**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**KHOA CÔNG NGHỆ THÔNG TIN**

****



**đỒ ÁN KẾT THÚC HỌC PHẦN**

**Môn: PHÂN TÍCH DỮ LIỆU**

Mã môn học: 841432

Khoa: Công nghệ thông tin

Chuyên ngành: Hệ thống thông tin

Nhóm nghiên cứu:

Trần Nguyên Lộc – 3120410297

Võ Đăng Quang – 3120410429

**Giảng viên phụ trách:**

**PHAN THÀNH HUẤN**

**Thành phố Hồ Chí Minh, tháng 12 năm 2023**

# **Lời cảm ơn**

Trước hết em xin gửi đến lời cảm ơn chân thành và sâu sắc nhất đến thầy TS. Phan Thành Huấn người trực tiếp hướng dẫn và tận tình chỉ bảo cho nhóm chúng em cho tới khi em hoàn thành đồ án của mình.

Tiếp đến em xin giành lời cảm ơn đến quý thầy cô Trường Đại học Sài Gòn – khoa Công nghệ thông tin đã truyền đạt cho em những kiến thức vô cùng quý báu và bổ ích trong suốt quá trình nghiên cứu và học tập tại trường.

Xin chân thành cảm ơn tới những người bạn đã luôn sát cánh cùng em, những lời động viên, những lần hỗ trợ những lúc cần thiết đã phần nào giúp em hoàn thành đồ án này.

Cuối cùng, em xin cảm ơn đến ba mẹ và người thân trong gia đình đã hỗ trợ và tạo điều kiện thuận lợi cho em trong suốt thời gian học tập và nghiên cứu tại Đại học Sài Gòn.

# **TÓM TẮT ĐỒ ÁN**

Đồ án này trình bày quá trình tìm kiếm, khám phá và phân tích thông tin từ tập dữ liệu tìm được. Trong đó có sử dụng các các thuật toán và phương pháp khai thác dữ liệu. Sau khi áp dụng các phương pháp, cần đánh giá hiệu suất của mô hình và lựa chọn mô hình tốt nhất dựa trên các tiêu chí như độ chính xác, độ tin cậy, độ phủ,… Cuối cùng, quá trình khai thác dữ liệu dẫn dến việc trích xuất tri thức từ dữ liệu và đưa ra quyết định thông minh có lợi cho việc marketing.

# **BẢNG ĐÁNH GIÁ MỨC ĐỘ THAM GIA**

|  |  |  |
| --- | --- | --- |
| Danh sách thành viên | MSSV | Mức độ tham gia |
| Trần Nguyên Lộc | 3120410297 | 100% |
| Võ Đăng Quang | 3120410429 | 100% |

# **Mục lục**

[**Lời cảm ơn** i](#_Toc150235405)

[**TÓM TẮT ĐỒ ÁN** ii](#_Toc150235406)

[**BẢNG ĐÁNH GIÁ MỨC ĐỘ THAM GIA** ii](#_Toc150235407)

[**Mục lục** iii](#_Toc150235408)

[**Danh mục hình ảnh** iv](#_Toc150235409)

[**Danh mục bảng biểu** v](#_Toc150235410)

[**Chương 1: KHÁI QUÁT ĐỒ ÁN** 1](#_Toc150235411)

[**1.1.** **Hành vi tiêu dùng của khách hàng** 1](#_Toc150235412)

[**1.2.** **Lí do chọn đề tài** 1](#_Toc150235413)

[**1.3.** **Mục tiêu đề tài** 1](#_Toc150235414)

[**1.4.** **Mô tả dữ liệu và cấu trúc dữ liệu** 2](#_Toc150235415)

[**Chương 2: QUY TRÌNH THỰC HIỆN VÀ KẾT QUẢ** 4](#_Toc150235416)

[**Tài liệu tham khảo** 5](#_Toc150235417)

# **Danh mục hình ảnh**

# **Danh mục bảng biểu**

Bảng 1.4a

Bảng 1.4b

# **Chương 1: KHÁI QUÁT ĐỒ ÁN**

## **Hành vi tiêu dùng của khách hàng**

Ngày nay, khi cuộc sống chúng ta ngày một phát triển thì việc mua bán, trao đổi hàng hoá ngày càng được chú trọng và quan tâm. Khách hàng ngày càng có nhiều sự lựa chọn, nhiều cơ hội để tìm cho mình một loại sản phẩm phù hợp nhất và giá cả lại phải chăng. Sự hiểu biết về hành vi tiêu dùng là chìa khóa cho một chiến lược marketing thành công cả trong nước và quốc tế. Những hoạt động marketing chủ yếu này sẽ hiệu quả hơn khi được đặt trên cơ sở một sự hiểu biết về hành vi tiêu dùng.

## **Lí do chọn đề tài**

Xuất phát từ nhận thức trên, nhận thấy được tầm quan trọng của sự thay đổi hành vi của người tiêu dùng của trong thời điểm hiện tại, nên nhóm chúng em quyết định lựa chọn đề tài: ”Phân tích các mẫu mua hàng trực tuyến để tìm hiểu các luật kết hợp giữa các mặt hàng” để nhận diện ra mặt hàng nào là tiềm năng, phát hiện những yếu tố quan trọng tác động đến hành vi mua sắm từ đó có những chiến lược Marketing phù hợp.

## **Mục tiêu đề tài**

Mục tiêu của đề tài này sẽ phân tích:

* Dữ liệu về lịch sử mua hàng: sử dụng để xác định các mặt hàng thường được mua cùng nhau, các mặt hàng thường được mua trước hoặc sau một mặt hàng khác, và các mặt hàng thường được mua bởi các khách hàng có chung sở thích.
* Dữ liệu về dữ liệu vị trí: sử dụng để xác định các mặt hàng thường được mua ở cùng một địa điểm, các mặt hàng thường được mua bởi các khách hàng ở cùng một khu vực, và các mặt hàng thường được mua vào cùng một thời điểm.
* Dữ liệu về dữ liệu thanh toán: sử dụng để xác định các mặt hàng thường được mua bởi cùng một khách hàng, các mặt hàng thường được mua cùng với một phương thức thanh toán cụ thể, và các mặt hàng thường được mua vào cùng một thời điểm trong năm.

## **Mô tả dữ liệu và cấu trúc dữ liệu**

Dữ liệu được tổng hợp từ các nguồn sau Wholesale & Retail Orders Dataset (kaggle.com):

<https://www.kaggle.com/datasets/gabrielsantello/wholesale-and-retail-orders-dataset?select=orders.csv>

<https://www.kaggle.com/datasets/gabrielsantello/wholesale-and-retail-orders-dataset?select=product-supplier.csv>

Đây là những nguồn được đánh giá là có những bộ dữ liệu đáng tin cậy để phục vụ cho việc nghiên cứu hành vi tiêu dùng.

*Bảng 1.4a. Bảng mô tả cấu trúc của bộ dữ liệu orders.csv*

|  |  |
| --- | --- |
| Thuộc tính | Ý nghĩa |
| Customer ID | Mã khách hàng |
| Customer Status | Trạng thái khách hàng thân thiết (Ví dụ: Gold, Slive, Platinum) |
| Date Order was placed | Ngày đặt hàng |
| Delivery Date | Ngày giao hàng |
| Order ID | Mã đặt hàng |
| Product ID | Mã sản phẩm |
| Quantity Ordered | Số lượng sản phẩm đặt |
| Total Retail Price for This Order | Tổng giá bán lẻ cho đơn hàng này |
| Cost Price Per Unit | Giá vốn của mỗi món hàng |

*Bảng 1.4b. Bảng mô tả cấu trúc của bộ dữ liệu product-supplier.csv*

|  |  |
| --- | --- |
| Thuộc tính | Ý nghĩa |
| Product ID | Mã sản phẩm |
| Product Line | Dòng sản phẩm |
| Product Category | Danh mục sản phẩm |
| Product Group | Nhóm sản phẩm |
| Product Name | Tên sản phẩm |
| Supplier Country | Nước phân phối |
| Supplier Name | Tên nhà phân phối |
| Supplier ID | Mã nhà phân phối |

# **Chương 2: QUY TRÌNH THỰC HIỆN VÀ KẾT QUẢ**

# **Tài liệu tham khảo**