số thuật toán khác trong việc khai phá mẫu thường xuyên, đặc biệt là khi xử lý các tập dữ liệu lớn.

**Ví dụ mô tả về cách hoạt động của thuật toán:**

Chúng ta có một tập dữ liệu bán lẻ và đưa vào cây fp-tree:

Giao dịch	items
GD1	{ 'E', 'C', 'B', 'D', 'A' }
GD2	{ 'D', 'F', 'E' }
GD3	{ 'C', 'D', 'A', 'E' }
GD4	{ 'B', 'E' }
GD5	{ 'B', 'D', 'C', 'E' }

Bảng 2.1 Ví dụ về cây FP-TREE

Tính toán số lần xuất hiện các items là:

item	support
------	---------

E	5
D	4
C	3
B	3
A	2
F	1

Bảng 1.2 Ví dụ về bảng Support item

Chọn min\_support là 0,5 tương đương 50%, xuất hiện với tỉ lệ 2.5 trên 5 giao dịch, vậy ta chọn tần suất là 3

Ta có:  $I = \{E:5, D:4, C:3, B:3\}$

Loại bỏ những item không cần thiết chúng ta sẽ có các giao dịch sau:

Giao dịch	items
GD1	{'E', 'C', 'B', 'D'}
GD2	{'D', 'E'}
GD3	{'C', 'D', 'E'}
GD4	{'B', 'E'}
GD5	{'B', 'D', 'C', 'E'}

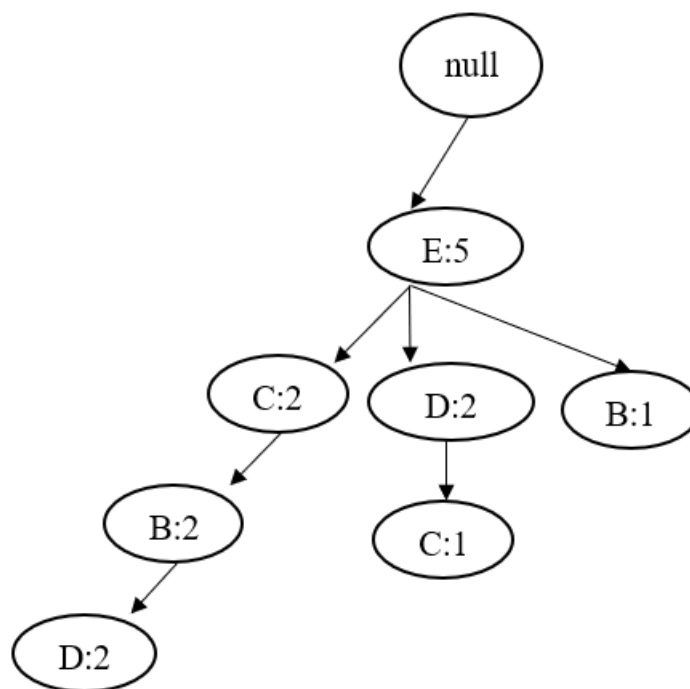
Bảng 2.2 Ví dụ về bảng giao dịch item

Bây giờ chúng ta có header table:

item	support
E	5
D	4
C	3
B	3

Bảng 2.3 Ví dụ bảng Header

Xây dựng cây FP-Tree, mỗi nút đại diện cho một mục và mỗi cạnh đại diện cho một sự kết hợp giữa các mục trong một giao dịch.



Hình 2. 2 Ví dụ cây FP-Tree

Ta có conditional pattern base:

item	Conditional pattern base
D	{E,C,B: 2}, {E:2}
B	{E,C: 2}, {E:1}
C	{E:2}, {E,D:1}
E	

Bảng 2.4 Conditional pattern base

Khai phá cây FP-tree, ta có min\_support là 0,5:

item	FP-tree	Mẫu phổ biến
D	E: 4	D, ED
B	E: 3	B, EB
C	E: 3	C, EB
E	5	E

Bảng 2.5 Bảng kết quả

Sau khi khai phá FP-Tree ta thấy được đã rút ra được các tập phổ biến. FP-growth sẽ giúp chúng ta khai phá được mối quan hệ tương quan, sự liên kết của các item từ những tập luật. Từ đó có thể giúp cho doanh nghiệp nắm bắt được xu hướng mua hàng và tối ưu hóa được hiệu suất kinh doanh.

## 2.2 Nghiên cứu liên quan

**Tên bài nghiên cứu:** Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons (Nghiên cứu so sánh về thuật toán Apriori và thuật toán tăng trưởng Fp với những ưu và nhược điểm).

**Tác giả:** M. Kavitha, S.T. Tamil Selvi, Khoa Khoa học Máy tính Cao đẳng dành cho nữ Tiruppur Kumaran Tiruppur, Tamil Nadu, Ấn Độ.

**Thời gian:** International Journal of Computer Science Trends and Technology (IJCS T) – Volume 4 Issue 4, Jul - Aug 2016 (Tạp chí Quốc tế về Xu hướng và Công nghệ Khoa học Máy tính (IJCS T) – Tập 4 Số 4, tháng 7 - tháng 8 năm 2016)

**Đường dẫn bài nghiên cứu:** [2] dưới danh mục tài liệu tham khảo

**Mục tiêu nghiên cứu:** Mục tiêu của bài báo này là thực hiện một nghiên cứu so sánh giữa hai thuật toán khai thác mẫu phổ biến là Apriori và FP-Growth trong lĩnh vực Khai thác Dữ liệu. Nghiên cứu này có thể nhằm đánh giá và so sánh hiệu suất, ưu điểm và nhược điểm của cả hai thuật toán khi áp dụng vào việc khai thác mẫu phổ biến từ cơ sở dữ liệu lớn. Bài báo tập trung vào việc so sánh cách mà Apriori và FP-Growth xử lý quá trình tạo và kiểm thử tập ứng viên, cũng như cách chúng khai thác và phân tích mẫu phổ biến từ dữ liệu. Mục đích cuối cùng có thể là cung cấp một cái nhìn chi tiết về hiệu suất và tính ứng dụng của cả hai thuật toán trong ngữ cảnh của khai thác dữ liệu.

**Kết luận:** Apriori, mặc dù phù hợp với cơ sở dữ liệu lớn, nhưng đối mặt với các thách thức về thời gian và bộ nhớ do quá trình tạo nhiều tập ứng viên. Trong khi đó, FP-Growth được mô tả là hiệu quả hơn đối với cơ sở dữ liệu lớn vì nó tiết kiệm bộ nhớ và thời gian thực thi.

## Chương 3: MÔ HÌNH LÝ THUYẾT

### 3.1 Mô tả bài toán

Bài toán xây dựng thuật toán FP-growth để đưa ra gợi ý sản phẩm là một vấn đề trong lĩnh vực khai phá dữ liệu và phân tích hành vi người dùng. Đối với các doanh nghiệp thương mại điện tử, cửa hàng trực tuyến, hoặc bất kỳ hệ thống nào thu thập dữ liệu giao dịch từ khách hàng, việc hiểu rõ xu hướng mua sắm và mối quan hệ giữa các sản phẩm có thể mang lại nhiều lợi ích. Trong ngữ cảnh thương mại điện tử và quản lý doanh nghiệp, bài toán xây dựng thuật toán FP-growth để đưa ra gợi ý sản phẩm đang trở thành một yếu tố quan trọng để tối ưu hóa trải nghiệm mua sắm và nâng cao hiệu suất kinh doanh. Bài toán này xoay quanh việc khai phá dữ liệu từ các giao dịch mua sắm của khách hàng để hiểu rõ hơn về xu hướng mua hàng và mối quan hệ giữa các sản phẩm.

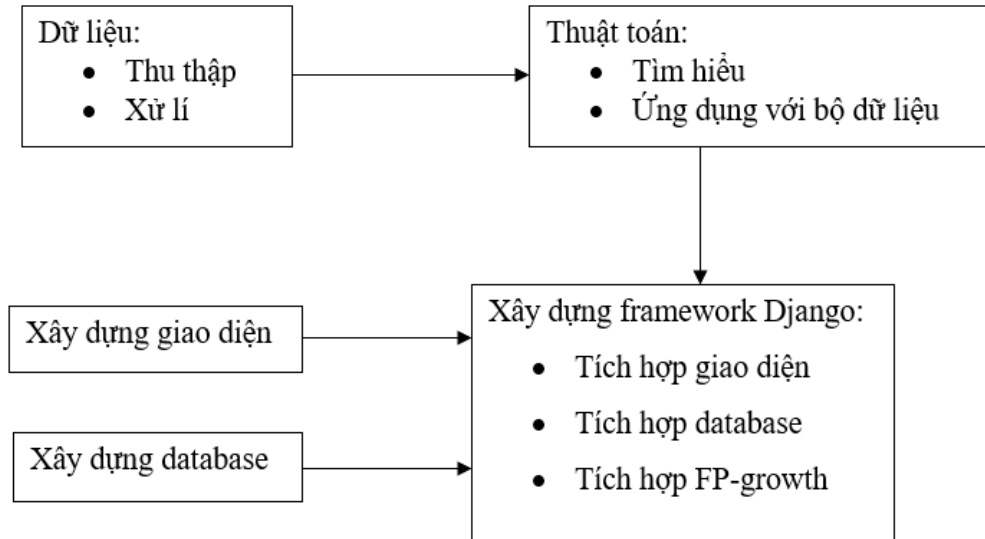
Dữ liệu đầu vào bao gồm thông tin về các giao dịch, trong đó mỗi giao dịch được biểu diễn dưới dạng danh sách các sản phẩm khách hàng đã mua. Mục tiêu chính là xây dựng một mô hình phân tích, có khả năng tìm ra các mẫu mua hàng thường xuyên (frequent itemsets). Để giải quyết bài toán này, chúng ta áp dụng thuật toán FP-growth, một phương pháp hiệu quả trong việc xây dựng cây FP-growth từ dữ liệu.

Thuật toán bắt đầu bằng việc loại bỏ những sản phẩm ít xuất hiện để tạo ra một cây FP-growth có cấu trúc đơn giản và dễ quản lý. Từ cây này, chúng ta xây dựng Conditional Pattern Base cho mỗi sản phẩm, tạo ra các cây con để tiếp tục phân tích sâu hơn. Quá trình lặp này được thực hiện cho đến khi không còn mẫu nào hoặc chỉ còn một mẫu.

Kết quả là chúng ta thu được các frequent itemsets, biểu thị nhóm sản phẩm thường xuất hiện cùng nhau trong các giao dịch. Từ những thông tin này, hệ thống có thể đưa ra gợi ý sản phẩm cho khách hàng dựa trên lịch sử mua hàng của họ. Nếu khách hàng đã mua một số sản phẩm trong một frequent itemset, hệ thống có thể đề xuất các sản phẩm còn lại trong itemset đó, tăng cơ hội mua sắm và tạo ra trải nghiệm cá nhân hóa.

Bằng cách này, doanh nghiệp có thể nắm bắt thông tin quý báu từ dữ liệu khách hàng để tạo ra chiến lược kinh doanh linh hoạt và hiệu quả.

### 3.2 Mô tả hệ thống mức cao



Hình 3. 1 Sơ đồ hệ thống mức cao

#### ***Mô tả hệ thống mức cao***

❖ Sơ đồ trên là mô tả về quy trình thực hiện của dự án này, nó sẽ trải qua các bước là:

❖ Đầu tiên là Dữ liệu: Thu thập dữ liệu, tìm dữ liệu trên các nguồn lớn đáng tin cậy sau đó lấy về và thực hiện xử lý, chuẩn hóa sao cho phù hợp với mô hình học máy để huấn luyện và đưa ra kết quả.

❖ Thuật toán: Tìm hiểu về thuật toán và quy trình hoạt động của thuật toán FP-growth, hiểu được quy luật của nó và ứng dụng nó với bộ dữ liệu để hỗ trợ đưa ra quyết định gợi ý.

❖ Về phần sản phẩm thì đầu tiên chúng ta phải xây dựng giao diện sao cho phù hợp với yêu cầu hệ thống.

❖ Tiếp theo đó chúng ta phải xây dựng database phù hợp với nhu cầu sử dụng của hệ thống.

Sau khi hoàn thành các bước trên thì chúng ta sẽ dựng khung Django và tích hợp giao diện, sau đó là database. Cuối cùng sẽ tích hợp thuật toán fp-growth vào hỗ trợ



đưa ra gợi ý các sản phẩm có liên quan với nhau giúp người tiêu dùng dễ dàng mua sắm và nâng cao được lợi nhuận cho cửa hàng.

### **3.2 Mô tả công nghệ**

Thuật toán FP-growth (Frequent Pattern-growth) là

Nguyên Tắc Hoạt Động:

Tạo cây FP-Tree từ tập dữ liệu, trong đó mỗi nút đại diện cho một mục mua sắm và các nhánh của nút biểu thị sự xuất hiện của mục mua sắm đó trong các giao dịch.

Xây dựng bảng header để theo dõi các nút của cây, giúp truy cập các nút liên quan.

Duyệt qua cây FP-Tree để xác định các mô hình tần suất, nơi mà mỗi đường đi từ nút lá lên đến gốc cây đại diện cho một mẫu tần suất.

#### **Lý Do Lựa Chọn FP-growth:**

FP-growth hiệu quả khi xử lý tập dữ liệu lớn, do không cần phải tạo tất cả các tập con của dữ liệu như các thuật toán khác như Apriori.

Tận dụng cấu trúc cây FP-Tree để lưu trữ thông tin mô hình một cách hiệu quả, đặc biệt là với các tập dữ liệu có độ thưa thớt.

FP-growth có khả năng mở rộng tốt, có thể xử lý các tập dữ liệu động và được cập nhật liên tục mà không cần phải chạy lại toàn bộ thuật toán.

Tạo ra kết quả tối ưu, giúp xây dựng mô hình mua sắm và gợi ý sản phẩm dựa trên mẫu mua sắm có ý nghĩa

#### **Tích Hợp Với Dữ Liệu Mua Sắm:**

- ❖ FP-growth được lựa chọn vì khả năng hiệu quả của nó trong việc phân tích và khai thác mẫu mua sắm từ dữ liệu giao dịch của khách hàng. Đặc biệt, thuật toán này phản ánh tốt được sự liên kết giữa các sản phẩm, giúp tạo ra gợi ý mua sắm có ý nghĩa và hiệu quả cho khách hàng.
- ❖ Lựa chọn FP-growth trong dữ liệu mua sắm không chỉ giúp tối ưu hóa hiệu suất mà còn tạo ra mô hình phù hợp với đặc tính và độ thưa thớt của dữ liệu, từ đó đảm bảo rằng hệ thống gợi ý sản phẩm hoạt động mạnh mẽ và hiệu quả.

### 3.3 Mô hình toán

Mô hình gợi ý sản phẩm của bài được thực hiện dựa trên thuật toán FP-growth. Nó sẽ tạo cây FP-tree từ các giao dịch, trong đó mỗi nút sẽ đại diện cho sản phẩm mua sắm và mỗi nhánh sẽ là một giao dịch.

Sau đó FP-growth sẽ khai thác tần suất xuất hiện từ các mẫu của cây FP-tree, các mẫu này sẽ biểu diễn các tập sản phẩm được mua cùng nhau.

Từ các mẫu tần suất, chúng ta sẽ xây dựng mô hình gợi ý sản phẩm bằng các quy luật kết hợp từ giữa các mục mua sắm.

Các mẫu có tần số support cao hơn được xem là quan trọng và được sử dụng trong việc đưa ra gợi ý.

## Chương 4: THỰC NGHIỆM

### 4.1 Thiết kế giao diện

Ở phần này, chúng ta sẽ đi vào mô tả phần giao diện của hệ thống. Mục tiêu của giao diện là hỗ trợ người dùng với các chức năng thuận tiện hơn. Giao diện sẽ có các phần như: trang sản phẩm, giỏ hàng, thanh toán.

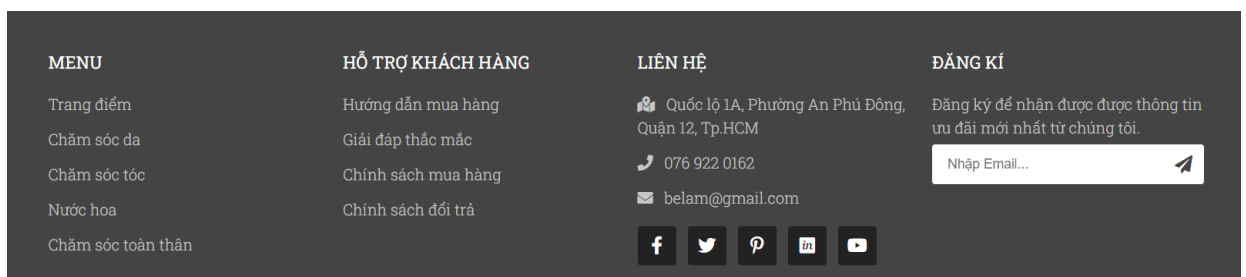
Giao diện của hệ thống này sẽ có những phần chung như phần header và footer.

Phần header bao gồm logo, thanh tìm kiếm, và thanh navigation bar có các chức năng dẫn đến trang khác. Khi click vào phần trang chủ thì sẽ đi đến trang trang chủ của hệ thống, click vào trang sản phẩm thì sẽ đi đến trang các sản phẩm, chức năng tin tức/liên hệ là đưa đến trang các tin tức/liên hệ. Các chức năng còn lại cũng vậy, click vào trang nào thì sẽ đi đến trang đó. Dưới đây là hình ảnh phần header:



Hình 4. 1 Phần Header

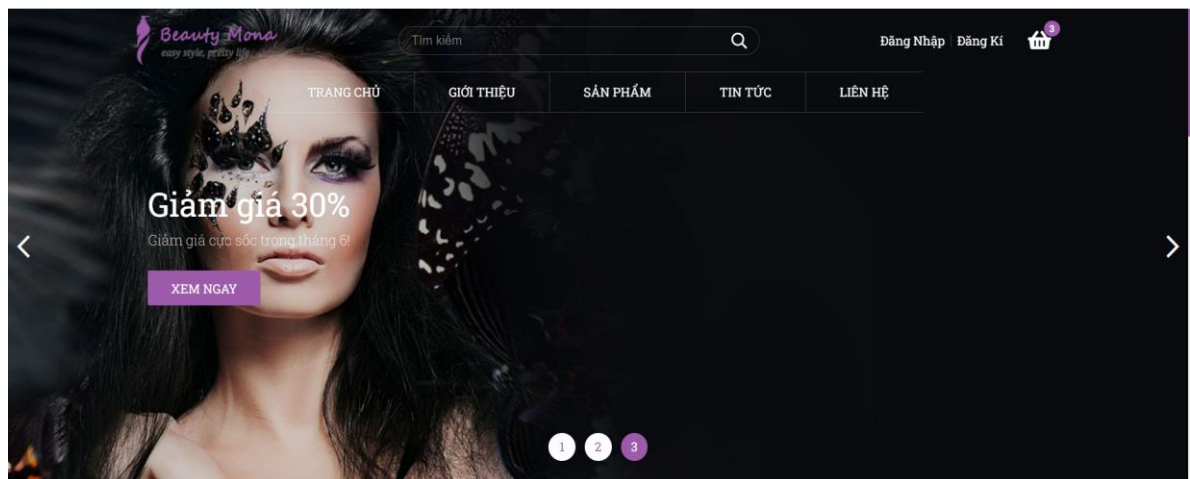
Còn về phần footer, hay còn gọi là phần chân trang, đây là phần dưới cùng của trang web. Nó có chức năng hiển thị các thông tin cần thiết về hệ thống của hàng. Như địa chỉ, số điện thoại để liên lạc, và một số thông tin cần thiết khác.



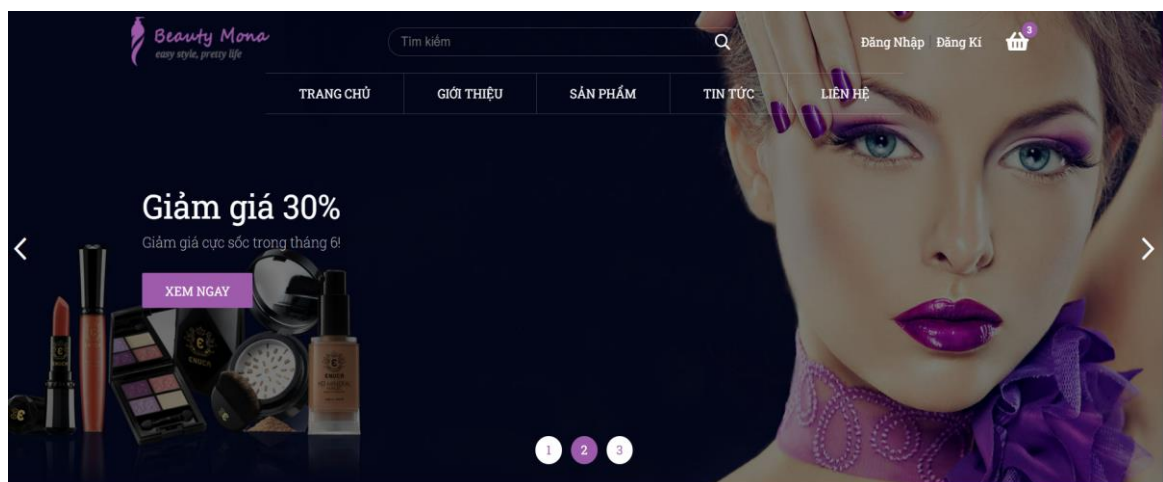
Hình 4. 2 Phần Footer

Bây giờ, chúng ta sẽ đi vào trang Trang chủ của hệ thống, trang này có chức năng hiển thị các thông tin chung, tin tức của hệ thống.

Dưới đây là phần slider của trang chủ, hiển thị các thông tin nổi bật của cửa hàng

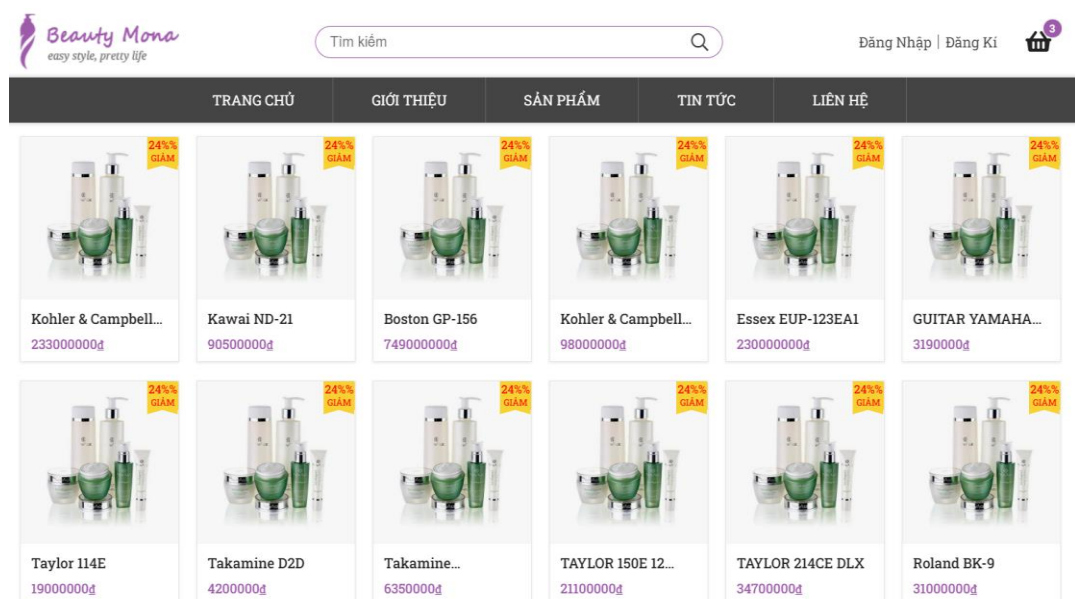


Hình 4. 3 Phần slider 1



Hình 4. 4 Phần slider 2

Hiển thị các sản phẩm của hệ thống và một số thông tin khác

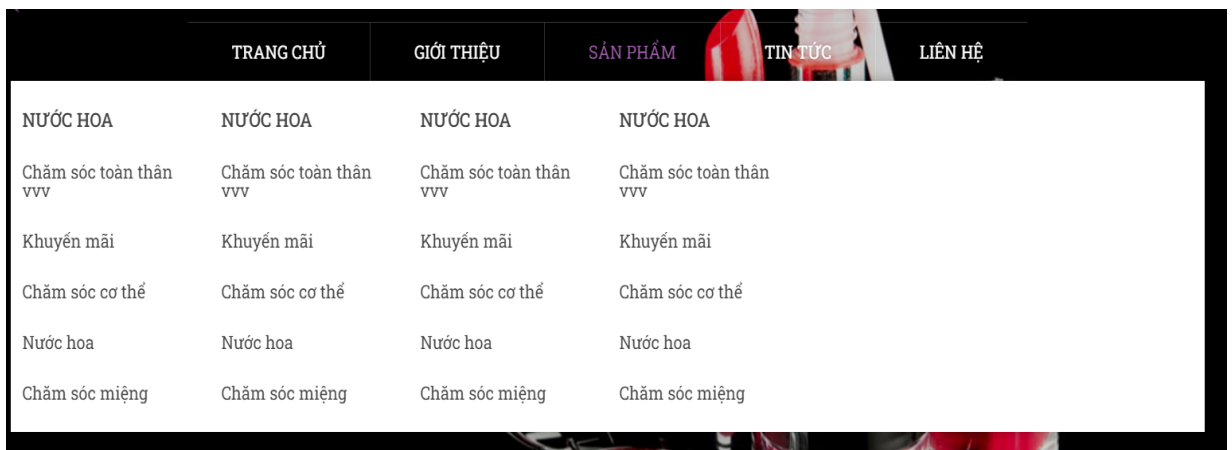


Hình 4. 5 Giao diện trang chủ sản phẩm

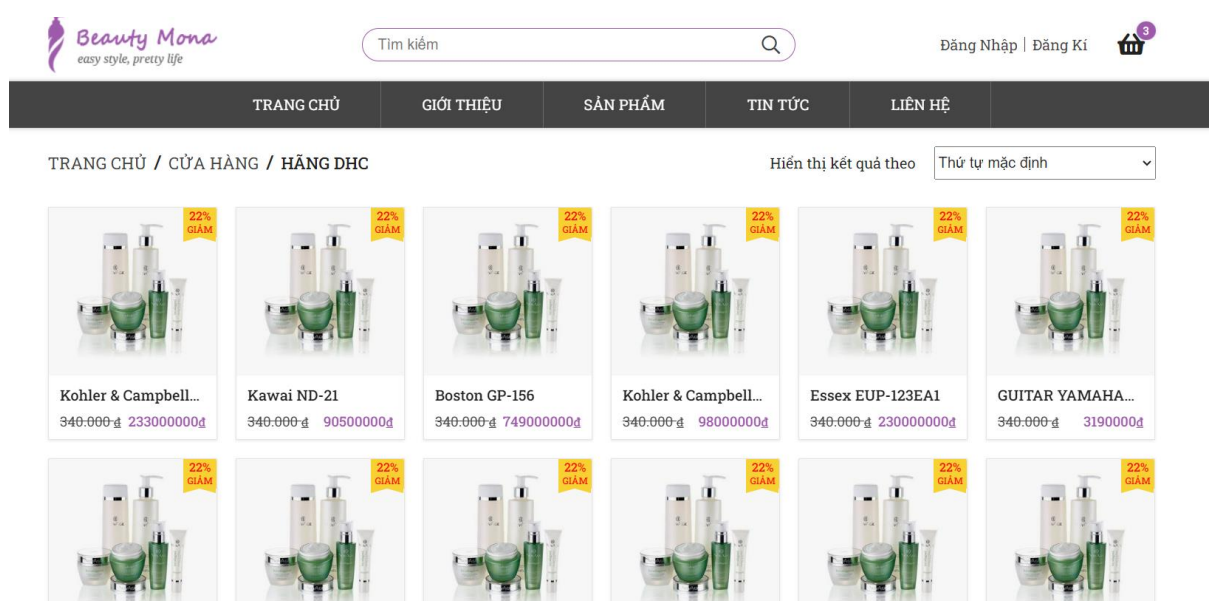


Hình 4. 6 Giao diện trang chủ tình tức

Chúng ta đến với trang sản phẩm, khi đưa chuột vào chức năng sản phẩm trên thanh navbar thì chúng sẽ hiển thị một số danh mục phân loại sản phẩm.



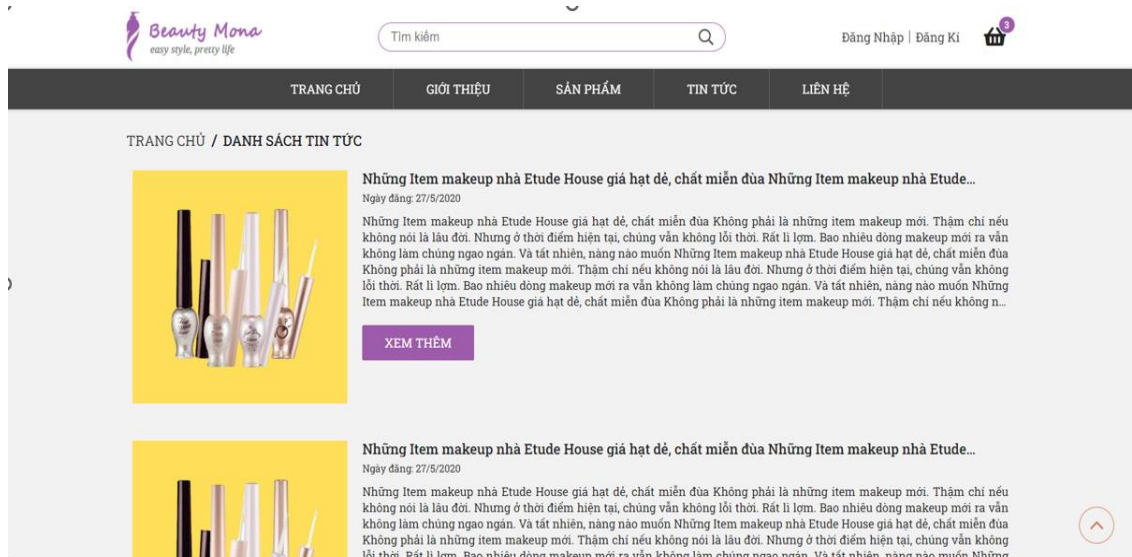
Hình 4. 7 Phần mục lục



Hình 4. 8 Giao diện sản phẩm

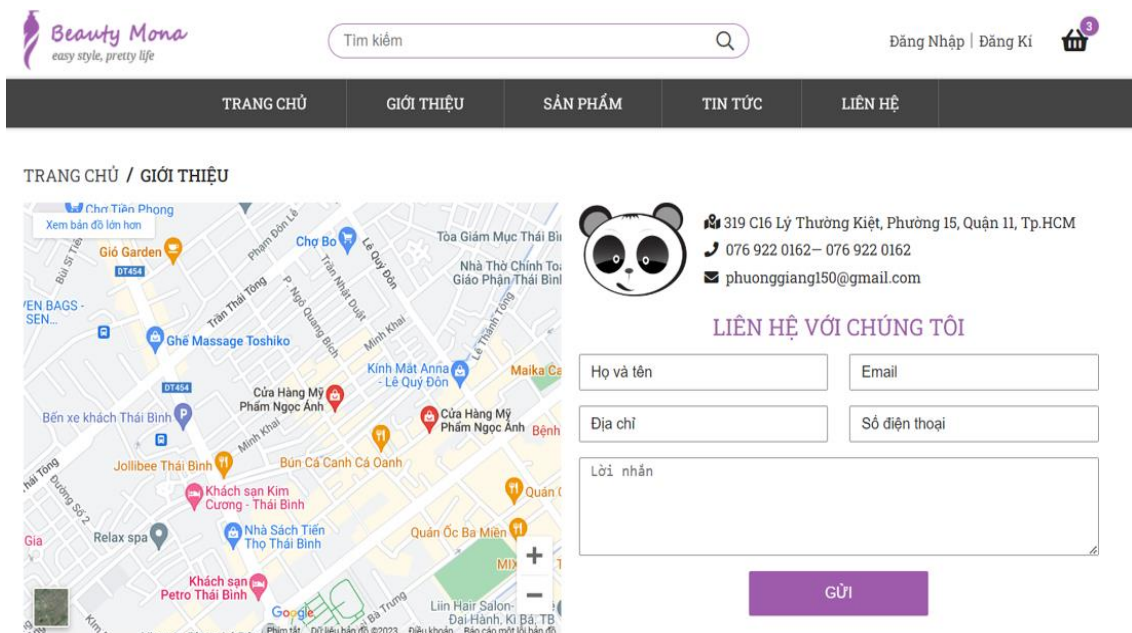
Đây là hình ảnh trang hiển thị sản phẩm của hệ thống cửa hàng.

Dưới đây là hình ảnh trang tin tức của hệ thống, có chức năng hiển thị các bài viết tin tức được đăng lên.



Hình 4. 8 Giao diện danh sách tin tức

Tiếp theo đây là trang liên hệ của hệ thống, giúp người dùng để lại thông tin cần hỗ trợ. Từ đó nhân viên sẽ nhận là liên hệ giúp họ.

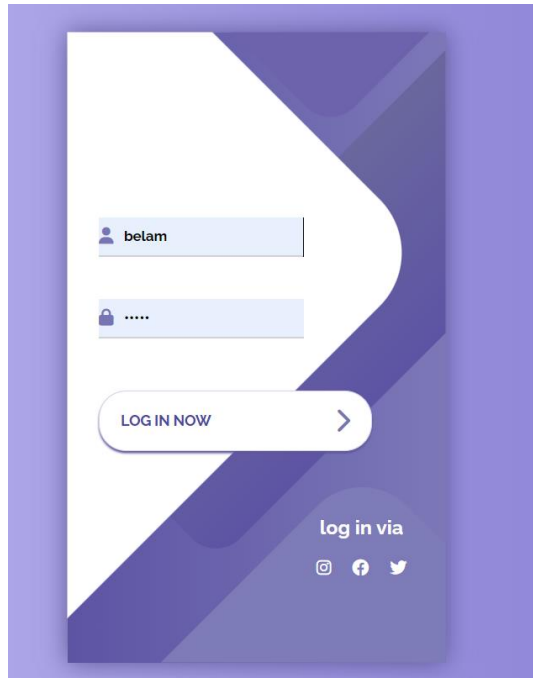


Hình 4. 9 Giao diện liên hệ

Chúng ta còn một số trang như đăng nhập, đăng ký, giúp người dùng quản trị hay khách hàng có thể đăng nhập vào.

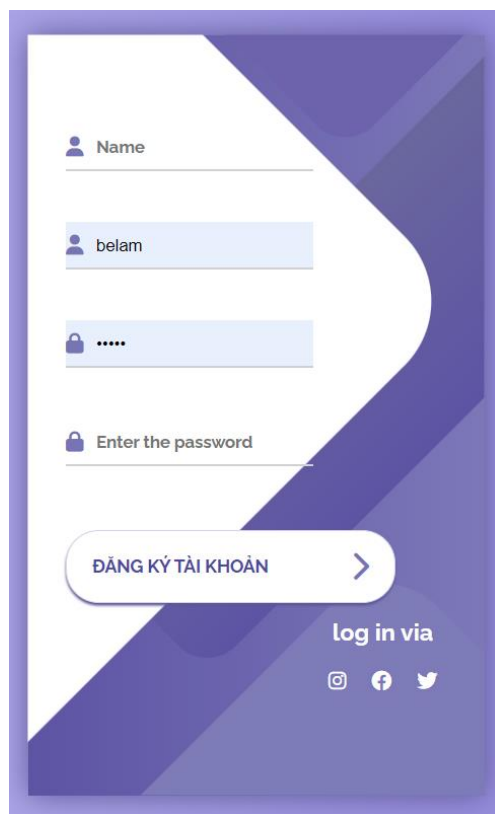
Dưới đây là trang đăng nhập.





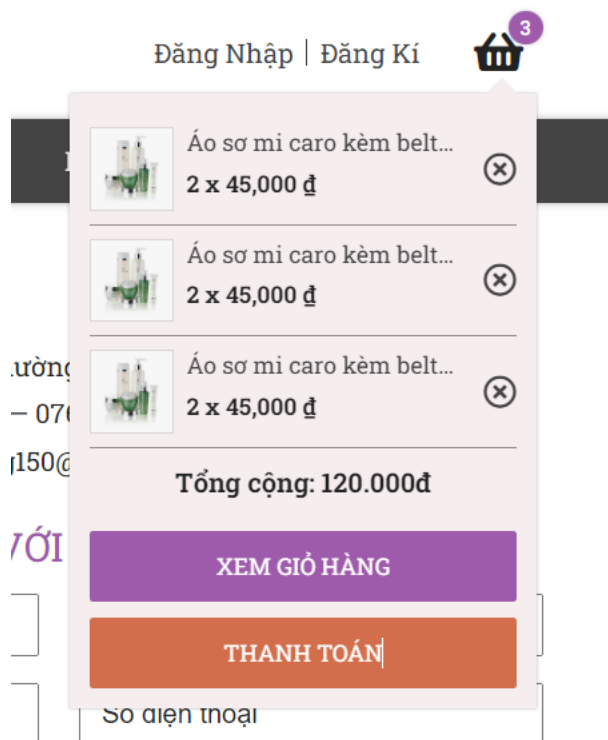
Hình 4. 10 Giao diện đăng nhập

Đây là trang đăng ký, giúp người dùng tạo user để đăng nhập vào cửa hàng.



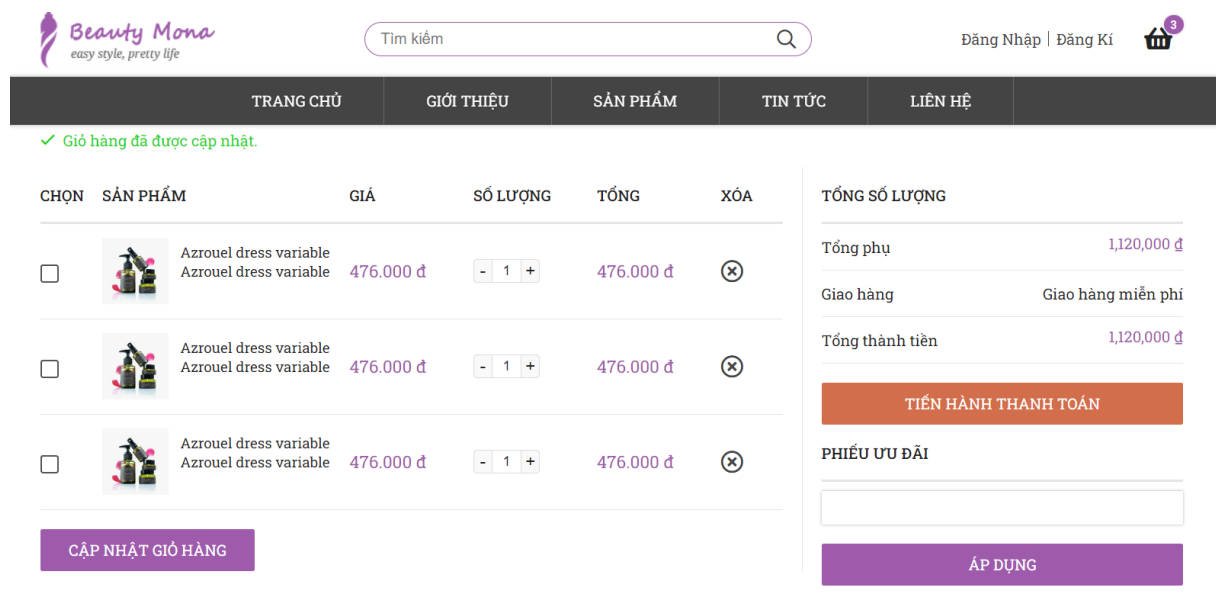
Hình 4. 11 Giao diện đăng kí

Sau khi nhấp chuột vào giỏ hàng trên phần header thì giỏ hàng sẽ hiện ra như thế này.



Hình 4. 12 Giỏ hàng mini

Click vào xem giỏ hàng sẽ hiển thị ra giỏ hàng lớn:



Hình 4. 13 Giao diện giỏ hàng

Phía sau giỏ hàng sẽ là trang thanh toán, trang này giúp người dùng nhập các thông tin cần thiết để liên hệ và giao hàng.



TRANG CHỦ	GIỚI THIỆU	SẢN PHẨM	TIN TỨC	LIÊN HỆ
-----------	------------	----------	---------	---------

### THÔNG TIN THANH TOÁN

Họ Tên \*

Địa chỉ \*

Tỉnh / Thành phố \*

Email \*

Số điện thoại \*

Ghi chú cho đơn hàng

### ĐƠN HÀNG CỦA BẠN

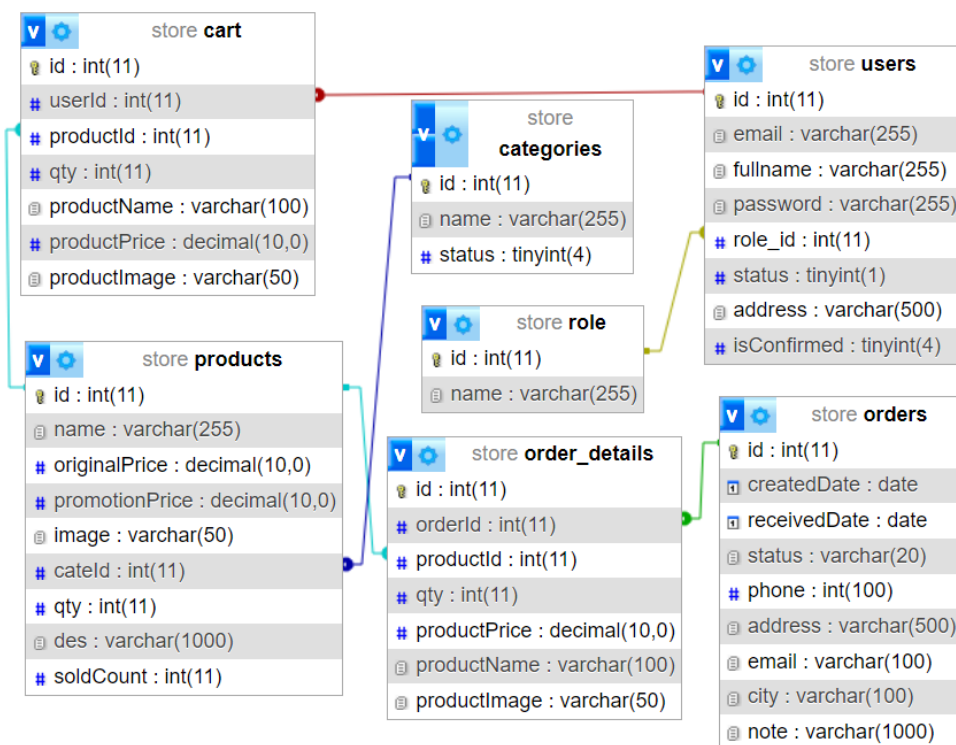
Azrouel dress variable		1,120,000 đ
Azrouel dress variable	x 3	1,120,000 đ
Azrouel dress variable		1,120,000 đ
Giao hàng	Giao hàng miễn phí	
Tổng thành tiền		1,120,000 đ

**ĐẶT HÀNG**

Hình 4. 14 Giao diện trang thanh toán

## 4.2 Mô tả database

Trong hệ thống quản lý cửa hàng thì chúng ta phải có một cơ sở dữ liệu để lưu trữ những dữ liệu của cửa hàng một cách có tổ chức, khi cần thì có thể truy xuất và thay đổi dễ dàng hơn. Nó giúp chúng ta quản lý thông tin và có thể ứng dụng cho các phân tích dữ liệu hỗ trợ chúng ta đưa ra quyết định trong kinh doanh hay trong các trường hợp cần thiết.



Hình 4. 15 Cơ sở dữ liệu quan hệ

Trên đây là bảng cơ sở dữ liệu của hệ thống. Nó được xây dựng dựa trên ngôn ngữ SQL. Bảng cơ sở dữ liệu trên thể hiện được các mối liên hệ trong cơ sở dữ liệu.

Đầu tiên, chúng ta sẽ nói về bảng USER, đây là bảng đại diện cho người dùng. Người dùng khi đăng nhập sẽ cần nhập các thông tin cần thiết hỗ trợ cho khi sử dụng hệ thống. Bảng user này có primary key là ID.

Bảng categories: đây là bảng phân loại sản phẩm, mỗi sản phẩm sẽ tương ứng với một loại sản phẩm của nó. Nó có primary key là khóa chính trong bảng.

Bảng Product (sản phẩm), nó lưu trữ các sản phẩm và các thông tin cần thiết của chúng, ví dụ như tên, mô tả, giá sản phẩm, hình ảnh, số lượng của chúng. Nó có khóa chính là id để xác nhận mỗi sản phẩm là riêng biệt để dễ dàng truy xuất và có foreign key với categories là trường cateID để tham chiếu đến danh mục sản phẩm của nó.

Bảng cart là giỏ hàng, mỗi giỏ hàng sẽ liên kết với một người dùng. Nó chứa các sản phẩm được khách hàng thêm vào và số lượng chúng được thêm. Trong bảng này có hai foreign key là userID và productID. UserID để tham chiếu đến người dùng, productID tham chiếu đến sản phẩm mà được thêm vào trong giỏ hàng.

Bảng order là bảng đơn hàng, mỗi id tương ứng với mỗi đơn hàng, chứa các thành phần như email, địa chỉ, số điện thoại, đây là các thông tin để giao hàng và đơn hàng được giao thành công thì nó sẽ cập nhật trạng thái là complete.

Bảng order\_details là mỗi hóa đơn sẽ có các sản phẩm và số lượng của nó trong đó, vậy nên bảng này thể hiện cho chúng ta thấy được trong hóa đơn có bao nhiêu sản phẩm, số lượng bao nhiêu. Nó có foreign key với bảng order là để tham chiếu đến hóa đơn chứa nó.

### **4.3 Mô tả dữ liệu**

Bộ dữ liệu này là các giao dịch trực tuyến của một cửa hàng bán lẻ, bộ dữ liệu được lấy từ Kaggle.

Nó thể hiện các giao dịch trong hóa đơn được mua bán trong hai năm là 2010 với 2011. Bộ dữ liệu này có 541909 dòng và 8 cột, dưới đây là thông tin về bộ dữ liệu.

```

RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   InvoiceNo              541909 non-null object
1   StockCode              541909 non-null object
2   Description            540455 non-null object
3   Quantity              541909 non-null int64
4   InvoiceDate            541909 non-null object
5   UnitPrice              541909 non-null float64
6   CustomerID            406829 non-null float64
7   Country                541909 non-null object
dtypes: float64(2), int64(1), object(5)
memory usage: 33.1+ MB

```

Hình 4. 16 Mô tả bộ dữ liệu

Bộ dữ liệu có 5 cột mang định dạng object, 1 cột int và 2 cột float.

Dưới đây là hình ảnh các cột của dữ liệu

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom

Hình 4. 17 Thông tin bộ dữ liệu

Cột InvoiceNo là đại diện cho các hóa đơn, vì mỗi dòng là mỗi sản phẩm được bán ra nên các sản phẩm có số hóa đơn trùng nhau là các sản phẩm được mua trong cùng một hóa đơn.

Cột StockCode là mã của các phẩm.

Cột Description là mô tả về các sản phẩm.

Cột Quantity là số lượng của các sản phẩm.

Cột InvoiceDate là thời gian các sản phẩm được bán ra.

Cột UnitPrice là giá của các sản phẩm

Cột CustomerID là mã của khách hàng.

Cột Country là quốc gia bán ra.

Từ bộ dữ liệu trên, chúng ta sẽ đi phân tích để hỗ trợ cho hệ thống.

## 4.4 Kết quả

### 4.4.1 Xử lý dữ liệu

Từ bộ dữ liệu hóa đơn giao dịch bán lẻ trên, em sẽ trích xuất các sản phẩm theo từng hóa đơn thành từng mảng. Từ đó sẽ đưa vào máy học rút ra tập luật cho các sản phẩm nào liên quan đến nhau, giúp hệ thống đưa ra gợi ý mua hàng dễ dàng hơn cho khách hàng và làm tăng doanh thu.

```
Invoice 0: ['WHITE HANGING HEART T-LIGHT HOLDER', 'WHITE METAL LANTERN', 'CREAM CUPID HEARTS COAT HANGER', 'KNITTED UNION FLAG HOT WATER BOTTLE', 'RED HANGING HEART T-LIGHT HOLDER']
Invoice 1: ['HAND WARMER UNION JACK', 'HAND WARMER RED POLKA DOT']
Invoice 2: ['ASSORTED COLOUR BIRD ORNAMENT', 'POPPY'S PLAYHOUSE BEDROOM', 'POPPY'S PLAYHOUSE KITCHEN', 'FELTCRAFT PRINCESS CHARLOTTE DOLL', 'IVORY DOLLY GIRL']
Invoice 3: ['JAM MAKING SET WITH JARS', 'RED COAT RACK PARIS FASHION', 'YELLOW COAT RACK PARIS FASHION', 'BLUE COAT RACK PARIS FASHION']
Invoice 4: ['BATH BUILDING BLOCK WORD']
Invoice 5: ['ALARM CLOCK BAKELIKE PINK', 'ALARM CLOCK BAKELIKE RED', 'ALARM CLOCK BAKELIKE GREEN', 'PANDA AND BUNNIES STICKER SHEET', 'STARS GIFT TAG']
Invoice 6: ['PAPER CHAIN KIT 50'S CHRISTMAS']
Invoice 7: ['HAND WARMER RED POLKA DOT', 'HAND WARMER UNION JACK']
Invoice 8: ['WHITE HANGING HEART T-LIGHT HOLDER', 'WHITE METAL LANTERN', 'CREAM CUPID HEARTS COAT HANGER', 'EDWARDIAN PARASOL RED', 'RETRO COFFEE MUG']
Invoice 9: ['VICTORIAN SEWING BOX LARGE']
Invoice 10: ['WHITE HANGING HEART T-LIGHT HOLDER', 'WHITE METAL LANTERN', 'CREAM CUPID HEARTS COAT HANGER', 'EDWARDIAN PARASOL RED', 'RETRO COFFEE MUG']
Invoice 11: ['HOT WATER BOTTLE TEA AND SYMPATHY', 'RED HANGING HEART T-LIGHT HOLDER']
Invoice 12: ['HAND WARMER RED POLKA DOT', 'HAND WARMER UNION JACK']
Invoice 13: ['JUMBO BAG PINK POLKADOT', 'JUMBO BAG BAROQUE BLACK WHITE', 'JUMBO BAG CHARLIE AND LOLA TOYS', 'STRAWBERRY CHARLOTTE BAG', 'RED 3 PIECE']
Invoice 14: ['JAM MAKING SET PRINTED']
Invoice 15: ['RETROSPOT TEA SET CERAMIC 11 PC', 'GIRLY PINK TOOL SET', 'JUMBO SHOPPER VINTAGE RED PAISLEY', 'AIRLINE LOUNGE METAL SIGN', 'WHITE SPOT']
Invoice 16: ['INFLATABLE POLITICAL GLOBE', 'VINTAGE SNAKES & LADDERS', 'CHOCOLATE CALCULATOR', 'JUMBO SHOPPER VINTAGE RED PAISLEY', 'RECYCLING BAG']
Invoice 17: ['WOOD BLACK BOARD ANT WHITE FINISH', 'COLOUR GLASS T-LIGHT HOLDER HANGING', 'HANGING METAL HEART LANTERN', 'HANGING MEDINA LANTERN SMALL']
Invoice 18: ['SET 3 WICKER OVAL BASKETS W LIDS', 'JAM MAKING SET PRINTED', 'JAM MAKING SET WITH JARS', 'JUMBO BAG DOLLY GIRL DESIGN', 'TRADITIONAL C']
Invoice 19: ['WHITE WIRE EGG HOLDER', 'JUMBO BAG BAROQUE BLACK WHITE', 'JUMBO BAG RED RETROSPOT']
Invoice 20: ['CHILLI LIGHTS', 'LIGHT GARLAND BUTTERFILES PINK', 'WOODEN OWLS LIGHT GARLAND', 'FAIRY TALE COTTAGE NIGHTLIGHT', 'RED TOADSTOOL LED NIGHTLIGHT']
Invoice 21: ['HOME BUILDING BLOCK WORD', 'LOVE BUILDING BLOCK WORD', 'DOORMAT FANCY FONT HOME SWEET HOME', 'HOME SMALL WOOD LETTERS', 'GINGHAM HEART']
Invoice 22: ['CHRISTMAS LIGHTS 10 REINDEER', 'VINTAGE UNION JACK CUSHION COVER', 'VINTAGE HEADS AND TAILS CARD GAME', 'SET OF 3 COLOURED FLYING DUCKS']
Invoice 23: ['CHRISTMAS LIGHTS 10 REINDEER', 'JAM MAKING SET WITH JARS', 'JAM MAKING SET PRINTED', 'JAM JAR WITH PINK LID', 'JAM JAR WITH GREEN LID']
Invoice 24: ['3 STRIPEY MICE FELTCRAFT', 'SET OF 6 SOLDIER SKITTLES', 'TRADITIONAL WOODEN SKIPPING ROPE', 'WOODEN BOX OF DOMINOES', 'RUSTIC SEVENTEENTH CENTURY']
...
Invoice 25896: ['VICTORIAN GLASS HANGING T-LIGHT', 'ZINC T-LIGHT HOLDER STARS SMALL']
Invoice 25897: ['Manual']
Invoice 25898: ['VICTORIAN SEWING BOX LARGE']
Invoice 25899: ['HANGING HEART JAR T-LIGHT HOLDER', '36 PENCILS TUBE RED RETROSPOT']
```

Hình 4. 18 Chuyển hóa dữ liệu

Hình trên là kết quả sau khi rút trích các sản phẩm trong cùng hóa đơn thành mảng. Ta thấy được mỗi mảng sẽ có các sản phẩm trùng mã hóa đơn sẽ được tích lại thành một mảng.

Sau đó sử dụng TransactionEncoder từ thư viện mlxtend để chuyển đổi giao dịch thành một dataframe được mã hóa one-hot. one-hot là một kỹ thuật biểu diễn dữ liệu categorical thành dạng số học. Dữ liệu categorical là dữ liệu mà có thể nhận một số giá trị giới hạn, không phải dữ liệu liên tục. Mã hóa one-hot chuyển đổi mỗi giá trị trong biến categorical thành một cột trong ma trận, và mỗi dòng của ma trận đại diện cho một giao dịch. Mỗi cột mới thường chỉ nhận giá trị 0 hoặc 1, tương ứng với việc có hay không có. Điều này tạo ra một ma trận mà ở mỗi cột chỉ có một giá trị là 1 (được gọi là "hot") và tất cả các giá trị khác là 0.

Về transactionEncoder ta có hai phương thức fit và transform. Fit sử dụng để xác định tất cả các mục duy nhất xuất hiện trong các giao dịch. Các mục duy nhất này

sẽ được sắp xếp và xác định thành các cột trong ma trận mã hóa one-hot. Sau khi quá trình fit hoàn tất, sử dụng TransactionEncoder để transform các tập dữ liệu mới có cấu trúc tương tự. Transform sẽ chuyển đổi dữ liệu giao dịch thành một ma trận mã hóa one-hot, trong đó mỗi hàng tương ứng với một giao dịch và mỗi cột tương ứng với một mục duy nhất.

#### 4.4.2 Thuật toán

Kết quả là danh sách các tập hợp mặt hàng (itemsets) cùng với giá trị hỗ trợ (support) tương ứng của chúng. Trong thuật toán FP-growth, các itemsets này biểu thị các nhóm sản phẩm thường xuất hiện cùng nhau trong dữ liệu mua sắm.

	support	itemsets
0	0.088880	(WHITE HANGING HEART T-LIGHT HOLDER)
1	0.017876	(KNITTED UNION FLAG HOT WATER BOTTLE)
2	0.016795	(RED WOOLLY HOTTIE WHITE HEART.)
3	0.014865	(SET 7 BABUSHKA NESTING BOXES)
4	0.012201	(WHITE METAL LANTERN)
...	...	...
1037	0.013475	(JUMBO BAG RED RETROSPOT, JUMBO BAG VINTAGE DO...
1038	0.010656	(JUMBO BAG VINTAGE DOILY, JUMBO BAG VINTAGE L...
1039	0.010039	(LUNCH BAG VINTAGE DOILY, JUMBO BAG VINTAGE D...
1040	0.010347	(LUNCH BAG VINTAGE DOILY, LUNCH BAG RED RETRO...
1041	0.010463	(HOT WATER BOTTLE KEEP CALM, CHOCOLATE HOT WAT...

Hình 4. 19 Kết quả của thuật toán FP - GROWTH

Đây là một số item có liên quan đến nhau sau khi đã chạy thuật toán FP-growth

```

187 0.0305019305019305,frozenset({'HOT WATER BOTTLE KEEP CALM'})
188 0.02413127413127413,frozenset({'ALARM CLOCK BAKELIKE RED ', 'ALARM CLOCK BAKELIKE GREEN'})"
189 0.020463320463320462,"frozenset({'WOODEN PICTURE FRAME WHITE FINISH', 'WOODEN FRAME ANTIQUE WHITE '})"
190 0.022702702702702703,"frozenset({'LUNCH BAG RED RETROSPOT', 'JUMBO BAG RED RETROSPOT'})"
191 0.03216216216216216,"frozenset({'JUMBO BAG PINK POLKADOT', 'JUMBO BAG RED RETROSPOT'})"
192 0.02,"frozenset({'JUMBO BAG PINK POLKADOT', 'JUMBO STORAGE BAG SUKI'})"
193 0.0283011583011583,"frozenset({'JUMBO BAG RED RETROSPOT', 'JUMBO STORAGE BAG SUKI'})"
194 0.022895752895752897,"frozenset({'JUMBO BAG RED RETROSPOT', 'JUMBO BAG BAROQUE BLACK WHITE'})"
195 0.020154440154440154,"frozenset({'JUMBO BAG RED RETROSPOT', 'JUMBO BAG PINK VINTAGE PAISLEY'})"
196 0.02637065637065637,"frozenset({'JUMBO BAG RED RETROSPOT', 'JUMBO SHOPPER VINTAGE RED PAISLEY'})"
197 0.02027027027027027,"frozenset({'JUMBO SHOPPER VINTAGE RED PAISLEY', 'JUMBO STORAGE BAG SUKI'})"
198 0.020733590733590733,"frozenset({'PAPER CHAIN KIT 50'S CHRISTMAS '", 'PAPER CHAIN KIT VINTAGE CHRISTMAS'})"
199 0.021235521235521235,"frozenset({'LUNCH BAG RED RETROSPOT', 'LUNCH BAG SPACEBOY DESIGN '})"
200 0.023667953667953667,"frozenset({'LUNCH BAG RED RETROSPOT', 'LUNCH BAG PINK POLKADOT'})"
201 0.02111969111969112,"frozenset({'LUNCH BAG BLACK SKULL.', 'LUNCH BAG PINK POLKADOT'})"
202 0.02054054054054054,"frozenset({'LUNCH BAG RED RETROSPOT', 'LUNCH BAG WOODLAND'})"
203 0.025019305019305018,"frozenset({'LUNCH BAG RED RETROSPOT', 'LUNCH BAG BLACK SKULL.'})"
204 0.020308880308880308,"frozenset({'LUNCH BAG CARS BLUE', 'LUNCH BAG BLACK SKULL.'})"
205 0.02166023166023166,"frozenset({'LUNCH BAG RED RETROSPOT', 'LUNCH BAG CARS BLUE'})"
206 0.021042471042471044,"frozenset({'JUMBO BAG RED RETROSPOT', 'JUMBO BAG STRAWBERRY'})"
207 0.020386100386100385,"frozenset({'SPACEBOY LUNCH BOX ', 'DOLLY GIRL LUNCH BOX'})"
208 0.020077220077220077,"frozenset({'REGENCY CAKESTAND 3 TIER', 'ROSES REGENCY TEACUP AND SAUCER '})"
209 0.029536679536679537,"frozenset({'GREEN REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER '})"
210 0.02,"frozenset({'REGENCY CAKESTAND 3 TIER', 'GREEN REGENCY TEACUP AND SAUCER'})"
211 0.020772200772200773,"frozenset({'JUMBO BAG RED RETROSPOT', 'JUMBO BAG WOODLAND ANIMALS'})"
212 0.020308880308880308,"frozenset({'CHARLOTTE PINK POLKADOT', 'RED RETROSPOT CHARLOTTE BAG'})"
213 0.024864864864864864,"frozenset({'GREEN REGENCY TEACUP AND SAUCER', 'PINK REGENCY TEACUP AND SAUCER'})"
214 0.023127413127413128,"frozenset({'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER '})"
215 0.020733590733590733,"frozenset({'GREEN REGENCY TEACUP AND SAUCER', 'PINK REGENCY TEACUP AND SAUCER', 'ROSES REGENCY TEACUP AND SAUCER '})"
216 0.021351351351351352,"frozenset({'LUNCH BAG RED RETROSPOT', 'LUNCH BAG SUKI DESIGN '})"
217 0.02142857142857143,"frozenset({'JUMBO BAG APPLES', 'JUMBO BAG RED RETROSPOT'})"

```

Hình 4. 20 Các mẫu phổ biến

Khi nhập vào một sản phẩm cần gợi ý thì kết quả sẽ cho ra:

```
Top predicted items based on user input ['JUMBO BAG APPLES']:  
{ 'JUMBO BAG APPLES', 'JUMBO BAG RED RETROSPOT' }
```

Hình 4. 21 Kết quả gợi ý sao khi dự đoán

Nó sẽ cho ra những sản phẩm có khả năng cao nhất nhờ vào bảng support

#### 4.5 Phân tích và nhận xét

Các mặt hàng có giá trị hỗ trợ cao như "WHITE HANGING HEART T-LIGHT HOLDER" và "KNITTED UNION FLAG HOT WATER BOTTLE" đều có xu hướng phổ biến trong dữ liệu mua sắm. Điều này có thể là các sản phẩm phổ biến và được nhiều khách hàng chọn mua.

Các itemset như "JUMBO BAG RED RETROSPOT, JUMBO BAG VINTAGE DOILY" có giá trị hỗ trợ cao, cho thấy xu hướng mua sắm nhóm các sản phẩm này cùng nhau. Điều này có thể thể hiện tính tương tác cao giữa các mặt hàng.

Các itemset có thể được sử dụng để đưa ra gợi ý sản phẩm. Ví dụ, nếu một khách hàng mua "LUNCH BAG VINTAGE DOILY", chúng ta có thể đề xuất gói sản phẩm có chứa "JUMBO BAG VINTAGE DOILY".

Các thông tin này có thể hỗ trợ chiến lược bán hàng và quảng cáo, giúp doanh nghiệp tối ưu hóa cách họ trình bày và kết hợp sản phẩm.

## Chương 5: KẾT LUẬN VÀ KIẾN NGHỊ

Dựa trên việc phân tích và gợi ý sản phẩm trong thương mại điện tử sử dụng thuật toán FP-growth, chúng ta có thể thấy rằng:

### **Mục tiêu chính:**

Thu thập và xử lý dữ liệu thành công để áp dụng thuật toán.

Xây dựng thành công giải thuật fp-growth để đưa ra gợi ý sản phẩm.

Áp dụng thuật toán FP-growth đã giúp chúng ta hiểu rõ mẫu mua sắm và xu hướng tương tác giữa các sản phẩm trong dữ liệu thương mại điện tử.

### **Mục tiêu phụ:**

Xây dựng thành công cơ sở dữ liệu.

Ứng dụng Django để phát triển web.

Thành công xây dựng giao diện người dùng.

Nhận biết được những sản phẩm phổ biến và thường xuyên xuất hiện trong các đơn hàng, giúp doanh nghiệp tập trung vào quảng cáo và khuyến mãi cho những mặt hàng quan trọng này.

Kết quả từ FP-growth có thể hỗ trợ xây dựng chiến lược bán hàng, đặc biệt là trong việc đề xuất sản phẩm và tăng cường trải nghiệm mua sắm của khách hàng.

Quy luật kết hợp và độ tin cậy của chúng có thể làm nổi bật những mối quan hệ tương đối mạnh mẽ giữa các mặt hàng, giúp doanh nghiệp hiểu rõ hơn về sự liên kết giữa sản phẩm.

Các itemsets có thể tích hợp vào hệ thống gợi ý sản phẩm, giúp tối ưu hóa trải nghiệm mua sắm trực tuyến và thúc đẩy doanh số bán hàng.

Thông qua việc phân tích và gợi ý sản phẩm, chúng ta có thể tìm hiểu rõ hơn về xu hướng mua sắm và điều chỉnh chiến lược kinh doanh theo thời gian.

Việc áp dụng FP-growth trong phân tích mẫu mua sắm và gợi ý sản phẩm có thể mang lại những thông tin quan trọng giúp doanh nghiệp hiểu rõ hơn về thị trường và tối ưu hóa chiến lược kinh doanh của họ.

**Dựa trên kết quả và trải nghiệm thu được từ nghiên cứu, em đề xuất một số điểm để cải thiện và phát triển hệ thống:**

Nghiên cứu thêm về cách tối ưu hóa tham số của thuật toán FP-growth để đảm bảo hiệu suất cao và khả năng mở rộng tốt trên các bộ dữ liệu lớn.

Xây dựng một cơ chế phản hồi người dùng để cải thiện tính cá nhân hóa của mô hình gợi ý sản phẩm và tăng cường sự tương tác.

Xem xét khả năng phát triển ứng dụng di động để mở rộng phạm vi sử dụng và cung cấp trải nghiệm tốt hơn cho người dùng di động.

Nghiên cứu và phát triển khả năng tích hợp với các nền tảng thương mại điện tử phổ biến khác nhau để mở rộng ảnh hưởng của hệ thống.

Sau cùng em muốn gửi lời cảm ơn đến thầy Bùi Tiến Đức đã giúp em hoàn thành bài khóa luận. Xin cảm ơn các quý thầy cô khoa công nghệ thông tin và các bạn đã hỗ trợ em hoàn thành bài khóa luận một cách thành công.



## DANH MỤC TÀI LIỆU THAM KHẢO

Các đường link:

Mã nguồn của bài: <https://github.com/belam29/web-django-store-fpGrowth>

[1] VIJAYKUMAR UMMADISETTY, 2018, Bộ dữ liệu hóa đơn của cửa hàng bán lẻ trực tuyến được sử dụng, Đường dẫn:

<https://www.kaggle.com/datasets/vijayuv/onlineretail?resource=download>

[2] M. Kavitha, S. T. Tamil, tháng 8 năm 2016, “Selvi Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons”. Đường dẫn:

<https://www.ijcstjournal.org/volume-4/issue-4/IJCST-V4I4P28.pdf> [3] Le Hoang,

Thuật toán FP-growth, 15/12/2023. Đường dẫn:

<https://viblo.asia/p/khai-pha-du-lieu-va-lop-bai-toan-khai-thac-cac-tap-pho-bien-p2-m68Z0W06KkG>

[4] W3School, 23/12/2023, Django FrameWork. Đường dẫn:

<https://www.w3schools.com/django/>