

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

---



**TRẦN PHÚC ĐỊNH**

**NGHIÊN CỨU PHÂN TÍCH HÀNH VI NGƯỜI DÙNG  
BỎ GIỎ HÀNG TRÊN CÁC TRANG THƯƠNG MẠI  
ĐIỆN TỬ**

**Chuyên ngành: Khoa học máy tính**

**Mã số: 8.48.01.01**

**TÓM TẮT LUẬN VĂN THẠC SĨ**

**Hà Nội - 2021**

Luận văn được hoàn thành tại:  
**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

Người hướng dẫn khoa học: PGS. TS. Trần Đình Quế

Phản biện 1: PGS. TS. Nguyễn Linh Giang

Phản biện 2: PGS. TS. Nguyễn Đức Dũng

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông  
Vào lúc: 14 giờ 00 ngày 28 tháng 08 năm 2021

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông

# MỞ ĐẦU

## 1. Lý do chọn đề tài

Trong những năm gần đây, cuộc cách mạng khoa học công nghệ 4.0 đã và đang đem lại nhiều thay đổi tích cực trên mọi mặt đời sống của con người. Nhu cầu mua bán, giao dịch trực tuyến qua mạng Internet dần trở nên phổ biến và ngày càng phát triển. Bên cạnh những thuận lợi và tiện nghi không thể phủ nhận, thương mại điện tử vẫn bộc lộ nhiều vấn đề cho khách mua hàng như lòng tin với sản phẩm, bảo mật dữ liệu, sự phụ thuộc vào đơn vị vận chuyển, hay sự sẵn sàng của sản phẩm. Đối với người bán hàng, thách thức lớn nhất không thể không đề cập là vấn đề người dùng bỏ rơi giỏ hàng khi mua sắm trực tuyến.

Bên cạnh những thuận lợi và tiện nghi không thể phủ nhận, thương mại điện tử vẫn bộc lộ nhiều vấn đề cho khách mua hàng như lòng tin với sản phẩm, bảo mật dữ liệu, sự phụ thuộc vào đơn vị vận chuyển, hay sự sẵn sàng của sản phẩm. Đối với người bán hàng, thách thức lớn nhất không thể không đề cập là vấn đề người dùng bỏ rơi giỏ hàng khi mua sắm trực tuyến. Theo một thống kê độc lập từ học viện

Baymard, 69% tổng số lượng giỏ hàng bị bỏ rơi và không được thanh toán trong các giao dịch mua sắm thương mại điện tử. Tỷ lệ bỏ rơi giỏ hàng cao tất yếu sẽ giảm tỷ lệ chuyển đổi đơn hàng, từ đó giảm doanh thu cho người bán nói riêng và cho trang thương mại điện tử nói chung, gia tăng chi phí lưu kho sản phẩm cùng nhiều tác động tiêu cực khác. Không chỉ vậy, người mua hàng sẽ có định kiến tiêu cực và nảy sinh thái độ thận trọng trước tần suất giao dịch hạn chế của người dùng đó tại sàn thương mại điện tử.

## **2. Mục đích nghiên cứu**

- Tìm hiểu kỹ thuật phân loại dữ liệu sử dụng thuật toán học máy có giám sát rừng ngẫu nhiên.
- Nghiên cứu và phát triển một hệ thống tích hợp giữa trang thương mại điện tử và dịch vụ phân tích dự đoán hành vi bỏ rơi giỏ hàng.
- Xây dựng khuôn mẫu hành vi đặc trưng cho phiên mua sắm có giỏ hàng bị bỏ rơi.

### **3. Đối tượng và phạm vi nghiên cứu**

- Đối tượng nghiên cứu là các khách hàng thực hiện giao dịch mua sắm tại trang thương mại điện tử.
- Nghiên cứu trên dữ liệu hoạt động mua sắm và lịch sử giao dịch của người tiêu dùng trực tuyến.
- Tập dữ liệu sử dụng cho huấn luyện và dự đoán là tập dữ liệu nhật ký hành động và lịch sử mua sắm tại trang thương mại điện tử.

### **4. Phương pháp nghiên cứu**

- Nghiên cứu, tìm hiểu lý thuyết về quy trình phân loại của thuật toán rừng ngẫu nhiên.
- Tìm hiểu và xây dựng bảng chỉ mục dữ liệu phục vụ huấn luyện cho tác vụ phân loại hành vi bỏ rơi giỏ hàng.
- Xây dựng hệ thống phân tích và ứng dụng thuật toán rừng ngẫu nhiên để phân loại và dự đoán hành vi bỏ rơi giỏ hàng

## **5. Cấu trúc luận văn**

### **CHƯƠNG 1 - Tổng quan về hành vi người dùng trong thương mại điện tử**

Trình bày tổng quan về thương mại điện tử và tiềm năng khai thác dữ liệu người dùng.

### **CHƯƠNG 2 - Phân tích hành vi bỏ rơi giỏ hàng**

Khái niệm bỏ rơi giỏ hàng và ứng dụng thuật toán rừng ngẫu nhiên trong phân loại dữ liệu.

### **CHƯƠNG 3 - Thử nghiệm và đánh giá**

Ứng dụng thuật toán rừng ngẫu nhiên để dự đoán và đưa ra những phân tích về hành vi bỏ rơi giỏ hàng.

# **CHƯƠNG 1 - TỔNG QUAN VỀ HÀNH VI NGƯỜI DÙNG TRONG THƯƠNG MẠI ĐIỆN TỬ**

## **1.1. Giới thiệu khái niệm, sự ra đời, hình thành và phát triển của thương mại điện tử**

- Khái niệm về thương mại điện tử.
- Sự ra đời, hình thành và phát triển của thương mại điện tử.

## **1.2. Tiềm năng khai thác dữ liệu người dùng trong thương mại điện tử**

- Phân tích giỏ hàng điện tử.
- Dự đoán nhu cầu thị trường.
- Đánh giá phân khúc thị trường.
- Phòng chống gian lận thương mại.

## **1.3. Giới thiệu giỏ hàng điện tử và hành vi bỏ rơi giỏ hàng**

- Khái niệm giỏ hàng điện tử.
- Khuynh hướng sử dụng giỏ hàng điện tử của người tiêu dùng.
- Bỏ rơi giỏ hàng trong mua sắm trực tuyến.

#### **1.4.Kết luận**

Trong chương một, luận văn đã trình bày một cách tổng quan về khái niệm, sự ra đời, hình thành và phát triển của thương mại điện tử trên thế giới. Dựa trên cơ sở đó, luận văn đề cập những tiềm năng khai thác dữ liệu và hành vi tiêu dùng trên các trang thương mại điện tử. Từ những khái niệm và tiềm năng khai thác dữ liệu, chương một cũng giới thiệu đến khái niệm giỏ hàng điện tử trong mua sắm trực tuyến và khuyến hướng sử dụng giỏ hàng điện tử tại Việt Nam; nêu lên hành vi bỏ rơi giỏ hàng được đề để làm cơ sở nghiên cứu trong chương hai.



## **CHƯƠNG 2: PHÂN TÍCH HÀNH VI BỎ RƠI GIỎ HÀNG**

### **2.1. Các yếu tố chính quyết định bỏ rơi giỏ hàng**

- Trải nghiệm người dùng không tốt.
- Chi phí vận chuyển cao, đơn hàng không minh bạch.
- Nhận thức rủi ro từ người dùng trực tuyến.

### **2.2. Thuật toán cây quyết định và rừng ngẫu nhiên**

- Thuật toán cây quyết định.
- Thuật toán rừng ngẫu nhiên.

### **2.3. Ứng dụng học máy trong dự đoán người dùng bỏ rơi giỏ hàng**

- Dữ liệu và bối cảnh thử nghiệm.
- So sánh kết quả thuật toán.

### **2.4. Kết luận**

Từ hành vi bỏ rơi giỏ hàng trong chương một, chương hai của luận văn đã làm rõ hơn các nguyên nhân chính dẫn đến hiện tượng người dùng bỏ rơi giỏ hàng điện

tử. Trong đó, luận văn đề cập đến ba yếu tố chính là trải nghiệm người dùng, chi phí vận chuyển cao, không minh bạch và khả năng nhận thức rủi ro của người dùng. Đồng thời, thuật toán rừng ngẫu nhiên cũng được giới thiệu và đề cập cùng thuật toán cây quyết định là đơn vị căn bản và cốt lõi của thuật toán rừng ngẫu nhiên. Một số thuật toán phổ biến trong phân loại dữ liệu cũng được đề cập và so sánh tốc độ và hiệu năng xử lý tác vụ phân tích dữ liệu. Trên những cơ sở so sánh đó, thuật toán rừng ngẫu nhiên thể hiện được sự hài hòa giữa thời gian xử lý cũng như độ chính xác của phân loại, từ đó cho thấy lý do ứng dụng rừng ngẫu nhiên trong phân loại và dự đoán hành vi bỏ rơi giỏ hàng.

## CHƯƠNG 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ

### 3.1. Phát biểu bài toán

- Bài toán phân tích và dự đoán phân luồng trực tiếp

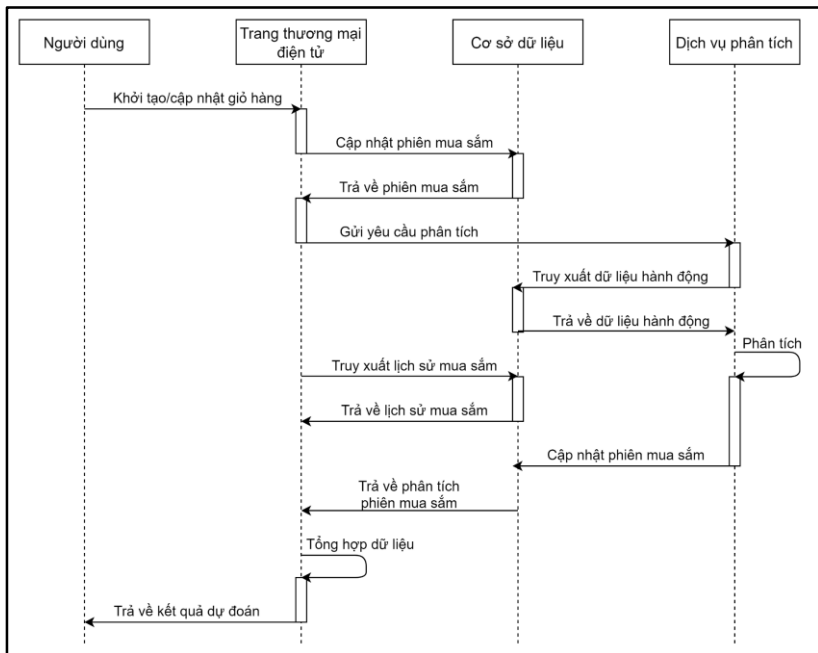
Bài toán phân tích và dự đoán phân luồng trực tiếp nhận đầu vào là các dữ liệu tổng hợp hoạt động của khách hàng trong các phiên mua sắm tại trang thương mại điện tử. Tập dữ liệu chủ yếu tổng hợp những hoạt động xem sản phẩm và sử dụng giỏ hàng của người tiêu dùng. Đây sẽ là tập dữ liệu được ứng dụng thuật toán rừng ngẫu nhiên để phân loại và dự đoán hành vi bỏ rơi giỏ hàng. Kết quả của bài toán là một giá trị dự đoán nhị phân 1 hoặc 0 với giá trị bằng 1 là người dùng sẽ hoàn tất giỏ hàng và 0 là người dùng sẽ bỏ rơi giỏ hàng.

- Bài toán thống kê dữ liệu phân luồng gián tiếp.

Bài toán thống kê dữ liệu phân luồng gián tiếp nhận đầu vào là các dữ liệu được thống kê và tổng hợp từ các lịch sử mua sắm, những hoạt động và tương tác của khách hàng. Từ dữ liệu đã được tổng hợp, bài toán sẽ cho ra kết quả là một tỷ lệ phần trăm cho khả năng thực hiện mua hàng sắm tới của

từng khách hàng hiện tại trong hệ thống. Kết quả dự đoán chủ yếu dựa trên số lượng giỏ hàng hoàn thành và số lượng đơn hàng thành công của khách hàng tại trang thương mại điện tử.

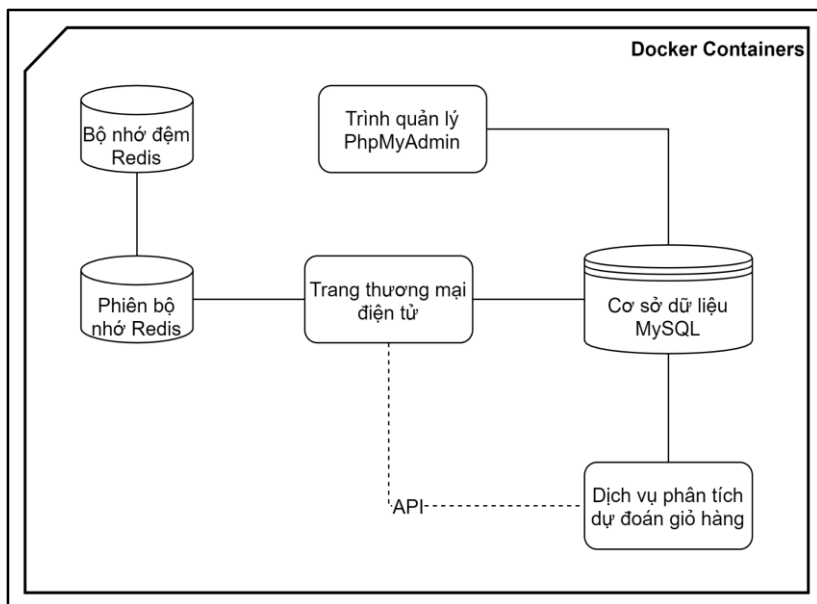
- Tổng hợp bài toán và trình tự phân tích.



*Sơ đồ trình tự phân tích lịch sử và hoạt động trong phiên mua sắm*

### 3.2. Cấu trúc hệ thống và dữ liệu

- Cấu trúc trang thương mại điện tử và dịch vụ phân tích.



*Cấu trúc trang thương mại điện tử và dịch vụ dự đoán bỏ rơi giỏ hàng*

- **Cấu trúc dữ liệu.**

Đối với phân luồng trực tiếp, dữ liệu thử nghiệm trong luận văn sẽ bao gồm tất cả các phiên mua sắm của 188 khách hàng đăng nhập và các khách hàng vắng lai khác được tổng hợp từ một trang thương mại điện tử cung cấp vật liệu xây dựng và thiết bị vệ sinh phòng tắm trong khoảng thời gian từ ngày 01 tháng 12 năm 2020 đến ngày 15 tháng 02 năm 2021 - khoảng 2,5 tháng. Trong thời gian đó, Gần 10.000 hoạt động tiêu dùng

trực tuyến trên trang thương mại điện tử được ghi lại và tổng hợp thành 1.173 phiên mua sắm khác nhau. Dữ liệu trong phần luồng gián tiếp sẽ được thu thập và tổng hợp ngay sau khi dịch vụ dự đoán bỏ rơi giỏ hàng được tích hợp vào trang thương mại điện tử. Trong đó, các thành phần ngoại vi của dịch vụ dự đoán sẽ yêu cầu truy xuất lịch sử mua sắm, các nhận xét và đánh giá của từng tài khoản khách hàng trong trang thương mại điện tử.

### **3.3. Thử nghiệm và đánh giá**

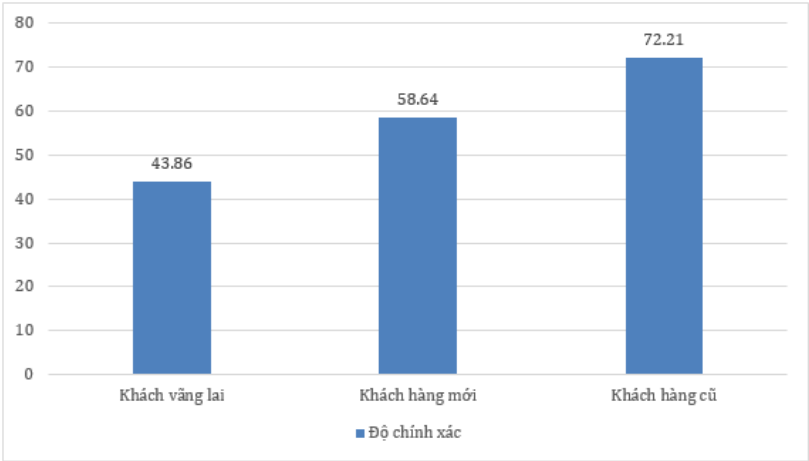
- Thống kê và phân tích khuôn mẫu dữ liệu.

Quá trình huấn luyện đòi hỏi dữ liệu được phân chia thành các nhóm dữ liệu khác nhau nhằm phục vụ mục đích huấn luyện, thử nghiệm và so sánh độ chính xác của dự đoán phân loại. Do vậy, ba đối số là kích thước thử nghiệm (test size), trạng thái ngẫu nhiên (random state) và số lượng ước tính (number of estimators) được cân nhắc để thuật toán không chỉ đưa ra được kết quả với độ chính xác cao mà thời gian thực hiện phải tối ưu nhất. Trong đó, các đối số có vai trò như sau:

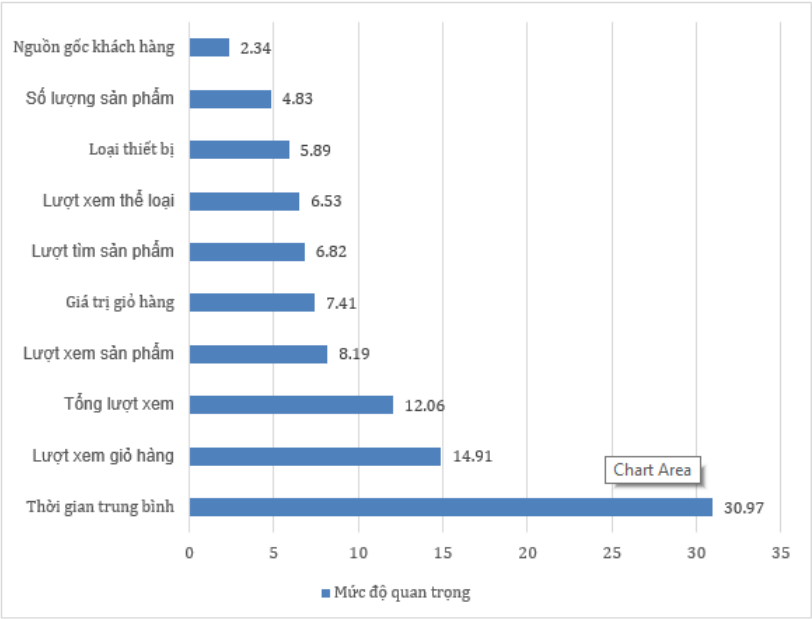
- *Kích thước thử nghiệm*: tỉ lệ phân chia giữa tập dữ liệu huấn luyện và tập dữ liệu thử nghiệm.
- *Trạng thái ngẫu nhiên*: sự nhất quán trong mỗi lần xáo trộn dữ liệu để đào tạo.
- *Số lượng ước tính*: số lượng các cây quyết định được khởi tạo trước khi đưa ra kết quả phân loại cuối cùng.

Từ so sánh kết quả chính xác và thời gian thực hiện tác vụ, có thể kết luận rằng với các đối số có giá trị lần lượt là: kích thước thử nghiệm – 0.1; trạng thái ngẫu nhiên – 10; số lượng ước tính – 10 sẽ cho kết quả tối ưu nhất với độ chính xác khoảng 90% và thời gian thực nghiệm trong khoảng 0,09 giây. Do vậy, ba giá trị trên sẽ được sử dụng trong quá trình ứng dụng thực tế việc phân loại người tiêu dùng bỏ rơi giỏ hàng

- Thử nghiệm thực tế.

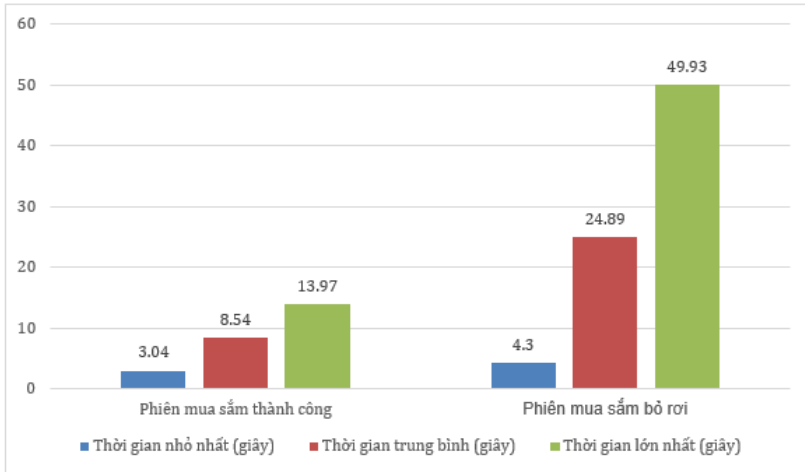


*Kết quả dự đoán trong ứng dụng dự đoán thực tế*



*Mức độ quan trọng của các thuộc tính trong phiên mua sắm*





*So sánh thời gian trung bình của phiên mua sắm thành công và bỏ rơi*

### 3.4. Kết luận

Trong chương ba, luận văn tập trung vào hai bài toán phân tích và thống kê dữ liệu tương ứng với hai phân luồng trực tiếp và gián tiếp trong dự đoán hành vi bỏ rơi giỏ hàng của người tiêu dùng. Dựa trên phát biểu của hai bài toán, chương ba cũng đã làm sáng tỏ trình tự phân tích và tổng hợp kết quả của hai phân luồng để đưa ra kết quả dự đoán cuối cùng. Đồng thời, luận văn đã trình bày cấu trúc hệ thống trang thương mại điện tử và dịch vụ phân tích cùng mô tả chi tiết, kích cỡ và cách thu thập dữ liệu thử nghiệm.

Trong thử nghiệm và đánh giá thực tế, luận văn đã tìm ra được giá trị tối ưu cho các đối số của thuật toán rừng ngẫu nhiên, từ đó đưa ra kết quả thử nghiệm thực tế với tập khách hàng cũ (khách hàng đã có tài khoản và đã từng có giao dịch thành công) có dự đoán bỏ rơi giỏ hàng chính xác nhất là 72,21%. Ngoài ra, thuật toán rừng ngẫu nhiên trong chương ba cũng đưa ra được tiêu chí thời gian trung bình giữa các hành động đóng vai trò quan trọng nhất trong quyết định bỏ rơi giỏ hàng của tiêu dùng. Trong đó, thời gian trung bình của phiên mua sắm thành công là 8,5 giây trong khi tại phiên mua sắm có hiện tượng bỏ rơi giỏ hàng, thời gian trung bình giữa các hành động cao hơn gần gấp 3 lần (24,9 giây).

# KẾT LUẬN CHUNG

## 1. Kết quả đạt được trong luận văn

Sau quá trình tìm hiểu, nghiên cứu và ứng dụng, luận văn “Nghiên cứu phân tích hành vi người dùng bỏ giỏ hàng trên các trang thương mại điện tử” đã cơ bản đáp ứng được các nội dung trình bày trong đề cương. Cụ thể, luận văn đã đạt được một số kết quả chính sau:

- Giới thiệu tổng quan về thương mại điện tử trên thế giới và ở Việt Nam trong những năm gần đây và làm rõ tiềm năng khai thác dữ liệu tiêu dùng và hoạt động mua sắm của khách hàng.
- Trình bày vấn đề bỏ rơi giỏ hàng điện tử trong mua sắm trực tuyến và những yếu tố chính dẫn đến hiện tượng này.
- Trình bày thuật toán học máy có giám sát cây quyết định và rừng ngẫu nhiên; đồng thời, so sánh hiệu năng rừng ngẫu nhiên với các thuật toán phân loại khác để làm sáng tỏ mức độ phù hợp trong tác vụ phân loại dự đoán bỏ rơi giỏ hàng.
- Ứng dụng thuật toán rừng ngẫu nhiên, xây dựng dịch vụ phân tích và tích hợp cùng hệ thống của

trang thương mại điện tử để thử nghiệm phân tích và dự đoán hành vi bỏ rơi giỏ hàng.

- Trên kết quả thu được từ thử nghiệm thực tế, luận văn đưa ra thời gian trung bình giữa các hành động là yếu tố chính trong quyết định bỏ rơi giỏ hàng của người dùng trong mua sắm trực tuyến.

## **2. Hướng nghiên cứu tiếp theo**

Trong tương lai, đề tài nghiên cứu và ứng dụng của luận văn có thể được mở rộng ở nhiều phương diện về thuật toán sử dụng cũng như nền tảng thương mại điện tử được áp dụng. Trong đó, hướng nghiên cứu phát triển tiếp theo có thể được cụ thể hóa như sau:

- Nghiên cứu thêm các thuật toán học máy mới để đa dạng hóa thuật toán sử dụng cũng như đối chiếu kết quả trong tác vụ dự đoán hành vi bỏ rơi giỏ hàng.
- Ứng dụng dịch vụ phân tích và dự đoán bỏ rơi giỏ hàng trên các nền tảng thương mại điện tử khác ví dụ: BigCommerce, Shopify, WooCommerce, v.v