|  |
| --- |
| **BỘ GIAO THÔNG VẬN TẢI**  **HỌC VIỆN HÀNG KHÔNG VIỆT NAM**  **KHOA CÔNG NGHỆ THÔNG TIN**    **TIỂU LUẬN**  **(ĐÁNH GIÁ LÀM VIỆC NHÓM)**  **KHAI THÁC DỮ LIỆU**  **HỆ THỐNG PHÂN TÍCH HÀNH VI MUA SẮM CỦA KHÁCH HÀNG**  **Giảng viên hướng dẫn: ThS. Trần Anh Tuấn**  **Sinh viên/ Nhóm sinh viên thực hiện: Nhóm 9**  **Mã số sinh viên:…………………………………………………………………….**  **Lớp: 010100087301**  **TP.Hồ Chí Minh, tháng 03/2025** |
| **BỘ GIAO THÔNG VẬN TẢI**  **HỌC VIỆN HÀNG KHÔNG VIỆT NAM**  **KHOA CÔNG NGHỆ THÔNG TIN**    **TIỂU LUẬN**  **(ĐÁNH GIÁ LÀM VIỆC NHÓM)**  **KHAI THÁC DỮ LIỆU**  **HỆ THỐNG PHÂN TÍCH HÀNH VI MUA SẮM CỦA KHÁCH HÀNG**  **Giảng viên hướng dẫn: ThS. Trần Anh Tuấn**  **Sinh viên/ Nhóm sinh viên thực hiện: Nhóm 9**  **Mã số sinh viên:…………………………………………………………………….**  **Lớp: 010100087301**  **TP.Hồ Chí Minh, tháng 03/2025** |

**Danh sách Nhóm:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Họ và tên** | **MSSV** | **Lớp** | **Ghi chú** |
| 1 | Dương Quốc Anh | 2254810294 | 22ĐHTT06 | Nhóm Trưởng |
| 2 |  |  |  | Thành Viên |
| 3 |  |  |  | Thành Viên |
| 4 |  |  |  | Thành Viên |
| 5 |  |  |  | Thành Viên |

|  |  |
| --- | --- |
| **Cán bộ chấm thi 1**  *(ký và ghi rõ họ tên)* | **Cán bộ chấm thi 2**  *(ký và ghi rõ họ tên)* |
| **Cán bộ chấm thi phúc khảo 1**  *(ký và ghi rõ họ tên)* | **Cán bộ chấm thi phúc khảo 2**  *(ký và ghi rõ họ tên)* |

**DANH MỤC CÁC KÝ HIỆU, CHỮ VIẾT TẮT**

|  |  |
| --- | --- |
| **Ký hiệu, chữ viết tắt** | **Chữ viết đầy đủ** |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# **MỤC LỤC**

### ****Danh mục các ký hiệu, chữ viết tắt****

### ****Danh mục các bảng, hình vẽ, đồ thị****

## **MỞ ĐẦU**

## **CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI**

### 1.1. Giới thiệu về đề tài

#### 1.1.1. Lý do chọn đề tài

#### 1.1.2. Mục tiêu nghiên cứu

### 1.2. Tổng quan về các nghiên cứu liên quan

#### 1.2.1. Các nghiên cứu trong nước

#### 1.2.2. Các nghiên cứu quốc tế

### 1.3. Phạm vi nghiên cứu

## **CHƯƠNG 2: CƠ SỞ LÝ THUYẾT**

### 2.1. Ngôn ngữ lập trình C# trong Unity

#### 2.1.1. Giới thiệu về C#

#### 2.1.2. Ứng dụng của C# trong Unity

#### 2.1.3. Điểm mạnh và điểm yếu của C# trong Unity

### 2.2. Blender – Công cụ thiết kế mô hình 3D

#### 2.2.1. Lịch sử hình thành và phát triển

#### 2.2.2. Khái niệm và đặc điểm

#### 2.2.3. Điểm mạnh và điểm yếu của Blender

### 2.3. Weka – Công cụ phân tích dữ liệu

#### 2.3.1. Giới thiệu về Weka

#### 2.3.2. Các thuật toán phân tích hành vi mua sắm trong Weka

#### 2.3.3. Ứng dụng của Weka trong phân tích dữ liệu

**2.4. Mô hình phát triển hệ thống phân tích hành vui mua sắm**

#### ****2.4.1. Tổng quan về hệ thống phân tích hành vi mua sắm****

### ****2.4.2. Các yếu tố quan trọng trong game nông trại****

### ****2.4.3. Điểm mạnh và điểm yếu của thể loại game nông trại****

## **CHƯƠNG 3: PHÂN TÍCH HỆ THỐNG VÀ XÂY DỰNG SẢN PHẨM**

**3.1. Phân tích hệ thống**

**3.1.1. Các tác nhân trong hệ thống**

**3.1.1.1. Khách hàng (Customer)**

**3.1.1.2. Hệ thống phân tích dữ liệu (Data Analytics System)**

**3.1.1.3. Hệ thống quản lý sản phẩm (Product Management System)**

**3.1.1.4. Hệ thống lập lịch (Scheduling System)**

**3.1.2. User Case**

**3.1.4. Mô hình hoạt động**

**3.1.4.1. Khách hàng**

**3.1.4.2. Dữ liệu mua sắm**

**3.1.4.3. Hệ thống phân tích hành vi**

**3.1.4.5. Hệ thống khuyến mãi và ưu đãi**

**3.1.4.6. Hệ thống thương mại điện tử**

**3.1.5. Sơ đồ tuần tự các chức năng**

**3.2. Thiết kế cơ sở dữ liệu**

**3.2.1. Bảng Users (Khách hàng)**

**3.2.2. Bảng Products (Sản phẩm)**

**3.2.3. Bảng Orders (Đơn hàng)**

**3.2.4. Bảng OrderDetails (Chi tiết đơn hàng)**

**3.2.5. Bảng Transactions (Thanh toán)**

**3.2.6. Bảng CustomerBehavior (Hành vi khách hàng)**

**3.2.7. Bảng Promotions (Chương trình khuyến mãi)**

## **3.2.8. Bảng UserPromotions (Khuyến mãi của khách hàng)**

## **CHƯƠNG 4: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

## **4.1. Kết luận**

### 4.2. Hướng phát triển

## **DANH MỤC TÀI LIỆU THAM KHẢO**

## **PHỤ LỤC**

**MỞ ĐẦU**

Trong thời đại số hóa, dữ liệu đóng vai trò quan trọng trong việc hỗ trợ các doanh nghiệp hiểu rõ hơn về khách hàng và đưa ra các chiến lược kinh doanh hiệu quả. Đặc biệt, hành vi mua sắm của khách hàng là một trong những yếu tố cốt lõi ảnh hưởng đến doanh thu, sự phát triển và cạnh tranh của doanh nghiệp. Việc khai thác dữ liệu để phân tích hành vi mua sắm không chỉ giúp doanh nghiệp nhận diện xu hướng tiêu dùng, mà còn tối ưu hóa chiến lược tiếp thị, cá nhân hóa trải nghiệm khách hàng và nâng cao hiệu quả kinh doanh.

Trong bối cảnh đó, bài tiểu luận này sẽ tập trung vào việc phân tích hành vi mua sắm của khách hàng thông qua các phương pháp khai thác dữ liệu. Cụ thể, nghiên cứu sẽ đề cập đến các kỹ thuật thu thập, xử lý và phân tích dữ liệu để xác định các mô hình mua sắm, yếu tố tác động đến quyết định mua hàng cũng như dự đoán xu hướng tiêu dùng trong tương lai.

Thông qua bài tiểu luận, người đọc sẽ có cái nhìn tổng quan về quá trình khai thác dữ liệu trong lĩnh vực thương mại, những ứng dụng thực tế của nó trong hoạt động kinh doanh và các thách thức trong việc phân tích dữ liệu khách hàng. Đây sẽ là cơ sở quan trọng để đưa ra các giải pháp nhằm nâng cao trải nghiệm mua sắm và tối ưu hóa chiến lược tiếp thị của doanh nghiệp.

**CHƯƠNG 1. GIỚI THIỆU**

**1.1. Lý do chọn đề tài**

Trong bối cảnh kinh tế số phát triển mạnh mẽ, hành vi mua sắm của khách hàng ngày càng trở nên phức tạp và đa dạng hơn. Việc khai thác và phân tích dữ liệu mua sắm không chỉ giúp doanh nghiệp hiểu rõ nhu cầu, sở thích của khách hàng mà còn đóng vai trò quan trọng trong việc xây dựng chiến lược kinh doanh hiệu quả, tối ưu hóa trải nghiệm khách hàng và gia tăng lợi nhuận.

Môn **Khai thác dữ liệu** cung cấp các phương pháp và kỹ thuật giúp trích xuất thông tin hữu ích từ dữ liệu, hỗ trợ doanh nghiệp trong việc dự đoán xu hướng mua sắm, phân loại khách hàng và đề xuất các chiến lược tiếp thị phù hợp. Việc nghiên cứu đề tài **“Phân tích hành vi mua sắm của khách hàng”** không chỉ giúp ứng dụng các thuật toán khai thác dữ liệu vào thực tế mà còn rèn luyện kỹ năng xử lý dữ liệu lớn, trực quan hóa thông tin và đưa ra các quyết định dựa trên dữ liệu.

Bên cạnh đó, đề tài này có tính ứng dụng thực tiễn cao, giúp doanh nghiệp tối ưu hóa hoạt động kinh doanh, cá nhân hóa trải nghiệm khách hàng và nâng cao khả năng cạnh tranh trên thị trường. Vì những lý do trên, việc nghiên cứu đề tài này không chỉ mang lại giá trị về mặt học thuật mà còn mở ra nhiều cơ hội trong lĩnh vực khoa học dữ liệu và kinh doanh.

**1.2. Mục tiêu đề tài**

Đề tài **“Phân tích hành vi mua sắm của khách hàng”** hướng đến các mục tiêu sau:

**Tìm hiểu và ứng dụng các kỹ thuật khai thác dữ liệu** trong việc phân tích hành vi mua sắm của khách hàng.

**Thu thập, xử lý và phân tích dữ liệu mua sắm**, xác định các yếu tố ảnh hưởng đến quyết định mua hàng.

**Xây dựng mô hình dự đoán xu hướng tiêu dùng** dựa trên các phương pháp khai thác dữ liệu như phân cụm khách hàng, phân tích giỏ hàng và dự đoán hành vi mua sắm.

**Đánh giá hiệu quả và ứng dụng thực tiễn** của việc phân tích dữ liệu trong chiến lược kinh doanh và marketing.

**1.3. Phạm vi đề tài**

Phạm vi nghiên cứu của đề tài tập trung vào:

**Ứng dụng các phương pháp khai thác dữ liệu** trong việc phân tích hành vi mua sắm của khách hàng, bao gồm thu thập, xử lý và phân tích dữ liệu từ các giao dịch mua sắm.

**Đối tượng nghiên cứu chủ yếu** là các kỹ thuật khai thác dữ liệu như phân cụm khách hàng, phân tích giỏ hàng (Market Basket Analysis) và dự đoán hành vi tiêu dùng dựa trên dữ liệu lịch sử.

**Đề tài chỉ tập trung vào khai thác dữ liệu khách hàng**, không đi sâu vào các yếu tố tâm lý hay xã hội ảnh hưởng đến hành vi mua sắm và không mở rộng sang các lĩnh vực ngoài thương mại điện tử hoặc bán lẻ.

**1.4 Đối tượng nghiên cứu**

Đối tượng nghiên cứu của đề tài bao gồm:

**Các kỹ thuật khai thác dữ liệu** trong phân tích hành vi mua sắm, bao gồm phương pháp phân cụm khách hàng, phân tích giỏ hàng (Market Basket Analysis) và dự đoán xu hướng tiêu dùng.

**Dữ liệu mua sắm của khách hàng**, có thể bao gồm thông tin giao dịch, tần suất mua hàng, danh mục sản phẩm và các yếu tố ảnh hưởng đến quyết định mua sắm.

**Các bên liên quan trong quá trình phân tích dữ liệu**, bao gồm doanh nghiệp, nhà tiếp thị, nhà khoa học dữ liệu và khách hàng – những người trực tiếp tạo ra và chịu ảnh hưởng bởi dữ liệu mua sắm.

**1.5. Phương pháp nghiên cứu**

Để thực hiện đề tài, các phương pháp nghiên cứu sau sẽ được áp dụng:

**Phương pháp nghiên cứu tài liệu:** Thu thập và phân tích các tài liệu, sách, bài viết khoa học về khai thác dữ liệu, các mô hình phân tích hành vi mua sắm và các thuật toán liên quan.

**Phương pháp khảo sát và thống kê:** Nghiên cứu dữ liệu mua sắm từ các nguồn thực tế, phân tích xu hướng tiêu dùng, tần suất mua hàng và các yếu tố ảnh hưởng đến quyết định mua sắm.

**Phương pháp phân tích và tổng hợp:** So sánh các phương pháp khai thác dữ liệu để lựa chọn mô hình phù hợp trong việc phân tích hành vi mua sắm của khách hàng.

**Phương pháp thử nghiệm mô hình:** Xây dựng, kiểm thử và đánh giá hiệu quả của các thuật toán khai thác dữ liệu trong việc phân loại khách hàng và dự đoán xu hướng tiêu dùng.

.**1.6. Bố cục đề tài**

Báo cáo gồm các chương sau:

**Chương 1: Giới thiệu đề tài** – Trình bày lý do chọn đề tài, mục tiêu nghiên cứu, phạm vi nghiên cứu, đối tượng nghiên cứu, phương pháp nghiên cứu và bố cục báo cáo.

**Chương 2: Cơ sở lý thuyết** – Tổng quan về khai thác dữ liệu, các khái niệm liên quan đến hành vi mua sắm của khách hàng, các thuật toán khai thác dữ liệu và phương pháp phân tích dữ liệu trong thương mại điện tử.

**Chương 3: Phân tích và triển khai mô hình** – Phân tích dữ liệu mua sắm, áp dụng các thuật toán khai thác dữ liệu để phân loại khách hàng, dự đoán hành vi mua sắm và đánh giá hiệu quả mô hình.

**Kết luận** – Tổng kết những điểm chính của nghiên cứu, đánh giá kết quả đạt được và đề xuất hướng phát triển trong tương lai cho việc ứng dụng khai thác dữ liệu trong phân tích hành vi mua sắm.

**CHƯƠNG 2. CƠ SỞ LÝ THUYẾT**

**2.1. Giới thiệu về Unity**

**2.1.1. Lịch sử hình thành và phát triển**

Khai thác dữ liệu (Data Mining) bắt đầu phát triển từ những năm 1980, khi các doanh nghiệp nhận thấy tiềm năng của việc sử dụng dữ liệu để đưa ra quyết định kinh doanh. Ban đầu, kỹ thuật này chỉ tập trung vào thống kê và xử lý dữ liệu đơn giản. Tuy nhiên, với sự phát triển của công nghệ máy tính và trí tuệ nhân tạo, khai thác dữ liệu đã mở rộng sang nhiều lĩnh vực như tài chính, y tế và thương mại điện tử. Ngày nay, các thuật toán khai thác dữ liệu đóng vai trò quan trọng trong việc phân tích hành vi mua sắm, giúp doanh nghiệp hiểu rõ hơn về khách hàng và tối ưu hóa chiến lược kinh doanh.

**2.1.2. Khái niệm và đặc điểm**

Khai thác dữ liệu là quá trình tìm kiếm và trích xuất thông tin hữu ích từ tập dữ liệu lớn, hỗ trợ đưa ra các quyết định kinh doanh dựa trên dữ liệu. Trong phân tích hành vi mua sắm, khai thác dữ liệu giúp xác định xu hướng tiêu dùng, phân loại khách hàng và dự đoán quyết định mua hàng. Một số đặc điểm chính của khai thác dữ liệu bao gồm:

**Phân tích dựa trên dữ liệu lớn:** Xử lý và phân tích một lượng lớn dữ liệu mua sắm từ các hệ thống bán hàng và thương mại điện tử.

**Sử dụng các thuật toán thông minh:** Ứng dụng các phương pháp như phân cụm khách hàng (Clustering), phân tích giỏ hàng (Market Basket Analysis) và học máy (Machine Learning).

**Hỗ trợ ra quyết định:** Giúp doanh nghiệp cá nhân hóa chiến lược tiếp thị, đề xuất sản phẩm và tối ưu hóa trải nghiệm khách hàng.

**Ứng dụng rộng rãi:** Được sử dụng trong các lĩnh vực như bán lẻ, thương mại điện tử, ngân hàng và tiếp thị kỹ thuật số.

### ****2.1.3. Điểm mạnh và điểm yếu của Unity****

**Điểm mạnh:**  
 Giúp doanh nghiệp hiểu rõ hơn về khách hàng và xu hướng tiêu dùng.  
 Hỗ trợ tối ưu hóa chiến lược marketing và cá nhân hóa sản phẩm.  
 Nâng cao hiệu quả kinh doanh và gia tăng lợi nhuận.  
 Ứng dụng linh hoạt trong nhiều lĩnh vực.

**Điểm yếu:**  
 Yêu cầu hệ thống lưu trữ và xử lý dữ liệu lớn, có thể tốn kém chi phí.  
 Các mô hình phân tích có thể gặp vấn đề về độ chính xác nếu dữ liệu không đầy đủ hoặc không được làm sạch kỹ lưỡng.  
 Việc khai thác dữ liệu có thể gây lo ngại về quyền riêng tư và bảo mật thông tin khách hàng.

**2.2. Ngôn ngữ lập trình C# trong Unity**

### ****2.2.1. Giới thiệu về Python****

Python là một ngôn ngữ lập trình bậc cao, hướng đối tượng, được phát triển bởi Guido van Rossum vào năm 1991. Python nổi bật với cú pháp đơn giản, dễ đọc và thư viện phong phú, giúp lập trình viên triển khai nhanh chóng các ứng dụng trong nhiều lĩnh vực, đặc biệt là khoa học dữ liệu và khai thác dữ liệu. Nhờ vào khả năng xử lý dữ liệu hiệu quả, Python đã trở thành một trong những ngôn ngữ phổ biến nhất trong phân tích hành vi mua sắm.

### ****2.2.2.**** Ứng dụng của Python trong khai thác dữ liệu

**Xử lý và phân tích dữ liệu:** Sử dụng các thư viện như Pandas, NumPy để thao tác và làm sạch dữ liệu mua sắm.

**Trực quan hóa dữ liệu:** Biểu diễn dữ liệu thông qua biểu đồ và đồ thị bằng Matplotlib, Seaborn.

**Áp dụng các thuật toán học máy:** Sử dụng Scikit-learn để phân cụm khách hàng, phân tích giỏ hàng và dự đoán xu hướng tiêu dùng.

**Xử lý dữ liệu lớn:** Kết hợp với Spark hoặc Dask để làm việc với dữ liệu lớn trong các hệ thống thương mại điện tử.

### ****2.2.3. Điểm mạnh và điểm yếu của C# trong Unity****

**Điểm mạnh:**  
 Cú pháp đơn giản, dễ học và dễ triển khai.  
 Hỗ trợ nhiều thư viện mạnh mẽ cho xử lý và phân tích dữ liệu.  
 Được sử dụng rộng rãi trong khoa học dữ liệu và trí tuệ nhân tạo.  
 Cộng đồng lớn, tài nguyên học tập phong phú.

**Điểm yếu:**  
 Tốc độ xử lý chậm hơn so với C++ hoặc Java trong các ứng dụng thời gian thực.  
 Cần tối ưu hóa khi làm việc với dữ liệu lớn để tránh sử dụng quá nhiều bộ nhớ.  
 Một số thư viện có thể không được tối ưu hóa tốt cho khai thác dữ liệu phức tạp

**2.3. Blender – Công cụ thiết kế mô hình 3D**

### ****2.3.1. Lịch sử hình thành và phát triển****

WEKA (Waikato Environment for Knowledge Analysis) là một phần mềm mã nguồn mở do Đại học Waikato (New Zealand) phát triển, được ra mắt lần đầu vào năm 1993. Ban đầu, WEKA được xây dựng nhằm hỗ trợ nghiên cứu về khai thác dữ liệu và học máy. Trải qua nhiều cải tiến, WEKA đã trở thành một trong những công cụ phổ biến nhất trong lĩnh vực phân tích dữ liệu và trí tuệ nhân tạo, đặc biệt được sử dụng rộng rãi trong giáo dục và nghiên cứu.

### ****2.3.2. Khái niệm và đặc điểm****

WEKA là một phần mềm khai thác dữ liệu mã nguồn mở, cung cấp nhiều công cụ để xử lý dữ liệu, phân tích mẫu và áp dụng các thuật toán học máy. Một số đặc điểm quan trọng của WEKA:

**Hỗ trợ nhiều thuật toán học máy**: Bao gồm phân loại (Classification), phân cụm (Clustering), hồi quy (Regression) và luật kết hợp (Association Rules).

**Tiền xử lý dữ liệu**: Cung cấp các công cụ để làm sạch, chuyển đổi và chuẩn hóa dữ liệu.

**Giao diện trực quan**: Người dùng có thể thao tác trên giao diện đồ họa mà không cần viết code.

**Tích hợp tốt với Java**: Cho phép lập trình viên mở rộng chức năng của WEKA bằng ngôn ngữ Java.

**Mã nguồn mở và miễn phí**: WEKA là một công cụ miễn phí và liên tục được cập nhật bởi cộng đồng nghiên cứu.

### ****2.3.3. Điểm mạnh và điểm yếu của Blender****

##### ****Điểm mạnh:****

**Miễn phí và mã nguồn mở**, có thể mở rộng và tùy chỉnh.  
 **Hỗ trợ nhiều thuật toán học máy**, phù hợp cho phân tích dữ liệu và trí tuệ nhân tạo.  
 **Giao diện trực quan**, dễ sử dụng ngay cả với người không có nhiều kinh nghiệm lập trình.  
 **Cung cấp các công cụ tiền xử lý dữ liệu**, giúp làm sạch và chuẩn bị dữ liệu hiệu quả.  
 **Tích hợp với Java**, có thể sử dụng trong các hệ thống lớn.

##### ****Điểm yếu:****

**Không mạnh về trực quan hóa dữ liệu**, không thể tạo báo cáo đẹp mắt như Power BI hoặc Tableau.  
 **Hiệu suất bị giới hạn**, không phù hợp với dữ liệu quá lớn do chạy trên RAM.  
 **Cần hiểu thuật toán học máy**, để sử dụng hiệu quả cần có kiến thức về khai thác dữ liệu.

**2.4. Mô hình phát triển hệ thống phân tích hành vui mua sắm**

#### ****2.4.1. Tổng quan về hệ thống phân tích hành vi mua sắm****

Hệ thống phân tích hành vi mua sắm của khách hàng là một ứng dụng khai thác dữ liệu nhằm thu thập, phân tích và dự đoán xu hướng mua hàng. Hệ thống này giúp doanh nghiệp hiểu rõ hơn về sở thích, nhu cầu và hành vi của khách hàng, từ đó tối ưu hóa chiến lược tiếp thị, nâng cao doanh thu và cải thiện trải nghiệm khách hàng. Một số ứng dụng phổ biến của hệ thống này bao gồm:

**Phân loại khách hàng** (Customer Segmentation) để cá nhân hóa chiến lược marketing.

**Dự đoán nhu cầu mua sắm** (Predictive Analytics) giúp quản lý hàng tồn kho hiệu quả.

**Phát hiện hành vi bất thường** (Anomaly Detection) nhằm ngăn chặn gian lận hoặc thay đổi xu hướng tiêu dùng.

### ****2.4.2. Các yếu tố quan trọng trong game nông trại****

**Thu thập dữ liệu khách hàng**: Bao gồm dữ liệu lịch sử mua sắm, hành vi truy cập website, thông tin nhân khẩu học, v.v.

**Tiền xử lý dữ liệu**: Làm sạch, chuẩn hóa và biến đổi dữ liệu để phù hợp với mô hình phân tích.

**Phân tích dữ liệu và khai thác mẫu**: Sử dụng các thuật toán học máy để nhận diện xu hướng, phân nhóm khách hàng và dự đoán nhu cầu mua sắm.

**Trực quan hóa kết quả**: Cung cấp bảng điều khiển (dashboard) và báo cáo để hỗ trợ ra quyết định.

### ****2.4.3. Điểm mạnh và điểm yếu của thể loại game nông trại****

##### ****Điểm mạnh:****

**Cải thiện chiến lược kinh doanh**: Hỗ trợ doanh nghiệp hiểu rõ khách hàng, tối ưu hóa chiến lược marketing.  
 **Cá nhân hóa trải nghiệm khách hàng**: Đề xuất sản phẩm phù hợp dựa trên hành vi mua sắm.  
 **Tăng hiệu quả quản lý hàng hóa**: Dự đoán nhu cầu giúp kiểm soát tồn kho hợp lý.  
 **Ngăn chặn gian lận và phát hiện hành vi bất thường**: Bảo vệ doanh nghiệp khỏi các hoạt động mua sắm không minh bạch.

##### ****Điểm yếu:****

##### ****Yêu cầu lượng dữ liệu lớn và chất lượng cao:**** để đảm bảo mô hình phân tích chính xác. ****Cần kiến thức về khoa học dữ liệu và học máy:**** để khai thác tối đa giá trị từ hệ thống. ****Có thể gặp vấn đề về quyền riêng tư và bảo mật dữ liệu:**** nếu không có chính sách quản lý phù hợp.

**CHƯƠNG 3. PHÂN TÍCH HỆ THỐNG VÀ XÂY DỰNG SẢN PHẨM**

**3.1. Phân tích hệ thống**

**3.1.1. Các tác nhân trong hệ thống**

**3.1.1.1. Khách hàng (Customer)**

Thực hiện các hoạt động mua sắm, bao gồm tìm kiếm sản phẩm, thêm vào giỏ hàng, thanh toán và đánh giá sản phẩm.

Tương tác với hệ thống qua các gợi ý sản phẩm dựa trên lịch sử mua sắm và sở thích cá nhân.

Nhận các chương trình khuyến mãi, đề xuất cá nhân hóa dựa trên phân tích hành vi mua hàng..

**3.1.1.2. Hệ thống phân tích dữ liệu (Data Analytics System)**

Thu thập và xử lý dữ liệu từ lịch sử mua sắm, hành vi duyệt web và các thông tin liên quan của khách hàng.

Áp dụng các thuật toán khai thác dữ liệu (như phân cụm khách hàng, dự đoán nhu cầu mua sắm) để tạo ra các báo cáo hữu ích.

Cung cấp thông tin chi tiết cho bộ phận tiếp thị và quản lý để tối ưu hóa chiến lược kinh doanh.

**3.1.1.3. Hệ thống quản lý sản phẩm (Product Management System)**

Kiểm soát kho hàng, quản lý số lượng sản phẩm còn lại và cập nhật thông tin sản phẩm trên nền tảng.

Đưa ra đề xuất nhập hàng dựa trên dự đoán nhu cầu tiêu thụ của khách hàng.

Tích hợp với hệ thống phân tích để tự động điều chỉnh danh mục sản phẩm theo xu hướng thị trường

**3.1.1.4. Hệ thống lập lịch (Scheduling System)**

Cung cấp danh sách sản phẩm phù hợp dựa trên sở thích và hành vi mua sắm của từng khách hàng.

Ứng dụng các thuật toán lọc cộng tác (Collaborative Filtering) hoặc lọc nội dung (Content-Based Filtering) để cá nhân hóa trải nghiệm mua sắm.

Tự động cập nhật danh mục gợi ý dựa trên xu hướng mua sắm thời gian thực

**3.1.2. User Case**

**Phân tích hành vi mua sắm của khách hàng**

**Mô tả ngắn**: Hệ thống phân tích dữ liệu mua sắm để đưa ra gợi ý sản phẩm, dự đoán nhu cầu và phân nhóm khách hàng.

**Tác nhân tham gia**: Khách hàng, hệ thống phân tích dữ liệu.

**Tiến trình**:

Hệ thống thu thập dữ liệu từ lịch sử mua sắm và hành vi duyệt web của khách hàng.

Phân tích dữ liệu bằng các thuật toán phân cụm (clustering) và dự đoán nhu cầu mua sắm.

Gợi ý sản phẩm phù hợp dựa trên sở thích cá nhân.

Hiển thị khuyến mãi và chiến dịch tiếp thị cá nhân hóa.



*Hình 3.1.2.Sơ đồ use case*

#### ****Gợi ý sản phẩm cá nhân hóa****

**Mô tả ngắn**: Hệ thống đề xuất sản phẩm cho khách hàng dựa trên sở thích và lịch sử mua hàng.

**Tác nhân tham gia**: Khách hàng, hệ thống gợi ý sản phẩm.

**Tiến trình**:

Khách hàng truy cập nền tảng mua sắm trực tuyến.

Hệ thống phân tích lịch sử tìm kiếm và mua hàng.

Đề xuất danh sách sản phẩm phù hợp lên trang chủ hoặc trang sản phẩm.

Khách hàng có thể mua ngay hoặc lưu sản phẩm vào danh sách yêu thích.

**3.1.3. Đặc tả chức năng của User Case**

#### ****Use Case 1: Phân tích hành vi mua sắm của khách hàng****

**Tên chức năng**: Phân tích dữ liệu khách hàng.

**Mô tả chức năng**: Hệ thống thu thập và phân tích dữ liệu để hiểu hành vi mua sắm và dự đoán nhu cầu.

**Tiến trình**:

Thu thập dữ liệu từ lịch sử giao dịch, lượt xem sản phẩm, giỏ hàng.

Áp dụng thuật toán phân tích dữ liệu (phân cụm khách hàng, dự đoán nhu cầu).

Hiển thị báo cáo phân tích cho quản trị viên hoặc bộ phận marketing.

Điều chỉnh chiến lược tiếp thị dựa trên kết quả phân tích.

#### ****Use Case 2: Gợi ý sản phẩm cá nhân hóa****

**Tên chức năng**: Hệ thống gợi ý sản phẩm.

**Mô tả chức năng**: Cung cấp danh sách sản phẩm được cá nhân hóa dựa trên sở thích khách hàng.

**Tiến trình**:

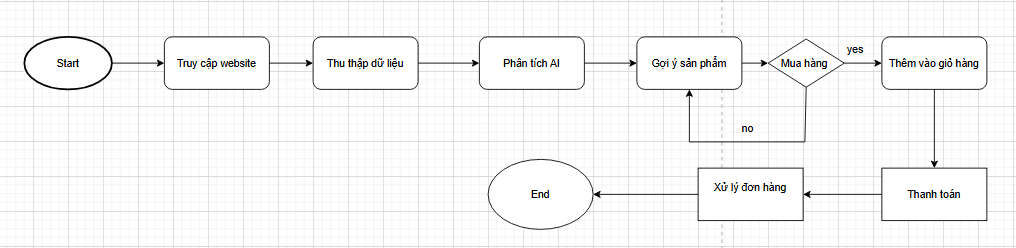
Lấy dữ liệu về lịch sử tìm kiếm, mua hàng và đánh giá sản phẩm.

Áp dụng thuật toán lọc cộng tác (Collaborative Filtering) hoặc lọc nội dung (Content-Based Filtering).

Hiển thị danh sách sản phẩm gợi ý trên trang web hoặc ứng dụng.

Cập nhật danh sách gợi ý dựa trên hành vi mới của khách hàng.

**3.1.4. Mô hình hoạt động**



*Hình 3.1.4.Mô hình hoạt động*

**3.1.4.1. Khách hàng**

**Hoạt động:**

Duyệt sản phẩm trên website/app thương mại điện tử.

Thêm sản phẩm vào giỏ hàng và thanh toán.

Đánh giá sản phẩm sau khi mua.

**Hệ thống cá nhân hóa:**

Gợi ý sản phẩm dựa trên lịch sử mua sắm và sở thích.

Tạo danh sách yêu thích và chương trình khách hàng thân thiết

**3.1.4.2. Dữ liệu mua sắm**

**Dữ liệu thu thập:**

Lịch sử mua hàng của khách hàng.

Hành vi duyệt web, click vào sản phẩm.

Thời gian mua sắm và xu hướng theo mùa.

**Xử lý dữ liệu:**

Phân loại sản phẩm thành các nhóm khác nhau.

Xác định nhóm khách hàng tiềm năng.

Phát hiện xu hướng mua sắm theo thời gian.

**3.1.4.3. Hệ thống phân tích hành vi**

**Các yếu tố phân tích:**

**Tần suất mua hàng:** Khách hàng có xu hướng mua hàng thường xuyên hay không?

**Giỏ hàng bỏ quên:** Phát hiện và gửi nhắc nhở cho khách hàng.

**Xu hướng sản phẩm:** Sản phẩm nào đang được quan tâm nhiều nhất?

**Phân nhóm khách hàng:**

Khách hàng trung thành, khách hàng tiềm năng, khách hàng mới.

Gợi ý chiến dịch khuyến mãi phù hợp với từng nhóm.

**3.1.4.4. Hệ thống thời gian thực**

**Phân tích theo thời gian:**

Sự thay đổi xu hướng theo mùa, ngày lễ, sự kiện đặc biệt.

Tối ưu hóa kho hàng dựa trên nhu cầu dự đoán.

**Chiến lược tiếp thị:**

Đề xuất giảm giá vào thời điểm phù hợp.

Gửi email/SMS cá nhân hóa đến khách hàng.

**3.1.4.5. Hệ thống khuyến mãi và ưu đãi**

**Loại khuyến mãi:**

Giảm giá theo nhóm khách hàng.

Mã giảm giá cho khách hàng trung thành.

Ưu đãi theo dịp đặc biệt như sinh nhật, ngày lễ.

**Quản lý khuyến mãi:**

Đánh giá hiệu quả chiến dịch marketing.

Điều chỉnh ưu đãi dựa trên phản hồi khách hàng.

**3.1.4.6. Hệ thống thương mại điện tử**

**Mua bán trực tuyến:**

Đặt hàng, thanh toán và giao hàng.

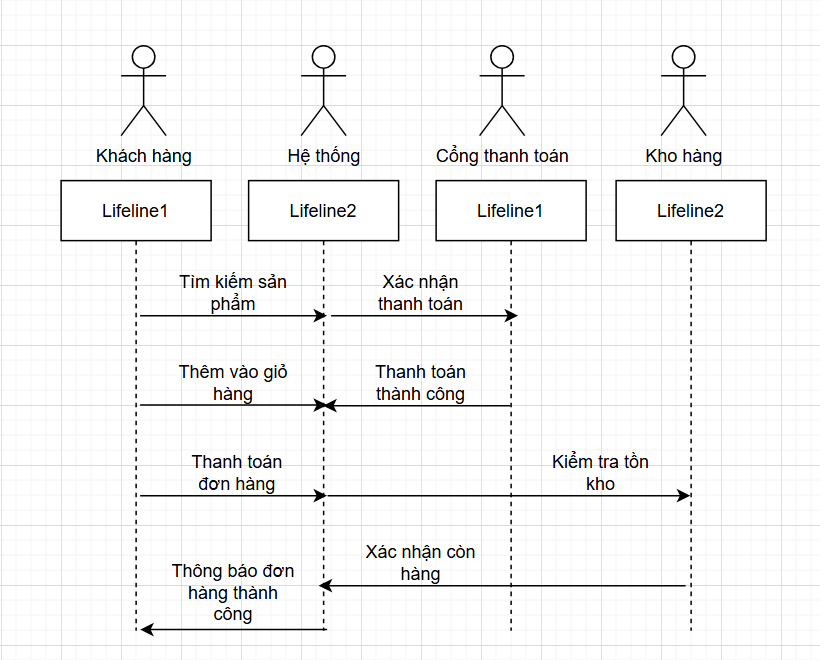
Theo dõi đơn hàng theo thời gian thực.

**Giao dịch & thanh toán:**

Hỗ trợ nhiều phương thức thanh toán (thẻ, ví điện tử, COD).

Đảm bảo bảo mật thông tin thanh toán.

**3.1.5. Sơ đồ tuần tự các chức năng**



**3.2. Thiết kế cơ sở dữ liệu**

Các thực thể chính trong hệ thống:

**Users**: Thông tin khách hàng.

**Products**: Danh mục sản phẩm.

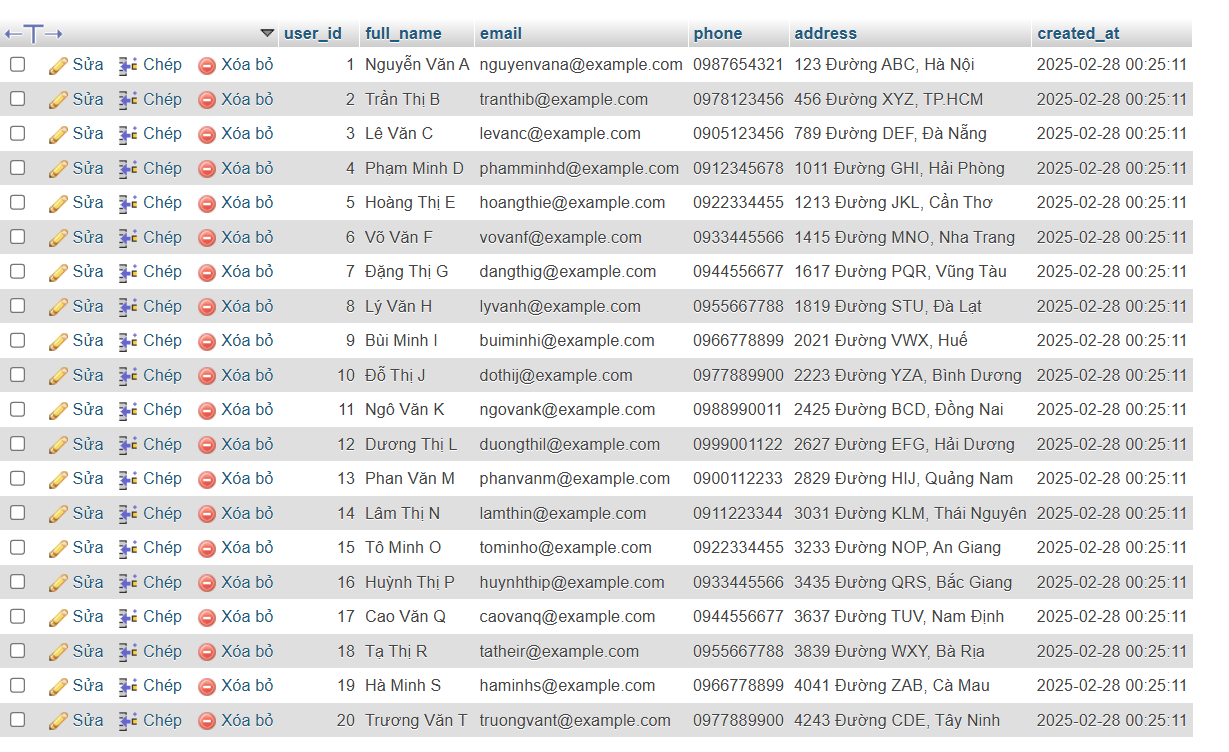
**Orders**: Đơn hàng của khách.

**Transactions**: Giao dịch thanh toán.

**CustomerBehavior**: Lưu hành vi duyệt web và mua hàng.

****Promotions**:** Chương trình khuyến mãi.

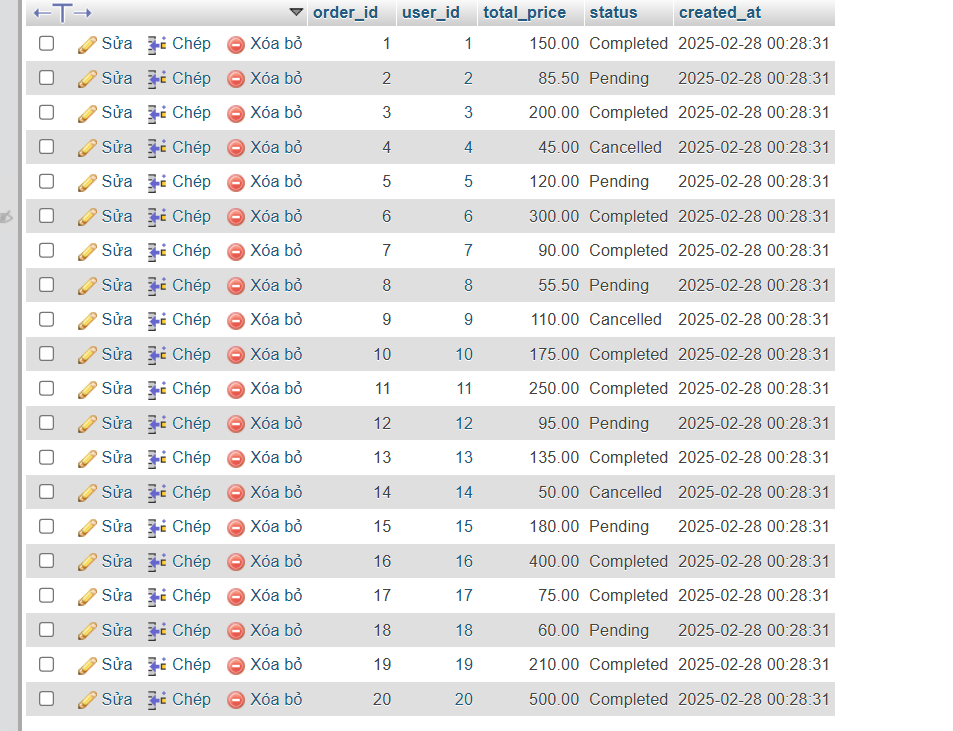
**3.2.1. Bảng Users (Khách hàng)**



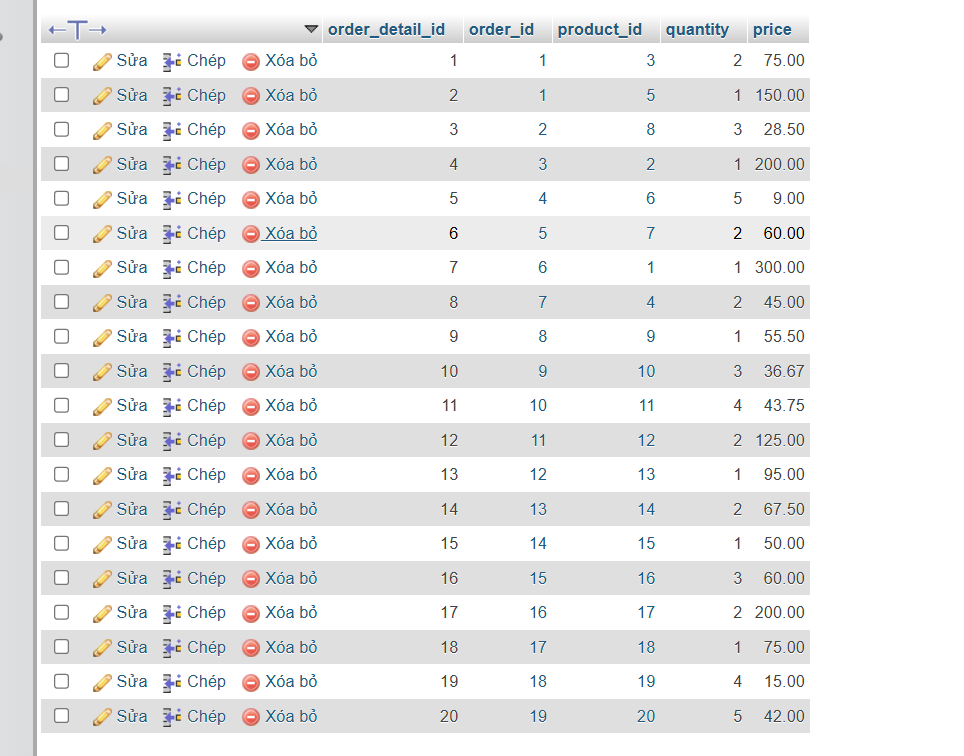
**3.2.2. Bảng Products (Sản phẩm)**



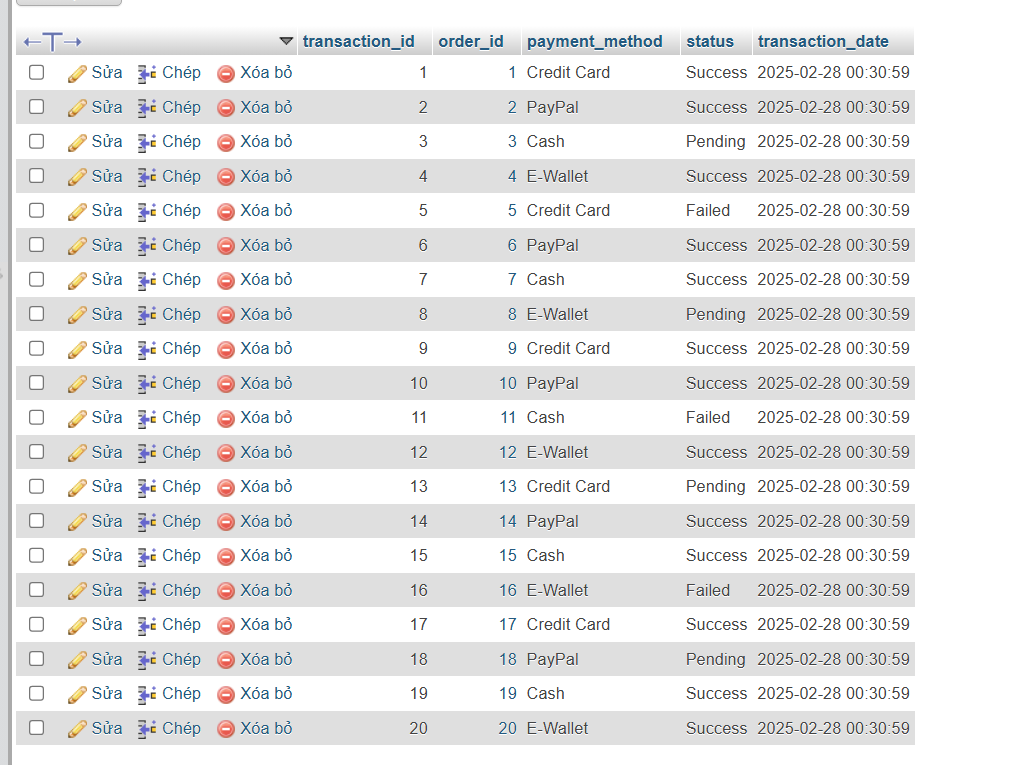
**3.2.3. Bảng Orders (Đơn hàng)**



**3.2.4. Bảng OrderDetails (Chi tiết đơn hàng)**



**3.2.5. Bảng Transactions (Thanh toán)**



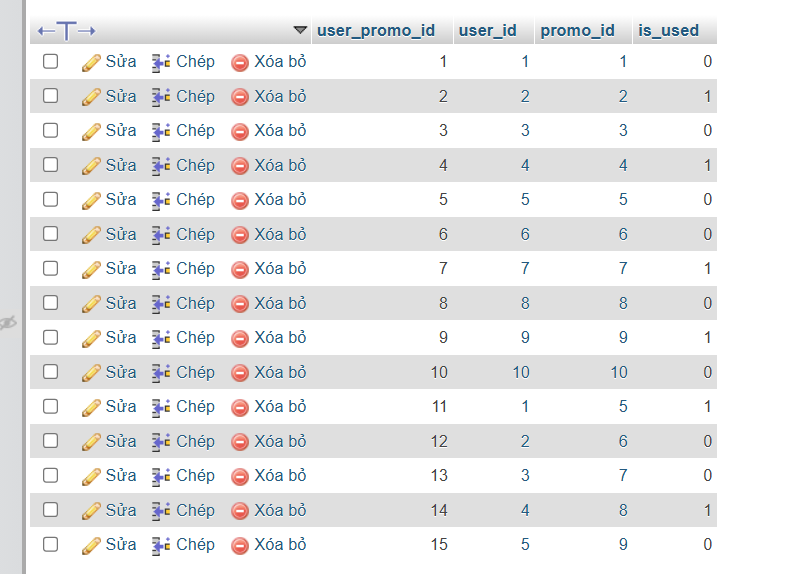
**3.2.6. Bảng CustomerBehavior (Hành vi khách hàng)**



**3.2.7. Bảng Promotions (Chương trình khuyến mãi)**



**3.2.8. Bảng UserPromotions (Khuyến mãi của khách hàng)**



**CHƯƠNG 4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

**4.1. Kết luận**

Trong quá trình thực hiện đề tài "Hệ thống phân tích hành vi mua sắm của khách hàng", nhóm đã nghiên cứu và phát triển một hệ thống có khả năng thu thập, phân tích dữ liệu mua sắm nhằm hỗ trợ doanh nghiệp tối ưu hóa chiến lược kinh doanh.

Cụ thể, hệ thống đã hoàn thành các chức năng chính sau:

**Xây dựng cơ sở dữ liệu** để lưu trữ thông tin khách hàng, sản phẩm, đơn hàng và hành vi mua sắm.

**Ghi nhận hành vi người dùng** như xem sản phẩm, thêm vào giỏ hàng, mua hàng.

**Phân tích dữ liệu** để nhận diện xu hướng tiêu dùng và đề xuất chương trình khuyến mãi phù hợp.

Bên cạnh những kết quả đạt được, hệ thống vẫn còn một số hạn chế:

Chưa áp dụng các mô hình học máy nâng cao để dự đoán hành vi khách hàng.

Chưa tích hợp hệ thống với các nền tảng thương mại điện tử lớn.

Chưa tối ưu hiệu suất xử lý dữ liệu trên tập dữ liệu lớn.

Những kinh nghiệm quan trọng rút ra từ quá trình thực hiện đề tài:

Việc thiết kế cơ sở dữ liệu hợp lý giúp tối ưu hiệu suất truy vấn dữ liệu.

Dữ liệu hành vi khách hàng cần được thu thập đầy đủ và có chiến lược lọc dữ liệu nhiễu.

Các mô hình phân tích cần được điều chỉnh để phù hợp với từng loại dữ liệu cụ thể.

**4.2. Hướng phát triển**

Để cải thiện hệ thống và mở rộng tính năng trong tương lai, nhóm đề xuất các hướng phát triển sau:

### ****1. Hoàn thiện các chức năng hiện có****

**Tối ưu hóa cơ sở dữ liệu**: Cải thiện khả năng truy vấn với các tập dữ liệu lớn bằng cách sử dụng chỉ mục (index) và tối ưu hóa query SQL.

**Cải thiện mô-đun phân tích hành vi**: Bổ sung các thuật toán phân cụm (clustering) như K-means, DBSCAN để nhóm khách hàng theo hành vi.

**Cập nhật hệ thống báo cáo**: Xây dựng giao diện trực quan với dashboard để doanh nghiệp dễ dàng theo dõi dữ liệu mua sắm.

### ****2. Phát triển các tính năng mới****

**Tích hợp trí tuệ nhân tạo (AI)**: Áp dụng các mô hình Machine Learning (như Random Forest, XGBoost) để dự đoán nhu cầu khách hàng và gợi ý sản phẩm.

**Hệ thống cá nhân hóa (Recommendation System)**: Sử dụng thuật toán Collaborative Filtering để đề xuất sản phẩm dựa trên lịch sử mua hàng.

**Phân tích cảm xúc khách hàng**: Thu thập và xử lý phản hồi, đánh giá của khách hàng để đo lường mức độ hài lòng.

### ****3. Mở rộng và tích hợp hệ thống****

**Tích hợp với các nền tảng TMĐT**: Kết nối hệ thống với Shopee, Lazada, Tiki để thu thập dữ liệu thực tế từ nhiều nguồn.

**Triển khai trên môi trường Cloud**: Sử dụng AWS hoặc Google Cloud để tăng hiệu suất xử lý dữ liệu lớn.

**Xây dựng API mở**: Cho phép các nền tảng khác truy xuất dữ liệu từ hệ thống để hỗ trợ việc đồng bộ hóa và mở rộng quy mô.

**TÀI LIỆU THAM KHẢO**

[1] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Morgan Kaufmann, 2011.

[2] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Morgan Kaufmann, 2016.

[3] J. R. Quinlan, "Induction of Decision Trees," Machine Learning, vol. 1, no. 1, pp. 81-106, 1986. DOI: 10.1023/A:1022643204877

[4] P. Tan, M. Steinbach, A. Karpatne, and V. Kumar, Introduction to Data Mining, 2nd ed. Pearson, 2018.

[5] C. Borgelt, "Frequent Item Set Mining," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 2, no. 6, pp. 437-456, 2012. DOI: 10.1002/widm.1074

[6] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," in Proc. 20th Int. Conf. Very Large Data Bases (VLDB'94), Santiago, Chile, 1994, pp. 487-499.

[7] Website chính thức của Weka: https://www.cs.waikato.ac.nz/ml/weka/

[8] Website hướng dẫn sử dụng Weka: https://waikato.github.io/weka-wiki/

**PHỤ LỤC** (nếu có)