

**BỘ GIÁO DỤC ĐÀO TẠO
TRƯỜNG ĐẠI HỌC ĐẠI NAM**



ĐỒ ÁN TỐT NGHIỆP

**ĐỀ TÀI: PHÁT TRIỂN WEBSITE TỰ ĐỘNG GỢI Ý,
TƯ VẤN KHÁCH HÀNG CHO CỬA HÀNG QUẦN ÁO**

SINH VIÊN : TẠ BÍCH LÂM

MÃ SINH VIÊN : 1451020134

LỚP : CNTT 14-05

NGƯỜI HƯỚNG DẪN: THS. PHẠM THỊ TỔNG

HÀ NỘI - 2024

**BỘ GIÁO DỤC ĐÀO TẠO
TRƯỜNG ĐẠI HỌC ĐẠI NAM**



TẠ BÍCH LÂM

**PHÁT TRIỂN WEBSITE TỰ ĐỘNG GỢI Ý,
TƯ VẤN KHÁCH HÀNG CHO CỬA HÀNG
QUẦN ÁO**

**CHUYÊN NGÀNH : CÔNG NGHỆ THÔNG TIN
MÃ SỐ : 74.80.201**

NGƯỜI HƯỚNG DẪN: THS. PHẠM THỊ TỔNG

HÀ NỘI - 2024

LỜI CAM ĐOAN

Em tên là Tạ Bích Lâm, em xin cam đoan rằng báo cáo đồ án tốt nghiệp về đề tài "Phát triển ứng dụng Website tự động gợi ý, tư vấn khách hàng cho cửa hàng quần áo" là sản phẩm của riêng em trong quá trình thực hiện đồ án tốt nghiệp dưới sự hướng dẫn của cô Phạm Thị Tố Nga.

Em xin cam đoan rằng trong quá trình thực hiện báo cáo này, em đã tuân thủ mọi quy định về nghiên cứu khoa học và trích dẫn tài liệu. Mọi thông tin, số liệu và kết quả được trình bày trong báo cáo đều được thu thập và xử lý một cách trung thực và đáng tin cậy.

Em xin chịu trách nhiệm về tính chính xác và độ tin cậy của báo cáo này. Báo cáo được trình bày như một tài liệu tham khảo dành cho mục đích học tập và nghiên cứu.

Hà Nội, ngày 16 tháng 06 năm 2024

Sinh viên thực hiện

Tạ Bích Lâm

LỜI CẢM ƠN

Trước hết, em xin bày tỏ lòng biết ơn chân thành đến tất cả các cá nhân và tổ chức đã hỗ trợ và giúp đỡ em trong suốt quá trình thực hiện đồ án tốt nghiệp này. Em muốn gửi lời cảm ơn đặc biệt đến các thầy cô giáo, những người đã dành thời gian và công sức để hướng dẫn và hỗ trợ em. Nhờ vào sự chỉ dẫn và động viên của thầy cô, em đã có cơ hội không chỉ làm việc một cách hiệu quả mà còn phát triển kỹ năng và kiến thức của mình.

Đặc biệt, em muốn gửi lời cảm ơn sâu sắc đến cô Phạm Thị Tố Nga, người đã tận tình hướng dẫn và giúp đỡ em trong quá trình thực hiện báo cáo thực tập. Sự hỗ trợ của cô đã giúp em hoàn thành báo cáo một cách thành công và tự tin.

Cuối cùng, em muốn bày tỏ lòng biết ơn đến trường Đại học Đại Nam vì đã tạo điều kiện cho sinh viên như em tiếp cận và áp dụng kiến thức thực tế, từ đó phát triển kỹ năng và chuẩn bị cho sự nghiệp trong tương lai.

Em nhận thức rằng do bản thân còn nhiều hạn chế về chuyên môn và thời gian, có thể sẽ gặp phải một số sai sót trong quá trình thực hiện. Em mong nhận được sự góp ý và phản hồi tích cực từ các thầy cô trong khoa và các bạn để báo cáo của em được hoàn thiện tốt hơn.

Em xin chân thành cảm ơn!

LỜI NÓI ĐẦU

Trong sự phát triển không ngừng của công nghệ và sự cạnh tranh gay gắt trong ngành thương mại điện tử và cửa hàng truyền thống, việc tối ưu hóa trải nghiệm mua sắm trở nên hết sức quan trọng. Để đáp ứng nhu cầu này, việc quyết định phát triển một ứng dụng website tự động gợi ý và tư vấn cho khách hàng cho cửa hàng quần áo. Mục tiêu của dự án là tạo ra một nền tảng thông minh và hiệu quả, giúp khách hàng dễ dàng tìm kiếm và lựa chọn sản phẩm phù hợp với sở thích và nhu cầu cá nhân của họ.

Tạo nên website gợi ý và tư vấn mang lại giá trị cao nhất cho cả khách hàng và doanh nghiệp thông qua việc cung cấp những trải nghiệm mua sắm độc đáo và tiện lợi. Trang sẽ sử dụng các công nghệ tiên tiến như trí tuệ nhân tạo và máy học để phân tích dữ liệu khách hàng, từ đó đưa ra những gợi ý sản phẩm chính xác và cá nhân hóa. Đồng thời, trang web cũng sẽ cập nhật thường xuyên các xu hướng thời trang mới nhất, giúp khách hàng luôn cập nhật với những điều mới mẻ và phong cách.

Qua đó, em hy vọng sẽ không chỉ nâng cao trải nghiệm mua sắm của khách hàng mà còn giúp cửa hàng quần áo tăng cường sự hài lòng và trung thành từ phía khách hàng, đồng thời cũng tạo ra một lợi thế cạnh tranh trong thị trường ngày càng khốc liệt hiện nay.

[illegible]

DANH MỤC VIẾT TẮT

| Từ viết tắt | Lý giải từ viết tắt |
|-------------|---------------------------------|
| AI | Artificial intelligence |
| DI | Dependency Injection |
| DOM | Document Object Model |
| FP | Functional Programming) |
| FRP | Functional Reactive Programming |
| JSX | JavaScript XML |
| KNN | K-Nearest Neighbors |
| OOP | Object-Oriented Programming |
| SVM | Support Vector Machine |

DANH MỤC HÌNH ẢNH

| | |
|--|----|
| Hình 2.1. Sử dụng thuật toán ID3 xem xét sự thành công qua một bộ phim | 12 |
| Hình 2.2. Identify the right hyper-plane (Scenario-1)..... | 14 |
| Hình 2.3. Identify the right hyper-plane (Scenario-2)..... | 15 |
| Hình 2.4. Identify the right hyper-plane (Scenario-3)..... | 16 |
| Hình 2.5. Identify the right hyper-plane (Scenario-4)..... | 16 |
| Hình 2.6. Identify the right hyper-plane (Scenario-5)..... | 17 |
| Hình 3.1. Giao diện trang chủ | 31 |
| Hình 3.2. Giao diện trang gợi ý khách hàng..... | 32 |
| Hình 3.3. Giao diện trang mô tả sản phẩm | 33 |
| Hình 3.4. Giao diện trang admin | 33 |
| Hình 3.5. Giao diện thêm người dùng | 34 |
| Hình 3.6. Giao diện sửa người dùng..... | 34 |
| Hình 3.7. Giao diện xóa..... | 35 |
| Hình 3.8. Giao diện thêm sản phẩm | 35 |
| Hình 3.9. Giao diện sửa sản phẩm..... | 36 |
| Hình 4.1. Mô hình Training và Test | 46 |
| Hình 4.2. Sơ đồ ROC SVM..... | 47 |
| Hình 4.3. Phân phối nhãn style..... | 47 |
| Hình 4.4. Kết quả chạy SVM | 47 |
| Hình 4.5. Sơ đồ ROC KNN..... | 48 |
| Hình 4.6. Hình Scatter Plot of Feature 1 vs Feature 2..... | 48 |
| Hình 4.7. Hình Scatter Plot of Feature 2 vs Feature 3..... | 49 |
| Hình 4.8. Sơ đồ ROC Decision Tree | 46 |
| Hình 4.9. Hình Distribution of Classes in Training Data..... | 50 |
| Hình 4.10. Hình Distribution of Classes in Test Data..... | 51 |

DANH MỤC BẢNG BIỂU

| | |
|--|----|
| Bảng 3.1. Xác định Actor | 29 |
| Bảng 3.2. Thuộc tính và mô tả | 37 |
| Bảng 3.3. Thống kê các trường dữ liệu ban đầu..... | 39 |
| Bảng 3.4. Dữ liệu sau khi làm sạch | 39 |
| Bảng 3.5. Bước chi tiết cho từng thuộc tính..... | 43 |
| Bảng 4.1. So sánh SVM, KNN và Decision Tree | 51 |

MỤC LỤC

| | |
|--|----|
| CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI | 1 |
| 1.1. Giới thiệu về đề tài | 1 |
| 1.2. Lý do chọn đề tài | 1 |
| 1.3. Đối tượng khách hàng | 2 |
| 1.4. Phạm vi nghiên cứu của đề tài | 2 |
| 1.5. Mục đích nghiên cứu..... | 3 |
| 1.6. Phương pháp nghiên cứu..... | 3 |
| CHƯƠNG 2. TỔNG QUAN KIẾN THỨC..... | 5 |
| 2.1. Học máy | 5 |
| 2.1.1. Lịch sử phát triển của học máy | 5 |
| 2.1.2. Phân loại về học máy | 6 |
| 2.2. ReactJS | 7 |
| 2.2.1. Khái niệm..... | 7 |
| 2.2.2. Ưu và nhược điểm | 7 |
| 2.3. API | 8 |
| 2.3.1. Khái niệm..... | 8 |
| 2.3.2. Ưu và nhược điểm | 8 |
| 2.4. NestJS..... | 9 |
| 2.4.1. Khái niệm:..... | 9 |
| 2.4.2. Ưu và nhược điểm | 9 |
| 2.5. Một số mô hình học máy..... | 10 |
| 2.5.1. Decision Trees (Cây quyết định) | 10 |
| 2.5.2. Support Vector Machines (SVM):..... | 13 |
| 2.5.3. K-Nearest Neighbors (KNN):..... | 19 |
| 2.5.4. Random Forests | 21 |
| CHƯƠNG 3. XÂY DỰNG WEBSITE PHÂN LOẠI VÀ GỢI Ý SẢN PHẨM | 24 |
| 3.1. Giới thiệu và mô tả về trang web “Aimée Clothing” | 24 |

| | |
|---|----|
| 3.1.1. Chức năng, ưu điểm và nhược điểm của hệ thống | 26 |
| 3.1.2. Giới thiệu đặc điểm nổi bật hơn của trang web đang xây dựng | 28 |
| 3.1.3. Xác định Actor..... | 28 |
| 3.1.4. Giao diện trang web..... | 31 |
| 3.2. Sử dụng mô hình học máy để xây dựng chức năng tự động gợi ý sản phẩm | 36 |
| 3.2.1. Mô tả chức năng tự động gợi ý sản phẩm..... | 36 |
| 3.2.2. Thống kê các trường dữ liệu trong Dataset Women Clothes | 39 |
| 3.2.3. Bộ dữ liệu sử dụng trong mô hình | 40 |
| 3.2.4. Phân chia dữ liệu huấn luyện..... | 40 |
| 3.2.5. Phân tích bộ dữ liệu | 41 |
| CHƯƠNG 4. TỐI ƯU HÓA MÔ HÌNH KỸ THUẬT HỌC MÁY..... | 46 |
| 4.1. Lựa chọn mô hình học máy | 46 |
| 4.2. Mô hình học máy phù hợp..... | 51 |
| 4.3. Đánh giá và kết quả | 52 |
| KẾT LUẬN | 54 |
| TÀI LIỆU THAM KHẢO | 55 |

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI

1.1. . Giới thiệu về đề tài

Việc phát triển một website tự động gợi ý và tư vấn cho khách hàng khi mua quần áo là một xu hướng không thể phủ nhận trong ngành bán lẻ. Môi trường với sự cạnh tranh giữa các cửa hàng truyền thống và các nền tảng thương mại điện tử đang thúc đẩy sự phát triển của các dịch vụ tư vấn trực tuyến, nhằm tạo ra trải nghiệm cá nhân hóa và thuận tiện cho khách hàng.

Các website tự động gợi ý và tư vấn sử dụng công nghệ học máy để tạo ra các gợi ý sản phẩm cá nhân hóa dựa trên sở thích và nhu cầu cụ thể của từng khách hàng. Thông qua việc phân tích dữ liệu, thông tin cá nhân các mô hình học máy có thể dự đoán được những mặt hàng mà khách hàng có thể quan tâm.

Website tự động gợi ý và tư vấn này không chỉ giúp khách hàng tiết kiệm thời gian trong việc tìm kiếm sản phẩm phù hợp mà còn giúp họ tìm được những món đồ mà họ có thể không nhận biết được thông qua việc duyệt qua các danh mục truyền thống. Việc phát triển một website tự động gợi ý và tư vấn cho khách hàng khi tư vấn quần áo không chỉ đáp ứng nhu cầu của người tiêu dùng trong thời đại số hóa mà còn là một cơ hội lớn cho các doanh nghiệp bán lẻ để tối ưu hóa trải nghiệm tư vấn.

Bên cạnh đó, việc xây dựng một Website tự động gợi ý và tư vấn cũng mở ra những cơ hội mới trong việc tương tác và tạo mối quan hệ giữa cửa hàng và khách hàng. Thông qua việc cung cấp các dịch vụ tư vấn và hỗ trợ cá nhân hóa, cửa hàng có thể tạo ra một trải nghiệm tư vấn trực tuyến tốt hơn và nâng cao sự hài lòng của khách hàng.

1.2. Lý do chọn đề tài

Một trong những lý do chính để lựa chọn đề tài này là vì nhu cầu càng tăng của người tiêu dùng trong sự trải nghiệm mua sắm. Thay vì phải tự mò mẫm trong hàng trăm sản phẩm trên website, khách hàng mong muốn có được các gợi ý cụ thể và phù hợp với sở thích và nhu cầu của họ.

Bên cạnh đó, việc tạo ra một trải nghiệm mua sắm thuận tiện và dễ dàng cũng là một lý do không thể phủ nhận. Thương mại điện tử trong giai đoạn mà sự tiện lợi và tốc độ trở thành yếu tố quyết định trong quá trình mua sắm. Một ứng dụng tư vấn tự động giúp khách

hàng tiết kiệm thời gian và năng lượng trong việc tìm kiếm sản phẩm phù hợp, từ đó tạo ra sự hài lòng và trải nghiệm tích cực cho họ.

Trong lĩnh vực kinh doanh quần áo, việc cung cấp một trải nghiệm tư vấn tốt hơn không chỉ là một yếu tố cạnh tranh mà còn là một cơ hội để tạo ra mối quan hệ chắc chắn hơn với khách hàng. Quần áo không chỉ là các sản phẩm mà còn là biểu hiện của phong cách và cá nhân của mỗi người. Do đó, việc cung cấp các gợi ý và tư vấn cá nhân hóa có thể giúp tạo ra trải nghiệm tư vấn đáng nhớ và độc đáo cho khách hàng.

Cuối cùng, khách hàng ngày càng đòi hỏi sự tiện lợi và cá nhân hóa trong trải nghiệm mua sắm của họ. Việc phát triển một website tự động gợi ý và tư vấn khách hàng cho cửa hàng quần áo sẽ giúp đáp ứng nhu cầu này, tạo ra một môi trường mua sắm trực tuyến thuận tiện và đáng tin cậy.

1.3. Đối tượng khách hàng

Đối tượng khách hàng của ứng dụng này bao gồm cả người tiêu dùng muốn tìm kiếm và tư vấn quần áo trực tuyến.

Đối với người tiêu dùng, ứng dụng này sẽ là một công cụ hữu ích giúp họ dễ dàng tìm kiếm và tư vấn các sản phẩm quần áo một cách thuận tiện và nhanh chóng từ bất kỳ đâu, bất kỳ khi nào. Họ sẽ được tận hưởng trải nghiệm tư vấn trực tuyến cá nhân hóa và được đề xuất những sản phẩm phù hợp với sở thích và nhu cầu của mình, từ đó tạo ra sự hài lòng và tin cậy trong việc tư vấn trực tuyến.

Ngoài ra đối với chủ cửa hàng bán quần áo, ứng dụng này sẽ là một công cụ quản lý và tương tác khách hàng hiệu quả. Họ có thể sử dụng dữ liệu và thông tin từ ứng dụng để hiểu rõ hơn về khách hàng của mình, từ đó cải thiện dịch vụ và sản phẩm để đáp ứng nhu cầu của họ một cách tốt nhất. Đồng thời, ứng dụng cũng sẽ giúp tăng cường tương tác và mối quan hệ với khách hàng, từ đó xây dựng và duy trì một cộng đồng khách hàng trung thành và phát triển lâu dài.

1.4. Phạm vi nghiên cứu của đề tài

Xây dựng hệ thống tự động gợi ý sản phẩm, một trong những mục tiêu hàng đầu của nghiên cứu là phát triển một hệ thống tự động có khả năng gợi ý sản phẩm phù hợp với nhu cầu và sở thích cụ thể của từng khách hàng. Điều này đòi hỏi việc áp dụng các phương

pháp và thuật toán học máy để phân tích dữ liệu khách hàng và hiểu rõ hơn về các yếu tố ảnh hưởng đến quyết định tư vấn của họ. Mục tiêu này giúp tăng cường trải nghiệm tư vấn bằng cách cung cấp cho khách hàng các gợi ý chính xác và hữu ích, từ đó tăng cơ hội cho việc tư vấn thành công và tạo ra một môi trường tư vấn trực tuyến thuận lợi và thú vị hơn.

Mục tiêu tiếp theo của nghiên cứu là tăng cường trải nghiệm tư vấn trực tuyến bằng cách cung cấp các gợi ý cá nhân hóa và tư vấn chính xác. Mục tiêu này giúp tạo ra một môi trường tư vấn trực tuyến tương tác và cá nhân hóa, từ đó thu hút và giữ chân khách hàng một cách hiệu quả.

Cách cung cấp các gợi ý sản phẩm chính xác và cá nhân hóa, hệ thống giúp tăng cơ hội cho việc tư vấn thành công. Đồng thời, trải nghiệm tư vấn trực tuyến tốt cũng giúp tạo ra một cảm giác tin cậy và sự hài lòng đối với khách hàng, từ đó xây dựng một cộng đồng khách hàng trung thành và đem lại lợi ích dài hạn cho doanh nghiệp.

1.5. Mục đích nghiên cứu

Mục đích của nghiên cứu này là phát triển một website tự động gợi ý và tư vấn khách hàng cho cửa hàng quần áo, nhằm nâng cao trải nghiệm mua sắm trực tuyến. Đầu tiên, hệ thống sẽ áp dụng các thuật toán học máy để phân tích dữ liệu khách hàng, từ đó hiểu rõ nhu cầu và sở thích cụ thể của từng cá nhân.

Điều này cho phép cung cấp các gợi ý sản phẩm chính xác và hữu ích, giúp tăng cơ hội tư vấn thành công. Bên cạnh đó, việc cung cấp gợi ý cá nhân hóa và tư vấn chính xác sẽ tạo ra một môi trường tư vấn trực tuyến tương tác và cá nhân hóa, thu hút và giữ chân khách hàng một cách hiệu quả. Mục tiêu cuối cùng là xây dựng một cộng đồng khách hàng trung thành, đem lại lợi ích dài hạn cho doanh nghiệp bằng cách tạo cảm giác tin cậy và sự hài lòng cao đối với khách hàng..

1.6. Phương pháp nghiên cứu

Phương pháp nghiên cứu trang web tư vấn gợi ý quần áo sử dụng học máy sẽ được bao gồm phương pháp phân tích và mô tả. Đầu tiên, về phương pháp phân tích, xác định người dùng khi sử dụng hệ thống tư vấn quần áo, bao gồm các yếu tố như thể loại, kích thước, màu sắc ưa thích, mùa và chất liệu. Sau đó, hệ thống sẽ đưa ra gợi ý quần áo hiện có trong cửa hàng.

Tiếp theo, quá trình tiền xử lý dữ liệu giúp chuẩn bị dữ liệu cho việc huấn luyện mô hình đảm bảo độ chính xác của dữ liệu, bao gồm loại bỏ dữ liệu trùng lặp và đưa dữ liệu về cùng một định dạng để dễ dàng sử dụng trong mô hình học máy, trong bài sẽ chuẩn hóa các giá trị màu sắc thành các mã số.

Sau khi dữ liệu đã được tiền xử lý, sẽ tiến hành xây dựng mô hình học máy. Trong đề tài này em lựa chọn thuật toán Support Vector Machine để xây dựng mô hình. Mô hình là các mô hình phân loại, giúp dự đoán và gợi ý quần áo phù hợp với người dùng dựa trên các đặc điểm cụ thể để chuẩn hóa thông tin sản phẩm, các thông tin về thể loại, màu sắc và chất liệu của các sản phẩm quần áo sẽ được chuẩn hóa, gợi ý sản phẩm để gợi ý sản phẩm dựa trên sở thích và yêu cầu của người dùng. Ví dụ, thông tin về thể loại, màu sắc và chất liệu của các sản phẩm quần áo sẽ được chuẩn hóa.

CHƯƠNG 2. TỔNG QUAN KIẾN THỨC

2.1. Học máy

Học máy là một lĩnh vực của trí tuệ nhân tạo (AI) mà nghiên cứu và xây dựng các phương pháp và thuật toán cho máy tính để học và cải thiện hiệu suất từ dữ liệu mà không cần phải được điều chỉnh một cách rõ ràng.

Trong học máy, máy tính được lập trình để tự động học từ dữ liệu và cải thiện hiệu suất của mình theo thời gian, thay vì phải tuân theo các quy tắc cụ thể được lập trình sẵn. Điều này giúp máy tính có khả năng phát triển khả năng tự học và tự điều chỉnh khi chúng được cung cấp với dữ liệu mới.

Các ứng dụng của học máy rất đa dạng và phổ biến trong nhiều lĩnh vực như nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên, dự đoán, phân loại và nhiều ứng dụng. [1]

2.1.1. Lịch sử phát triển của học máy

Các ứng dụng của học máy rất đa dạng và phổ biến trong nhiều lĩnh vực như nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên, dự đoán, phân loại và nhiều ứng dụng khác.

Lịch sử phát triển của học máy có nguồn gốc từ những nghiên cứu trong lĩnh vực trí tuệ nhân tạo (AI) và khoa học máy tính. Dưới đây là một số điểm quan trọng trong lịch sử phát triển của học máy:

- Thập niên 1950-1960: Những nghiên cứu đầu tiên về học máy đã bắt đầu trong thập kỷ này, với các nhà khoa học như Alan Turing và Marvin Minsky đưa ra các ý tưởng về máy tính có khả năng tự học.
- Thập niên 1970-1980: Các thuật toán học máy cổ điển như cây quyết định và mạng nơ-ron nhân tạo đã được phát triển trong giai đoạn này.
- Thập niên 1990-2000: Sự phát triển mạnh mẽ của Internet và sự gia tăng về khả năng tính toán đã tạo điều kiện cho việc sử dụng các phương pháp học máy trên quy mô lớn.
- Thập niên 2010-2020: Học sâu (deep learning) đã trở thành xu hướng quan trọng trong học máy, với việc áp dụng các mạng nơ-ron sâu để giải quyết nhiều bài toán phức tạp như nhận dạng hình ảnh và xử lý ngôn ngữ tự nhiên.

Trong những năm gần đây, học máy đã trở thành một lĩnh vực nghiên cứu quan trọng và đang phát triển mạnh mẽ, với nhiều ứng dụng thực tế rộng rãi. [2]

2.1.2. Phân loại về học máy

2.1.3.1. Dựa trên nguồn gốc dữ liệu:

- Phân loại có giám sát (Supervised Learning): Mô hình được huấn luyện trên các cặp dữ liệu đầu vào và đầu ra đã được gán nhãn.
- Phân loại không giám sát (Unsupervised Learning): Mô hình được huấn luyện trên dữ liệu không có nhãn. Mục tiêu là khám phá cấu trúc hoặc mẫu ẩn trong dữ liệu. [3]

2.1.3.2. Dựa trên phương pháp tiếp cận:

- Phân loại truyền thống (Traditional Machine Learning): Sử dụng các thuật toán truyền thống như hồi quy tuyến tính, cây quyết định, SVM, và Naive Bayes.
- Học sâu (Deep Learning): Sử dụng mạng nơ-ron sâu để học các mô hình phức tạp với nhiều lớp ẩn. [3]

2.1.3.3. Dựa trên mục tiêu của bài toán:

- Phân loại nhị phân (Binary Classification): Phân loại dữ liệu vào hai lớp.
- Phân loại đa lớp (Multiclass Classification): Phân loại dữ liệu vào ba hoặc nhiều lớp.
- Phân loại đa nhãn (Multilabel Classification): Mỗi mẫu có thể thuộc về một hoặc nhiều lớp. [3]

2.1.3.4. Dựa trên trên đặc điểm của mô hình:

- Phân loại tuyến tính (Linear Classification): Mô hình dựa trên các giả định tuyến tính về dữ liệu.
- Phân loại phi tuyến tính (Non-linear Classification): Mô hình không giới hạn bởi các giả định tuyến tính. [3]

2.1.3.5. Dựa trên việc sử dụng cụ thể của mô hình:

- Cây quyết định (Decision Trees): Sử dụng cây quyết định để tạo ra quy tắc phân loại.

- Mạng nơ-ron nhân tạo (Artificial Neural Networks): Sử dụng mạng nơ-ron để học các mô hình phân loại phức tạp. [3]

2.2. ReactJS

2.2.1. Khái niệm

React.js là một thư viện JavaScript mã nguồn mở được phát triển bởi Facebook để xây dựng giao diện người dùng. Nó được sử dụng rộng rãi để phát triển các ứng dụng web động, đặc biệt là các ứng dụng đơn trang (Single Page Applications - SPAs), nơi mà trạng thái của trang web thay đổi mà không cần phải tải lại trang.

Trong React.js, mọi thành phần của giao diện người dùng được xem như các "component" (thành phần), mỗi component đều có trạng thái riêng và có thể tái sử dụng một cách dễ dàng. React sử dụng JSX để viết mã, cho phép lập trình viên kết hợp JavaScript với HTML một cách linh hoạt, giúp tạo ra mã nguồn dễ đọc và dễ hiểu hơn.

Khả năng tái sử dụng component và quản lý trạng thái hiệu quả là những đặc điểm chính của React, giúp tạo ra các ứng dụng web mạnh mẽ, linh hoạt và dễ bảo trì. Đặc biệt, cộng đồng lớn và sự hỗ trợ mạnh mẽ từ Facebook và cộng đồng làm cho React trở thành một lựa chọn phổ biến cho việc phát triển ứng dụng web hiện đại. [4]

2.2.2. Ưu và nhược điểm

Ưu điểm

- Dễ học và ứng dụng: ReactJS được đánh giá dễ học và ứng dụng. Bất kỳ lập trình viên nào hiểu về nền tảng JavaScript đều có thể dễ dàng bắt đầu tạo ứng dụng web bằng ReactJS.
- Dễ dàng hơn trong tạo ứng dụng web động: Rất khó để lập trình viên có thể tạo một ứng dụng web động bằng HTML do đòi hỏi về mã hóa phức tạp. Tuy nhiên với ReactJS đã giúp giải quyết vấn đề này và giúp nó dễ dàng hơn nhờ cung cấp ít mã hóa hơn và nhiều chức năng hơn.
- Dễ dàng tái sử dụng các thành phần: Một ứng dụng web ReactJS được tạo từ nhiều thành phần và những thành phần này có thể tái sử dụng giúp người dùng phát triển và bảo trì ứng dụng dễ dàng hơn.

- Nâng cao hiệu suất: Nhờ DOM ảo giúp ReactJS cải thiện hiệu suất đáng kể. Cụ thể, React Virtual DOM tồn tại trong bộ nhớ và đại diện cho DOM của trình duyệt. Do đó, khi lập trình viết một thành phần React sẽ không ghi trực tiếp vào DOM mà các thành phần ảo phản ứng sẽ chuyển thành DOM giúp hiệu suất mượt mà hơn. [4]

Nhược điểm

- Học phức tạp ban đầu: JSX và kiến trúc dựa trên component có thể làm cho việc học React ban đầu khá phức tạp đối với những người mới bắt đầu.
- Quản lý trạng thái nếu không được điều chỉnh: Mặc dù React có các thư viện như Redux để quản lý trạng thái ứng dụng, nhưng nếu không được sử dụng và điều chỉnh đúng cách, việc quản lý trạng thái có thể trở nên phức tạp và dẫn đến các vấn đề về hiệu suất.
- Cần có kỹ năng JavaScript: React là một thư viện JavaScript, vì vậy để sử dụng nó hiệu quả, lập trình viên cần có kiến thức vững về JavaScript. [4]

2.3. API

2.3.1. Khái niệm

Là một tập hợp các quy tắc và giao diện mà các phần mềm ứng dụng có thể sử dụng để giao tiếp với nhau. Nó định nghĩa cách thức và phương pháp mà các thành phần phần mềm có thể tương tác với nhau, cho phép chúng trao đổi dữ liệu và chức năng một cách hiệu quả. [5]

2.3.2. Ưu và nhược điểm

Ưu điểm

- Tích hợp dễ dàng: API cho phép tích hợp các dịch vụ và chức năng từ các nguồn bên ngoài vào ứng dụng của bạn một cách dễ dàng. Điều này giúp mở rộng tính năng của ứng dụng mà không cần phải xây dựng lại từ đầu.
- Tiết kiệm thời gian và chi phí: Thay vì phải phát triển mọi chức năng từ đầu, sử dụng API giúp tiết kiệm thời gian và chi phí phát triển. Bạn có thể tận dụng các dịch vụ có sẵn từ các nhà cung cấp API khác nhau để nhanh chóng xây dựng các tính năng mới cho ứng dụng của mình.

- Dữ liệu chính xác và cập nhật: API thường cung cấp dữ liệu từ nguồn đáng tin cậy và được cập nhật định kỳ. Điều này giúp đảm bảo rằng thông tin và dữ liệu được gợi ý cho khách hàng là chính xác và phản ánh đúng xu hướng thị trường. [5]

Nhược điểm của API:

- Phụ thuộc vào bên thứ ba: Sử dụng API có nghĩa là phải phụ thuộc vào các nhà cung cấp API khác nhau. Nếu họ thay đổi hoặc ngưng hoạt động dịch vụ, điều này có thể ảnh hưởng đến hoạt động của ứng dụng.
- Rủi ro bảo mật: Khi tích hợp API từ bên thứ ba, có thể tồn tại rủi ro bảo mật. Nếu không kiểm soát được quyền truy cập và sử dụng dữ liệu từ API, thông tin khách hàng có thể bị đánh cắp hoặc sử dụng một cách không đúng đắn.
- Hạn chế về tùy chỉnh: Mặc dù API cung cấp các tính năng sẵn có từ nhà cung cấp, nhưng có thể gặp khó khăn trong việc tùy chỉnh hoặc mở rộng các chức năng này theo nhu cầu cụ thể của ứng dụng của bạn. [5]

2.4. NestJS

2.4.1. Khái niệm

NestJS là một framework Node.js mã nguồn mở được xây dựng trên TypeScript, cung cấp một cách cấu trúc ứng dụng có tổ chức và dễ bảo trì. Nó kết hợp các tính năng của OOP, FP (Functional Programming) và FRP (Functional Reactive Programming), tạo ra một môi trường phát triển ứng dụng hiệu quả.

Mục tiêu của NestJS là giúp phát triển ứng dụng server-side hiện đại và mạnh mẽ, với kiến trúc được cấu trúc rõ ràng và dễ dàng mở rộng. NestJS cung cấp các tính năng như Dependency Injection, Middleware, Exception Handling, và một cách tiếp cận modularity, giúp phát triển ứng dụng dễ dàng và linh hoạt.

Với NestJS, có thể xây dựng các ứng dụng web, API, và microservices một cách hiệu quả và linh hoạt, đặc biệt là khi kết hợp với các công nghệ như TypeScript và Express.js. [6]

2.4.2. Ưu và nhược điểm

Ưu điểm

- Kiến trúc gọn nhẹ và dễ hiểu: NestJS cung cấp một cấu trúc kiến trúc rõ ràng và tổ chức mô-đun, giúp dễ dàng hiểu và quản lý codebase.
- Tích hợp TypeScript: NestJS được viết bằng TypeScript, cho phép sử dụng các tính năng của TypeScript như kiểu dữ liệu tĩnh và mã TypeScript được biên dịch thành JavaScript chạy được trên Node.js.
- Hỗ trợ Dependency Injection : NestJS hỗ trợ Dependency Injection, giúp quản lý và tổ chức các thành phần của ứng dụng một cách hiệu quả, giảm sự phụ thuộc giữa các phần của ứng dụng.
- Tích hợp tốt với các thư viện và công nghệ phổ biến: NestJS tích hợp tốt với Express.js, Fastify và nhiều thư viện khác, cho phép bạn sử dụng các tính năng mạnh mẽ của các thư viện này trong ứng dụng của mình.
- Cộng đồng phát triển mạnh mẽ: NestJS có một cộng đồng phát triển đông đảo và tích cực, với nhiều tài liệu, ví dụ và plugin được cung cấp, giúp việc phát triển ứng dụng trở nên thuận lợi hơn. [6]

Nhược điểm của NestJS:

- Khả năng tương thích với các thư viện và công nghệ cũ hơn: Mặc dù NestJS có thể tích hợp tốt với nhiều thư viện và công nghệ hiện đại, nhưng có thể gặp khó khăn trong việc tích hợp với các thư viện và công nghệ cũ hơn hoặc không còn được duy trì.
- Khó khăn trong việc tìm kiếm giải pháp cho các vấn đề cụ thể: Do NestJS là một framework mới, có thể gặp khó khăn trong việc tìm kiếm giải pháp cho các vấn đề cụ thể hoặc không phổ biến. [6]

2.5. Một số mô hình học máy

2.5.1. *Decision Trees (Cây quyết định)*

2.5.1.1. *Khái niệm Decision Trees*

Decision tree hay còn gọi là Cây quyết định là một dạng sơ đồ phân cấp dựa theo quy luật của mối quan hệ nguyên nhân – kết quả, cho phép người sử dụng dự

báo được kết quả của các lựa chọn. Nhờ đó, Decision tree có ý nghĩa như một công cụ lập kế hoạch giúp điều hướng đến mục tiêu mong muốn. [7]

2.5.1.2. Cách ứng dụng và hoạt động của Decision Trees

Ứng dụng Decision Trees trong các lĩnh vực cụ thể như:

- Tài chính

Phân tích tín dụng: Decision Trees có thể được sử dụng để đánh giá rủi ro tín dụng của khách hàng bằng cách phân loại họ thành các nhóm rủi ro cao, trung bình, và thấp dựa trên các yếu tố như lịch sử tín dụng, thu nhập, và các khoản nợ hiện tại.

- Y tế

Chẩn đoán bệnh: Decision Trees có thể hỗ trợ các bác sĩ trong việc chẩn đoán bệnh bằng cách phân tích các triệu chứng và kết quả xét nghiệm để xác định khả năng mắc bệnh.

Dự đoán kết quả điều trị: Hỗ trợ dự đoán kết quả của các phương pháp điều trị khác nhau dựa trên dữ liệu bệnh nhân, giúp cá nhân hóa kế hoạch điều trị. [7]

- Marketing

Phân khúc khách hàng: Decision Trees giúp phân loại khách hàng thành các nhóm dựa trên hành vi mua sắm, sở thích, và các đặc điểm nhân khẩu học, từ đó tạo ra các chiến dịch marketing mục tiêu hiệu quả hơn.

Dự đoán hành vi mua sắm: Phân tích các dữ liệu lịch sử để dự đoán xu hướng mua sắm tương lai của khách hàng, giúp tối ưu hóa quản lý kho và chiến lược bán hàng. [7]

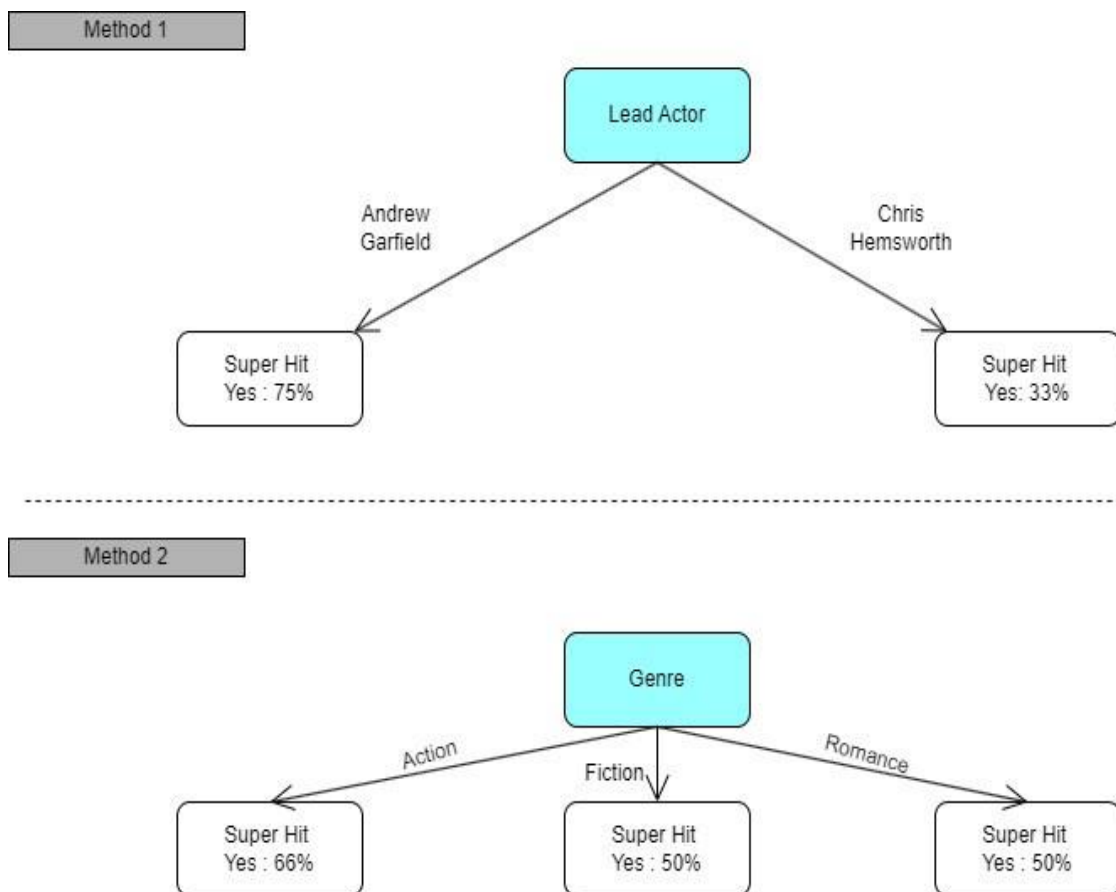
Cách hoạt động của Decision Trees

Bảng 2.1: Sử dụng thuật toán ID3 xem xét sự thành công qua một bộ phim

| Diễn viên chính | Thể loại phim | Hit(Y/N) |
|-----------------|---------------|----------|
| Andrew Garfield | Hành động | Y |
| Andrew Garfield | Viễn tưởng | Y |

| Diễn viên chính | Thể loại phim | Hit(Y/N) |
|-----------------|---------------|----------|
| Andrew Garfield | Lãng mạn | N |
| Andrew Garfield | Chiến tranh | Y |
| Chris Hemsworth | Hành động | Y |
| Chris Hemsworth | Viễn tưởng | Y |
| Chris Hemsworth | Lãng mạn | N |

Xác định độ thành công của bộ phim chỉ trên 1 yếu tố, sẽ có hai cách thực hiện sau: qua diễn viên chính của phim và qua thể loại phim. [8]



Hình 2.1. Sử dụng thuật toán ID3 xem xét sự thành công qua một bộ phim [8]

2.5.1.3. Ưu điểm và nhược điểm của Decision Trees

Ưu điểm của Decision Trees

- Dễ hiểu và diễn giải: Cây quyết định tạo ra các quy tắc quyết định rõ ràng và dễ hiểu, giúp diễn giải kết quả dễ dàng.
- Xử lý dữ liệu không cân bằng: Cây quyết định có thể xử lý tốt dữ liệu không cân bằng bằng cách cung cấp trọng số khác nhau cho các lớp hoặc sử dụng các kỹ thuật như oversampling hoặc undersampling.
- Hỗ trợ đa dạng đặc trưng: Cây quyết định có thể xử lý cả dữ liệu dạng số và dữ liệu phân loại một cách tự nhiên. [7]

Nhược điểm của Decision Trees

- Dễ bị quá mức: Cây quyết định có thể dễ dàng bị quá mức (overfitting) đặc biệt khi cây quá sâu hoặc khi có quá nhiều đặc trưng không quan trọng.
- Không ổn định: Các thay đổi nhỏ trong dữ liệu có thể dẫn đến các cây quyết định khác nhau, điều này làm cho chúng không ổn định so với các phương pháp khác.
- Khó xử lý dữ liệu liên tục: Cây quyết định không xử lý tốt các dữ liệu liên tục và có thể dẫn đến các cây có nhiều lá, tăng nguy cơ overfitting. [7]

2.5.2. Support Vector Machines (SVM)

2.5.2.1. Khái niệm Support Vector Machines

SVM (Support Vector Machine) là một thuật toán học máy phổ biến được sử dụng cho các bài toán phân loại và hồi quy. Ý tưởng cơ bản của SVM là tìm một ranh giới phẳng (hyperplane) trong không gian đặc trưng sao cho nó tối ưu hóa khoảng cách giữa các điểm dữ liệu của các lớp (trong trường hợp phân loại) hoặc giữa các điểm dữ liệu và đường hồi quy (trong trường hợp hồi quy). [9]

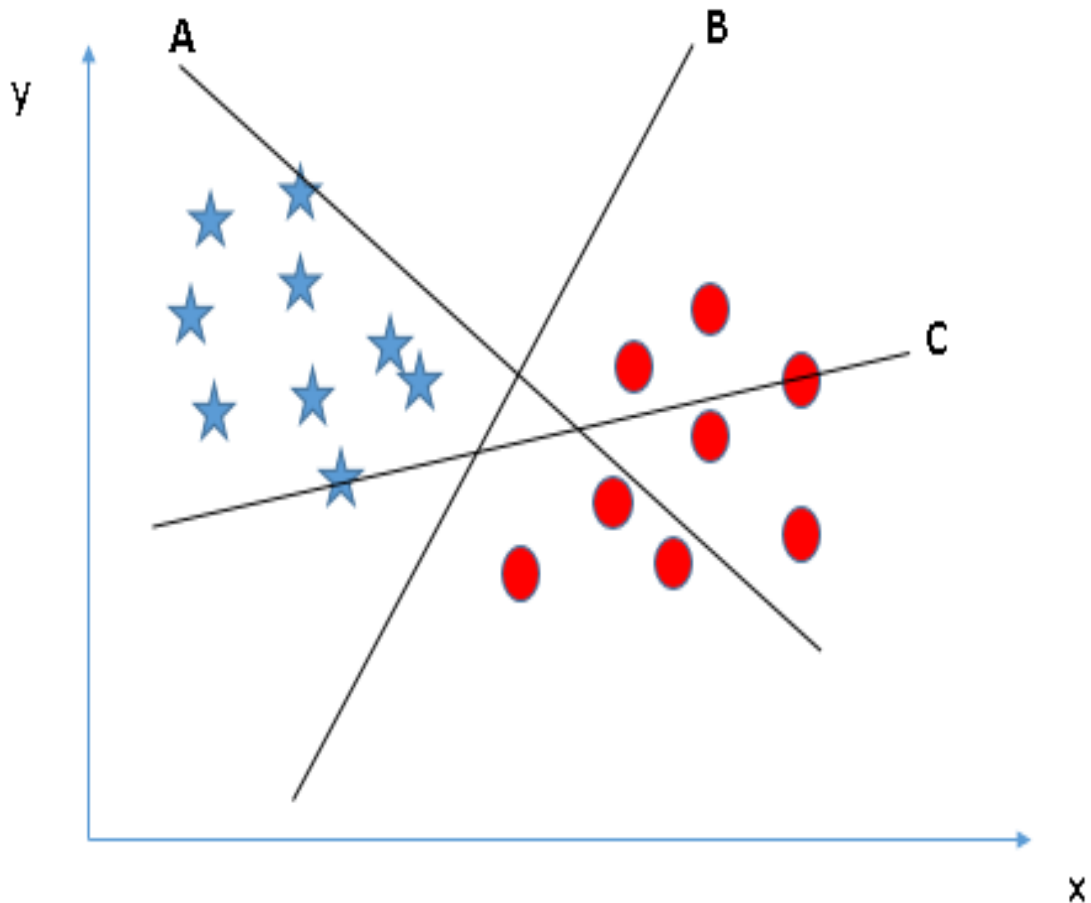
2.5.2.2. Cách ứng dụng và hoạt động của Support Vector Machines

Cách ứng dụng trong các lĩnh vực Support Vector Machines

- Đào tạo là việc xác định sự liên quan thật sự khách quan giữa các môn học trong chương trình đào tạo. Bên cạnh đó, sinh viên cũng cần sự tư vấn để lựa chọn các môn học nhằm đạt được kết quả học tập cao nhất. [9]

Cách hoạt động Support Vector Machines

- Identify the right hyper-plane (Scenario-1):

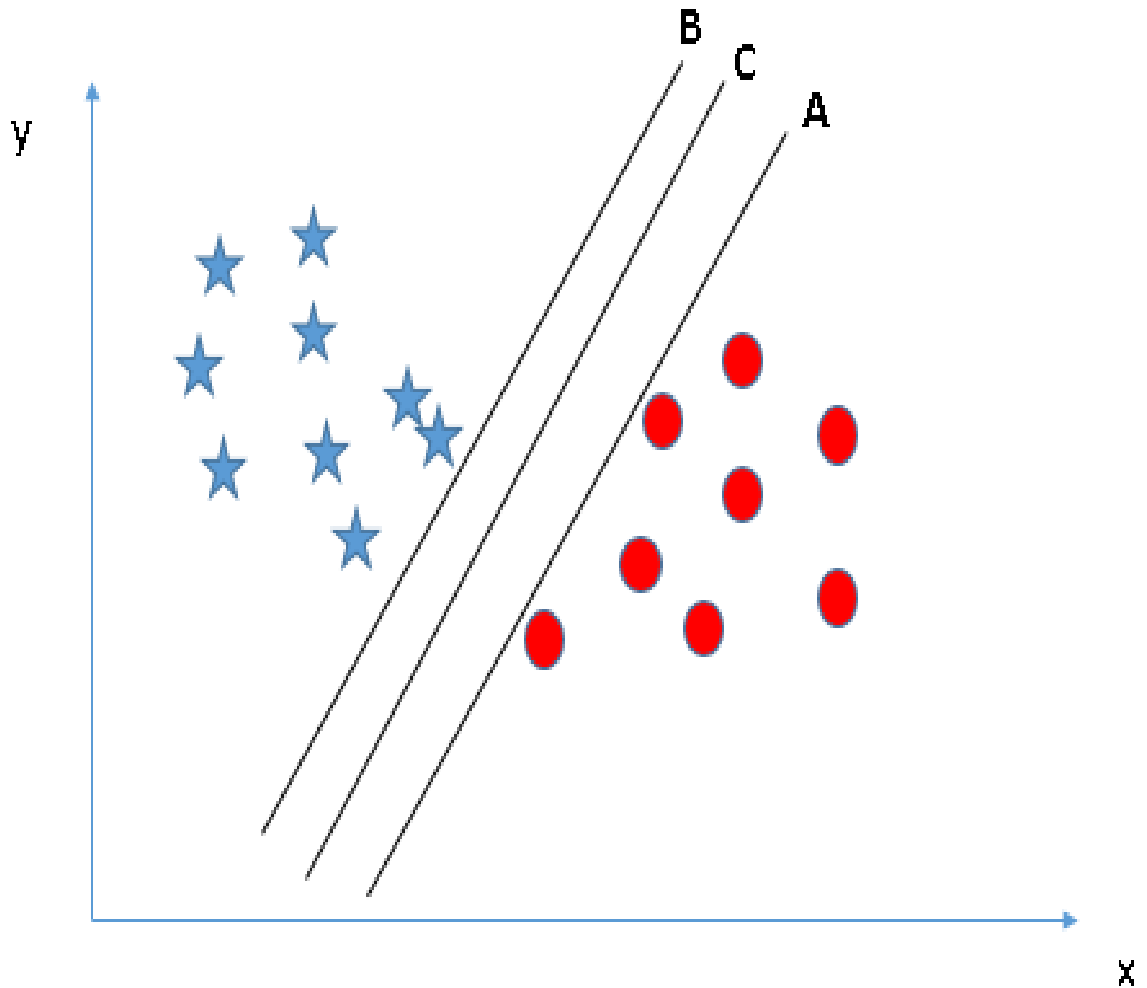


Hình 2.2. Identify the right hyper-plane (Scenario-1) [9]

Ở đây, có 3 đường hyper-lane (A,B and C). Bây giờ đường nào là hyper-lane đúng cho nhóm ngôi sao và hình tròn.

Quy tắc số một để chọn 1 hyper-lane, chọn một hyper-plane để phân chia hai lớp tốt nhất. Trong ví dụ này chính là đường B.

- Identify the right hyper-plane (Scenario-2):

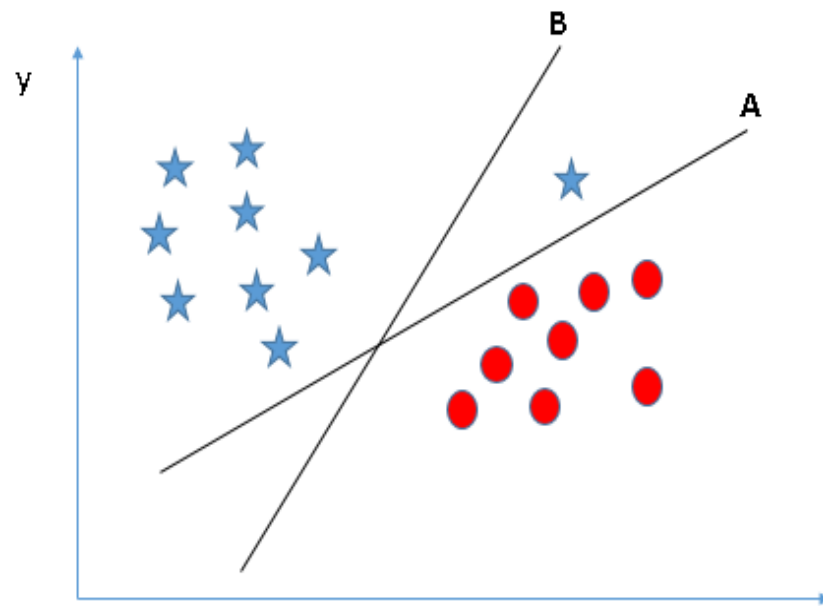


Hình 2.3. Identify the right hyper-plane (Scenario-2) [9]

Ở đây chúng ta cũng có 3 đường hyper-plane (A,B và C), theo quy tắc số 1, chúng đều thỏa mãn.

Quy tắc thứ hai chính là xác định khoảng cách lớn nhất từ điều gần nhất của một lớp nào đó đến đường hyper-plane. Khoảng cách này được gọi là "Margin", Hãy nhìn hình bên dưới, trong đây có thể nhìn thấy khoảng cách margin lớn nhất đây là đường C. Cần nhớ nếu chọn làm hyper-lane có margin thấp hơn thì sau này khi dữ liệu tăng lên thì sẽ sinh ra nguy cơ cao về việc xác định nhầm lớp cho dữ liệu.

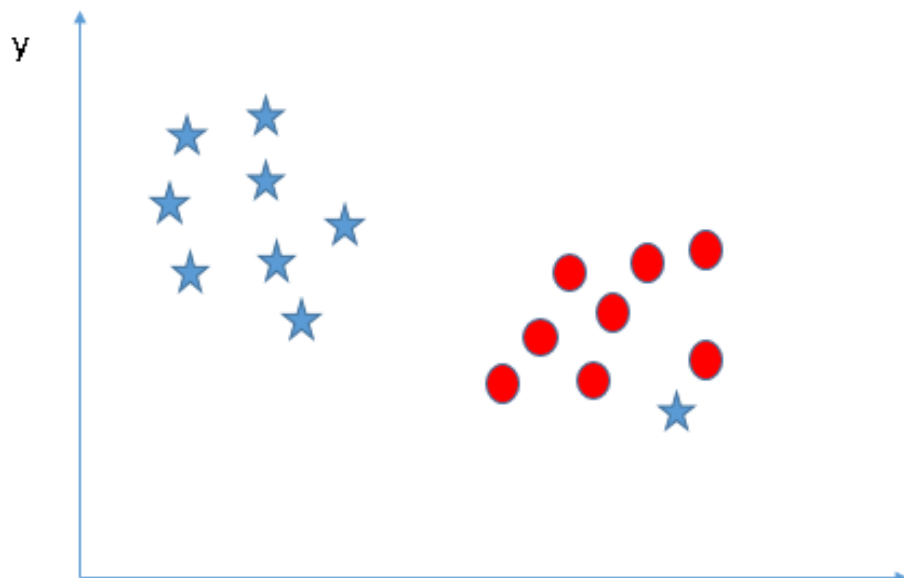
- Identify the right hyper-plane (Scenario-3):



Hình 2.4. Identify the right hyper-plane (Scenario-3) [9]

Sử dụng các nguyên tắc đã nêu trên để chọn ra hyper-plane, có thể có một vài người sẽ chọn đường B bởi vì nó có margin cao hơn đường A, nhưng đây sẽ không đúng bởi vì nguyên tắc đầu tiên sẽ là nguyên tắc số 1, chúng ta cần chọn hyper-plane để phân chia các lớp thành riêng biệt. Vì vậy đường A mới là lựa chọn chính xác.

- Can we classify two classes (Scenario-4)?

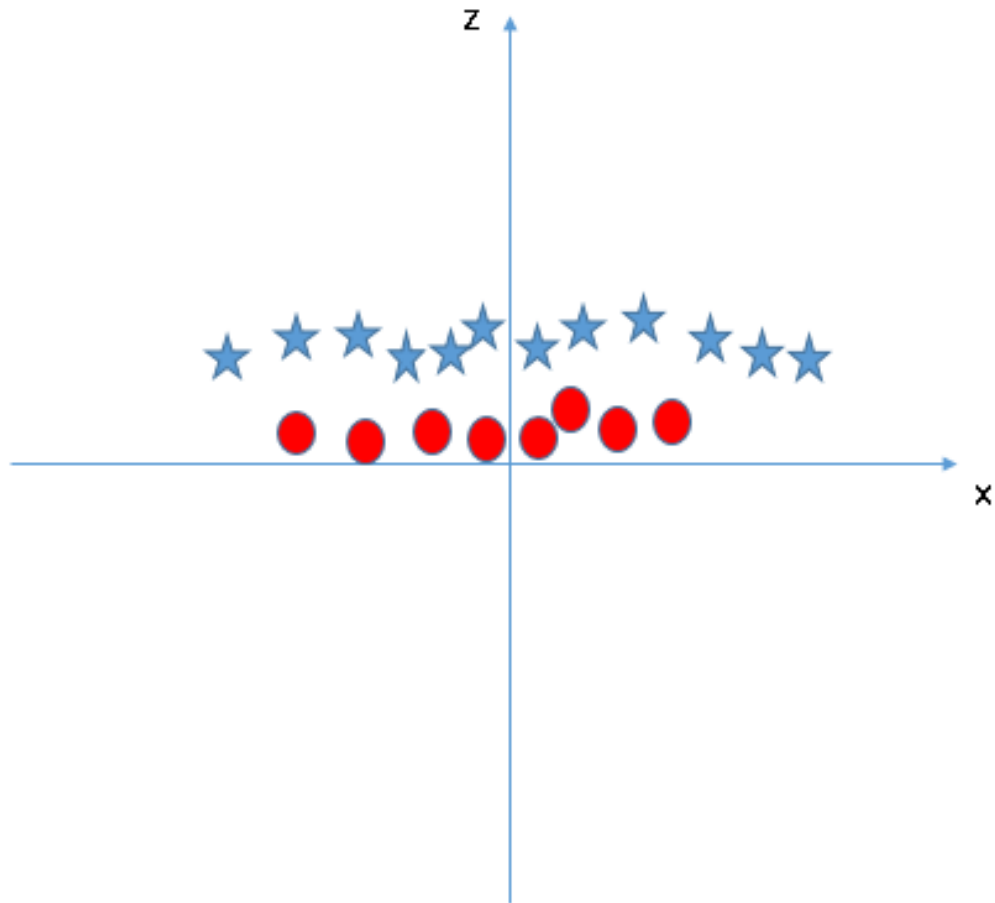


Hình 2.5. Identify the right hyper-plane (Scenario-4) [9]

Ở đây sẽ chấp nhận, một ngôi sao ở bên ngoài cuối được xem như một ngôi sao phía ngoài hơn, SVM có tính năng cho phép bỏ qua các ngoại lệ và tìm ra hyper-plane có biên giới tối đa . Do đó có thể nói, SVM có khả năng mạnh trong việc chấp nhận ngoại lệ.

-
- A scatter plot on a Cartesian coordinate system with x and y axes. The plot displays two classes of data points: red circles and blue stars. The red circles are clustered near the origin, while the blue stars are clustered further out along the axes.

SVM có thể giải quyết vấn đề này, Khá đơn giản, nó sẽ được giải quyết bằng việc thêm một tính năng, Ở đây chúng ta sẽ thêm tính năng $z = x^2 + y^2$. Bây giờ dữ liệu sẽ được biến đổi theo trục x và z như sau



Trong sơ đồ trên, các điểm cần xem xét là:

- Tất cả dữ liệu trên trục z sẽ là số dương vì nó là tổng bình phương x và y
- Trên biểu đồ các điểm tròn đỏ xuất hiện gần trục x và y hơn vì thế z sẽ nhỏ hơn \Rightarrow nằm gần trục x hơn trong đồ thị (z, x)

Trong SVM, rất dễ dàng để có một siêu phẳng tuyến tính (linear hyper-plane) để chia thành hai lớp. SVM có một kỹ thuật được gọi là kernel trick (kỹ thuật hạt nhân), đây là tính năng có không gian đầu vào có chiều sâu thẳm và biến đổi nó thành không gian có chiều cao hơn, tức là nó không phân chia các vấn đề thành các vấn đề riêng biệt, các tính năng này được gọi là kernel. Nói một cách đơn giản nó thực hiện một số biến đổi dữ liệu phức tạp, sau đó tìm ra quá trình tách dữ liệu dựa trên các nhãn hoặc đầu ra mà chúng ta đã xác định trước. [9]

2.5.2.3. Ưu điểm và nhược điểm của Support Vector Machines

Ưu điểm của Support Vector Machines

- Hiệu suất tốt trong không gian cao chiều: SVM hiệu quả trong việc xử lý dữ liệu có nhiều chiều, bởi vì nó chỉ phụ thuộc vào một số lượng nhỏ các điểm dữ liệu gần ranh giới quyết định (support vectors).
- Hiệu quả với dữ liệu lớn: SVM hiệu quả với các tập dữ liệu lớn vì việc tìm kiếm hyperplane chỉ phụ thuộc vào một số lượng hạn chế các điểm dữ liệu.
- Khả năng xử lý dữ liệu phi tuyến tính: Bằng cách sử dụng các hàm kernel, SVM có thể tạo ra các đường ranh giới phi tuyến tính, cho phép phân loại các dữ liệu không phân cách tuyến tính.
- Ít ảnh hưởng bởi nhiễu: SVM chọn ranh giới quyết định dựa trên các support vectors, các điểm dữ liệu gần biên quyết định. Do đó, nó ít bị ảnh hưởng bởi các điểm dữ liệu nhiễu. [9]

Nhược điểm của Support Vector Machines

- Đòi hỏi lượng tính toán lớn: Quá trình huấn luyện của SVM có thể tốn kém về mặt tính toán đối với các tập dữ liệu lớn, đặc biệt là khi sử dụng kernel phi tuyến tính.
- Độ phức tạp của siêu tham số: SVM có các siêu tham số như C và gamma cần được tinh chỉnh một cách chính xác để đạt được hiệu suất tốt nhất. Điều này có thể đòi hỏi nhiều thử nghiệm và đánh giá.
- Khó hiểu và khó diễn giải: Ranh giới quyết định tạo ra bởi SVM có thể là một hàm phi tuyến tính phức tạp, điều này khiến cho việc hiểu và diễn giải kết quả trở nên khó khăn đối với người dùng. [9]

2.5.3. *K-Nearest Neighbors (KNN)*

2.5.3.1. *Khái niệm K-Nearest Neighbors*

Thuật toán K láng giềng gần nhất trong tiếng Anh là K-Nearest Neighbor, viết tắt là KNN. K-Nearest Neighbors (K-NN) là một thuật toán học máy có kiến trúc đơn giản và thường được sử dụng cho các nhiệm vụ phân loại và dự đoán. Thuật toán này hoạt động dựa trên nguyên tắc rằng các điểm dữ liệu có thuộc tính giống nhau sẽ có xu hướng tập trung gần nhau trong không gian đa chiều. K-NN xác định lớp hoặc giá trị dự đoán của một

điểm dữ liệu mới dựa trên lớp hoặc giá trị của các điểm láng giềng gần nhất của nó trong không gian đặc trưng. [10]

2.5.3.2. Cách ứng dụng và hoạt động của *K-Nearest Neighbors*

Cách ứng dụng trong các lĩnh vực của K-Nearest Neighbors:

Thuật toán KNN có nhiều ứng dụng trong ngành đầu tư, bao gồm dự đoán phá sản, dự đoán giá cổ phiếu, phân bổ xếp hạng tín dụng trái phiếu doanh nghiệp, tạo ra chỉ số vốn và trái phiếu tùy chỉnh. [10]

Cách hoạt động

- Chuẩn bị dữ liệu: Mỗi điểm dữ liệu được biểu diễn như một vector đặc trưng trong không gian nhiều chiều, có thể là các thuộc tính, đặc trưng của dữ liệu.
- Tính khoảng cách: Sử dụng các phương pháp đo khoảng cách ví dụ: Euclidean, Manhattan, Cosine để tìm khoảng cách của các điểm dữ liệu.
- Chọn giá trị K: Chọn giá trị K là số lượng láng giềng gần nhất mà thuật toán sử dụng để đưa ra quyết định.
- Xác định láng giềng: Chọn K điểm dữ liệu gần nhất với điểm dựa trên khoảng cách đã tính.
- Xác định phân lớp hoặc giá trị:

Đối với bài toán phân loại: Đếm số lượng điểm thuộc mỗi lớp tập láng giềng, lớp của điểm mới được quyết định bằng cách chọn lớp có số lượng lớn nhất trong tập láng giềng.

Đối với bài toán dự đoán: Tính giá trị trung bình (hoặc trung vị) của các giá trị trong tập láng giềng. [10]

2.5.3.3. Ưu điểm của *K-Nearest Neighbors*

Ưu điểm của K-Nearest Neighbors

- Dễ hiểu và triển khai: K-NN là một thuật toán đơn giản và dễ hiểu. Không có quá nhiều giả định phức tạp, và nó có thể được triển khai một cách nhanh chóng.

- Không yêu cầu huấn luyện: K-NN là một thuật toán "lazy learning", có nghĩa là nó không yêu cầu quá trình huấn luyện. Thay vào đó, nó thực hiện dự đoán dựa trên dữ liệu thực tế khi cần thiết.
- Linh hoạt với dữ liệu không đồng nhất: K-NN có khả năng làm việc tốt với dữ liệu không đồng nhất và không có giả định về phân phối của dữ liệu.
- Hiệu suất tốt với số lượng đặc trưng nhỏ: Trong trường hợp dữ liệu có số lượng đặc trưng nhỏ, K-NN có thể là lựa chọn hiệu quả vì nó không bị ảnh hưởng bởi việc có quá nhiều đặc trưng.
- Độ chính xác tốt với dữ liệu lớn: Khi có một lượng lớn dữ liệu, đặc biệt là khi có sự đa dạng và phức tạp, K-NN có thể cung cấp kết quả dự đoán chính xác. [10]

2.5.3.3. *Nhược điểm của K-Nearest Neighbors*

Nhược điểm của K-Nearest Neighbors

- Nhược điểm với dữ liệu lớn: K-NN có thể trở nên chậm khi xử lý dữ liệu lớn vì phải tính toán khoảng cách giữa điểm dữ liệu và tất cả các điểm dữ liệu khác. Nhạy cảm với nhiễu và đặc trưng không quan trọng: K-NN có thể bị ảnh hưởng bởi nhiễu trong dữ liệu và đặc trưng không quan trọng. Các giá trị ngoại lai có thể ảnh hưởng đáng kể đến kết quả.
- Cần xác định giá trị K: Lựa chọn giá trị K có thể ảnh hưởng đến hiệu suất của thuật toán. Một K nhỏ có thể dẫn đến tăng cường độ nhiễu, trong khi một K lớn có thể dẫn đến mô hình quá đơn giản.
- Độ phức tạp không đồng đều: Trong không gian đa chiều, việc đo khoảng cách có thể trở nên phức tạp và dẫn đến việc giảm hiệu suất của thuật toán. [10]

2.5.4. *Random Forests*

2.5.4.1 *Khái niệm Random Forests*

Random Forest là một phương pháp học máy được sử dụng cho các bài toán phân loại và hồi quy. Nó được phát triển bởi Leo Breiman và Adele Cutler, và là một dạng mở rộng của phương pháp cây quyết định (Decision Trees). Random Forest sử dụng một tập

hợp (ensemble) của nhiều cây quyết định, kết hợp lại để đưa ra dự đoán chính xác hơn và giảm thiểu nguy cơ quá khớp (overfitting). [11]

2.5.4.2. Cách ứng dụng và hoạt động của Random Forests

Cách ứng dụng qua các lĩnh vực của Random Forests:

Ngân hàng: Lĩnh vực ngân hàng chủ yếu sử dụng thuật toán này để xác định rủi ro cho vay.

Y học: Với sự trợ giúp của thuật toán này, các xu hướng bệnh tật và nguy cơ của bệnh có thể được xác định.

Sử dụng đất: Chúng tôi có thể xác định các khu vực sử dụng đất tương tự bằng thuật toán này.

Tiếp thị: Các xu hướng tiếp thị có thể được xác định bằng cách sử dụng thuật toán này. [11]

Cách hoạt động của Random Forests

Random Forest xây dựng một "rừng" (forest) bao gồm nhiều cây quyết định, mỗi cây được huấn luyện trên một mẫu dữ liệu khác nhau được lấy từ tập dữ liệu gốc bằng phương pháp bootstrap (lấy mẫu ngẫu nhiên có hoàn lại). Các cây trong rừng hoạt động độc lập và kết quả cuối cùng được kết hợp lại để đưa ra dự đoán.

Các bước chính:

- Bootstrap Sampling: Từ tập dữ liệu gốc, lấy mẫu ngẫu nhiên có hoàn lại để tạo ra nhiều tập con khác nhau.
- Xây dựng Cây Quyết Định: Trên mỗi tập con, xây dựng một cây quyết định. Khi xây dựng mỗi cây, tại mỗi nút chia, chỉ một tập hợp ngẫu nhiên của các đặc tính được xem xét để tìm ra đặc tính tốt nhất để chia.
- Kết hợp Dự đoán: Đối với bài toán phân loại, lấy dự đoán cuối cùng là lớp có số phiếu bầu cao nhất từ tất cả các cây (majority voting). Đối với bài toán hồi quy, lấy trung bình các giá trị dự đoán từ các cây. [11]

2.5.4.3. Ưu điểm và nhược điểm Random Forests

Ưu điểm của Random Forests

- Giảm thiểu quá khớp (overfitting): Sự ngẫu nhiên trong việc lấy mẫu và lựa chọn đặc tính giúp giảm thiểu nguy cơ quá khớp so với việc sử dụng một cây quyết định duy nhất.
- Ổn định và mạnh mẽ: Random Forest thường có hiệu suất tốt ngay cả khi có nhiễu trong dữ liệu và không bị ảnh hưởng mạnh bởi các giá trị ngoại lệ.
- Xử lý dữ liệu đa dạng: Có thể xử lý cả dữ liệu phân loại và dữ liệu liên tục.
- Quan trọng của đặc tính: Random Forest cung cấp thông tin về tầm quan trọng của các đặc tính trong việc đưa ra quyết định, giúp hiểu rõ hơn về dữ liệu. [11]

Nhược điểm của Random Forests

- Tính toán phức tạp: Việc xây dựng nhiều cây quyết định và kết hợp kết quả có thể đòi hỏi tài nguyên tính toán lớn.
- Giải thích mô hình: Khó giải thích hơn so với cây quyết định đơn giản, do tính phức tạp và sự kết hợp của nhiều cây. [11]

CHƯƠNG 3. XÂY DỰNG WEBSITE PHÂN LOẠI VÀ GỢI Ý SẢN PHẨM

3.1. Giới thiệu và mô tả về trang web “Aimée Clothing”

Trang web quần áo Aimée Clothing là một trong tư vấn trực tuyến phổ biến. Aimée Clothing là một thương hiệu thời trang, nổi tiếng với các sản phẩm thời trang đa dạng quần áo nữ. Trang web Aimée cung cấp một trải nghiệm tư vấn sản phẩm trực tuyến tuyệt vời với hàng ngàn sản phẩm mới được cập nhật thường xuyên.

Mỗi sản phẩm trên trang web không chỉ là một món đồ mà còn là một câu chuyện về phong cách và cái đẹp. Từ những bộ trang phục hàng ngày đến những bộ sưu tập dành riêng cho các dịp đặc biệt cho phụ nữ, mỗi sản phẩm đều được thiết kế với sự tỉ mỉ và tinh tế, thể hiện tinh thần sáng tạo và đổi mới.

Trang web của Aimée Clothing không chỉ đơn thuần là một nơi tư vấn trực tuyến mà còn là một trải nghiệm đích thực cho người yêu thời trang. Tại đây, bạn sẽ được khám phá những xu hướng mới nhất, những bộ sưu tập độc đáo và những sản phẩm chất lượng, tạo nên không gian tư vấn trực tuyến đích thực cho những người đam mê thời trang.

Bên cạnh đó, dịch vụ khách hàng của Aimée Clothing cũng là một điểm nhấn đáng kể. Với đội ngũ nhân viên chuyên nghiệp và thân thiện cam kết mang đến cho khách hàng những trải nghiệm tư vấn trực tuyến tuyệt vời nhất, từ quá trình lựa chọn sản phẩm đến quá trình giao hàng và hậu mãi.

Trang web quần áo Aimée Clothing không chỉ là nơi tư vấn trực tuyến mà còn là biểu tượng của phong cách và cái đẹp, nơi mà người yêu thời trang có thể tự tin khám phá và thể hiện phong cách riêng của mình.

Đặc tả các chức năng của trang web:

Trang chủ:

- **Mô tả:** Trang chủ là trang đầu tiên khi người dùng truy cập vào trang web. Đây là nơi hiển thị các thông tin quan trọng như các sản phẩm mới nhất, các chương trình khuyến mãi, tin tức thời trang, và các thông báo quan trọng.

- **Chức năng:** Trang chủ giúp người dùng dễ dàng tiếp cận và khám phá các sản phẩm và thông tin mới nhất.

Trang giới thiệu:

- **Mô tả:** Trang giới thiệu cung cấp thông tin chi tiết về lịch sử và giá trị cốt lõi của thương hiệu. Đây là nơi mà người dùng có thể hiểu rõ hơn về nguồn gốc và triết lý kinh doanh của Aiméee Clothing.
- **Chức năng:** Trang giới thiệu giúp xây dựng lòng tin và tạo niềm tin cho khách hàng.

Trang sản phẩm

- **Mô tả:** Trang sản phẩm là nơi hiển thị danh sách các sản phẩm thời trang của Aiméee, được phân loại theo từng danh mục.
- **Chức năng:** Người dùng có thể xem và tìm kiếm các sản phẩm theo nhu cầu và sở thích cá nhân trên trang sản phẩm.

Trang tư vấn khách hàng:

- **Mô tả:** Trang dịch vụ khách hàng cung cấp thông tin về các dịch vụ hỗ trợ.
- **Chức năng:** Người dùng có thể tìm hiểu và sử dụng các dịch vụ hỗ trợ để có trải nghiệm tư vấn trực tuyến thuận lợi và dễ dàng hơn.

Trang mô tả sản phẩm:

- **Mô tả:** Trang mô tả sản phẩm cung cấp thông tin chi tiết về từng sản phẩm được bán trên trang. Trang bao gồm các hình ảnh, mô tả, giá cả, và các thông tin khác liên quan đến sản phẩm nhằm giúp khách hàng hiểu rõ hơn về sản phẩm trước khi mua..
- **Chức năng:** Hiển thị hình ảnh sản phẩm, mô tả chi tiết, thông tin sản phẩm, giá cả.

Trang Đăng ký và đăng nhập:

- **Mô tả:** Cho phép người dùng tạo tài khoản và đăng nhập để sử dụng các dịch vụ cá nhân hóa.
- **Chức năng:** Người dùng điền thông tin đăng ký hoặc thông tin đăng nhập. Hệ thống xác thực và cho phép truy cập vào tài khoản.

Trang quản lý người dùng:

- **Mô tả:** Trang quản lý người dùng là nơi quản lý có thể theo dõi và quản lý thông tin tài khoản của người dùng, bao gồm thông tin cá nhân và thêm sửa xóa người dùng
- **Chức năng:** Hiển thị danh sách tất cả người dùng đã đăng ký, cho phép quản trị viên chỉnh sửa thông tin cá nhân của người dùng, sửa thông tin người dùng, xóa tài khoản của người dùng khỏi hệ thống,

Trang quản lý sản phẩm:

- **Mô tả:** Trang quản lý sản phẩm là nơi quản lý có thể thêm mới, chỉnh sửa, xóa và quản lý các thông tin liên quan đến sản phẩm, bao gồm tên sản phẩm, giá cả, mô tả, hình ảnh.
- **Chức năng:** Cho phép quản lý thêm sản phẩm mới vào hệ thống, cập nhật thông tin của các sản phẩm hiện có, xóa sản phẩm khỏi hệ thống

Mô tả bài toán

Bài toán đặt ra là phát triển một ứng dụng website tự động gợi ý và tư vấn cho khách hàng khi họ tư vấn quần áo trực tuyến. Trong thị trường bán lẻ ngày càng chuyển hướng sang mô hình trực tuyến, việc cung cấp trải nghiệm tư vấn cá nhân hóa và tương tác tốt giữa cửa hàng và khách hàng trở thành yếu tố quyết định để thu hút và giữ chân khách hàng.

Ứng dụng sẽ hoạt động dựa trên việc thu thập và phân tích thông tin từ khách hàng như sở thích cá nhân và xu hướng thị trường. Các thuật toán và công nghệ hiện đại sẽ được áp dụng để từ các dữ liệu này, ứng dụng có thể đưa ra các gợi ý sản phẩm phù hợp và tư vấn thời trang cho từng khách hàng cá nhân.

Mục tiêu của bài toán là tạo ra một trải nghiệm trực tuyến tốt nhất cho khách hàng, giúp họ dễ dàng tìm kiếm và lựa chọn sản phẩm, đồng thời tăng cường tương tác và gắn kết giữa cửa hàng và khách hàng.

3.1.1. Chức năng, ưu điểm và nhược điểm của hệ thống

Chức năng của hệ thống:

Aiméee có một loạt các chức năng hệ thống trên trang web của họ để cung cấp trải nghiệm tư vấn trực tuyến thuận tiện và đáng tin cậy cho khách hàng.

- Tìm kiếm sản phẩm: Khách hàng có thể tìm kiếm sản phẩm theo từ khóa, danh mục.
- Gợi ý sản phẩm: Hệ thống sử dụng thuật toán và công nghệ để phân tích sở thích cá nhân và xu hướng thị trường của khách hàng, từ đó đề xuất những sản phẩm phù hợp nhất với từng người dùng.
- Tư vấn thời trang: Hệ thống cung cấp thông tin về xu hướng thời trang, cách phối đồ và các bộ sưu tập mới nhất để giúp khách hàng có cái nhìn tổng quan về thị trường thời trang và lựa chọn sản phẩm phù hợp
- Thông tin sản phẩm chi tiết: Cung cấp thông tin chi tiết về sản phẩm, bao gồm mô tả, hình ảnh, chất liệu.

Ưu điểm nổi bật:

- Sản phẩm đa dạng: Aiméee Clothing cung cấp một loạt các sản phẩm thời trang từ quần áo nữ. Sự đa dạng này giúp họ thu hút một đối tượng khách hàng rộng lớn.
- Thịnh hành và xu hướng: Aiméee Clothing thường xuyên cập nhật các sản phẩm mới theo xu hướng thời trang hiện đại. Điều này giúp họ thu hút người mua hàng đang tìm kiếm những sản phẩm thời trang mới nhất và phù hợp với phong cách cá nhân.
- Trải nghiệm tư vấn trực tuyến thuận tiện: Trang web của Aiméee Clothing được thiết kế một cách thông minh và dễ sử dụng, cung cấp trải nghiệm trực tuyến thuận tiện cho người tiêu dùng. Họ cung cấp các tính năng như tìm kiếm nhanh chóng.
- Dịch vụ khách hàng: Aiméee Clothing cung cấp một dịch vụ khách hàng chuyên nghiệp và thân thiện, từ việc hướng dẫn tư vấn đến hỗ trợ sản phẩm. Đội ngũ nhân viên luôn sẵn lòng hỗ trợ khách hàng mọi lúc.

Nhược điểm của hệ thống:

- Phụ thuộc vào dữ liệu đầu vào: Nếu dữ liệu về sản phẩm và khách hàng không đầy đủ hoặc không chính xác, hệ thống sẽ khó đưa ra các gợi ý phù hợp. Chất lượng gợi ý phụ thuộc mạnh vào độ chính xác và tính toàn diện của dữ liệu thu thập được từ khách hàng.

- Khả năng xử lý lỗi và ngoại lệ: Hệ thống có thể gặp khó khăn trong việc xử lý các tình huống ngoại lệ hoặc các trường hợp đặc biệt mà không có dữ liệu tương ứng..
- Cần cập nhật liên tục: Thị trường thời trang thay đổi liên tục với các xu hướng mới xuất hiện nhanh chóng. Hệ thống cần được cập nhật thường xuyên để theo kịp các xu hướng mới này. Việc không cập nhật kịp thời có thể làm giảm hiệu quả của hệ thống và làm mất lòng tin của khách hàng.

3.1.2. Giới thiệu đặc điểm nổi bật hơn của trang web đang xây dựng

Để có một trang web tư vấn quần áo, có một số đặc điểm nổi bật cần được xây dựng và thúc đẩy:

- Giao diện dễ sử dụng và hấp dẫn: Giao diện trang web cần phải thân thiện với người dùng, dễ sử dụng và hấp dẫn. Các sản phẩm nên được trình bày một cách rõ ràng và dễ nhìn, giúp người dùng dễ dàng tìm kiếm và chọn lựa. Sử dụng chatbot để hỗ trợ khách hàng 24/7.
- Hình ảnh chất lượng cao: Hình ảnh của sản phẩm cần phải rõ ràng và sắc nét, giúp người dùng có cái nhìn tổng quan và chính xác về sản phẩm trước khi mua hàng.
- Tính cá nhân hóa: Trang web có thể tăng cường trải nghiệm bằng cách cá nhân hóa nội dung và gợi ý sản phẩm dựa trên lịch sử và sở thích của người dùng.
- Bảo mật thông tin: Đảm bảo rằng trang web có các biện pháp bảo mật chặt chẽ để bảo vệ thông tin cá nhân và thanh toán của người dùng, giúp họ cảm thấy an tâm khi tư vấn trực tuyến.
- Cung cấp thông tin chi tiết: Mỗi sản phẩm cần có mô tả chi tiết, kích cỡ, chất liệu, hướng dẫn chăm sóc và một số hình ảnh chất lượng để người dùng có thể đánh giá sản phẩm một cách đầy đủ trước khi mua hàng
- Sử dụng học máy: Hệ thống học máy phân tích sở thích cá nhân của người dùng để đề xuất các sản phẩm phù hợp với sở thích cá nhân.

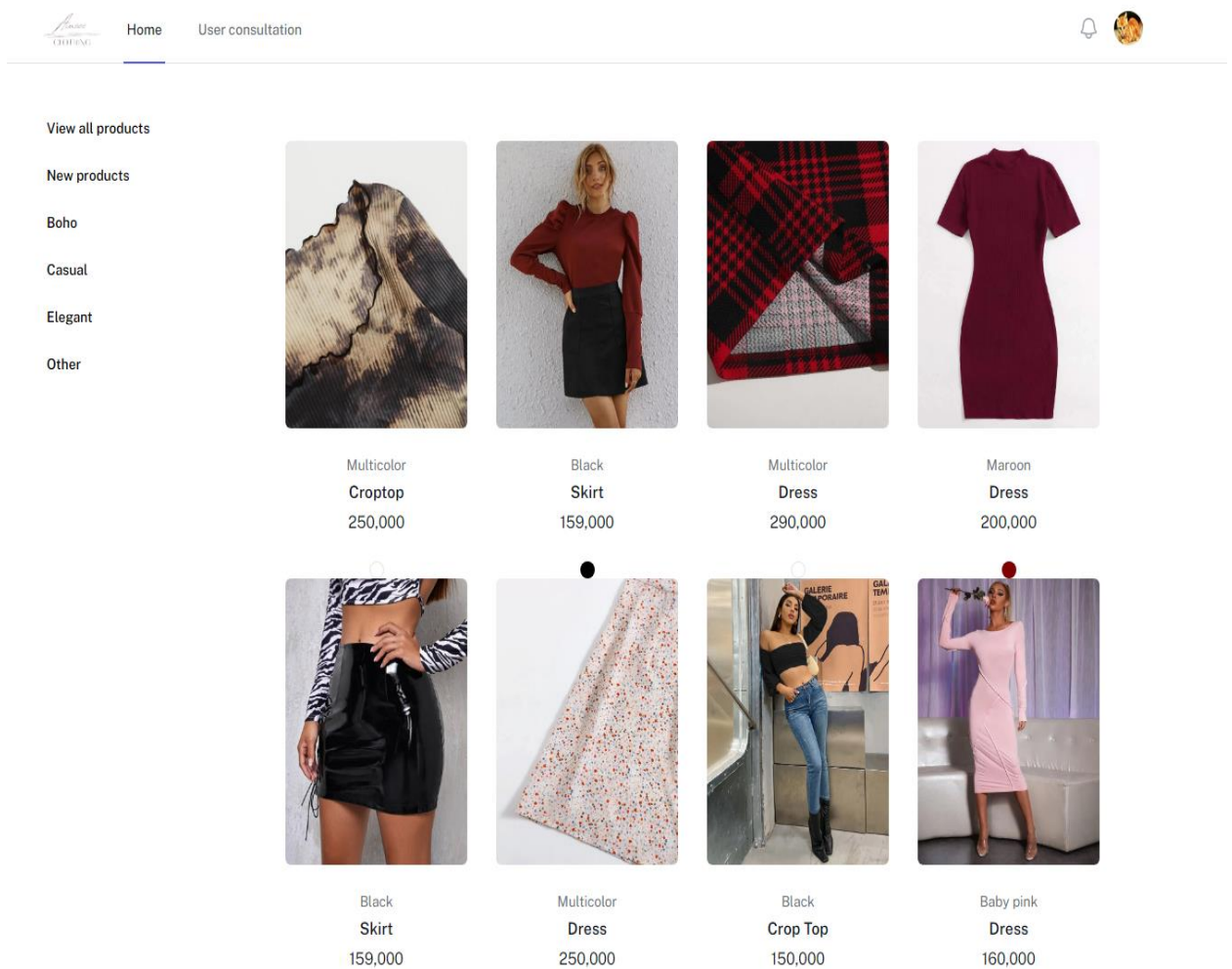
3.1.3. Xác định Actor

Bảng 3.1: Xác định Actor

| STT | Tên actor | Tên chức năng | Mô tả |
|-----|-----------|------------------------|--|
| 1 | Quản lý | Đăng nhập | Chức năng để quản lý đăng nhập của người dùng |
| | | Xem thông tin sản phẩm | Chức năng để người dùng xem chi tiết về các sản phẩm, bao gồm thông tin như tên sản phẩm, giá cả, mô tả, hình ảnh và các thông tin liên quan khác. |
| | | Xem thông tin cá nhân | Chức năng cho phép người dùng xem và cập nhật thông tin cá nhân của mình, bao gồm tên, địa chỉ, số điện thoại, email và các thông tin khác |
| | | Quản lý người dùng | Chức năng để thêm, sửa và xóa thông tin của người dùng, cũng như quản lý quyền truy cập của họ. |
| | | Quản lý sản phẩm | Chức năng để thêm, sửa và xóa sản phẩm trong cơ sở dữ liệu, cũng như quản lý thông tin và giá cả của sản phẩm. |
| | | Quản lý danh mục | Chức năng để quản lý các danh mục sản phẩm, bao gồm việc thêm, sửa và xóa danh mục cũng như phân loại sản phẩm. |

| STT | Tên actor | Tên chức năng | Mô tả |
|-----|------------|----------------------------|---|
| 2 | Người dùng | Đăng nhập | Cho phép người dùng truy cập vào hệ thống bằng cách cung cấp thông tin đăng nhập như tên đăng nhập và mật khẩu. |
| | | Đăng ký | Người dùng khi chưa có tài khoản sẽ tạo ra một tài khoản mới trên hệ thống bằng cách điền vào một biểu mẫu đăng ký và cung cấp thông tin cá nhân cần thiết. |
| | | Đăng xuất | Người dùng đăng xuất khỏi hệ thống, kết thúc phiên làm việc hiện tại và bảo vệ thông tin cá nhân. |
| | | Tìm kiếm | Người dùng tìm kiếm thông tin, sản phẩm trên trang web bằng cách sử dụng thanh công cụ tìm kiếm. |
| | | Xem sản phẩm | Người dùng xem thông tin chi tiết về một sản phẩm cụ thể, bao gồm hình ảnh, mô tả, giá cả và các thông tin khác để đưa ra quyết định mua hàng. |
| | | Gợi ý sản phẩm qua chatbox | Người dùng tương tác qua trang web thông qua giao diện chat. |

3.1.4. Giao diện trang web



Hình 3.1. Giao diện trang chủ

Mô tả: Trang chủ của trang web là điểm tiếp xúc đầu tiên của người dùng khi truy cập vào. Bao gồm các sản phẩm quần áo nổi bật, xu hướng thời trang mới nhất, và các thông tin quan trọng khác. Trang chủ phải hấp dẫn và dễ điều hướng để người dùng có trải nghiệm mua sắm trực tuyến tốt nhất. Trang chủ của trang web cải thiện thêm với tính năng chat box, giúp người dùng có thể tương tác hỗ trợ, nhận hỗ trợ và giải đáp thắc mắc một cách nhanh chóng.

Home

User consultation

View all products

New products

Casual

Boho

Elegant

Other

Style

Choose...

Color

Choose...

Size

Choose...

Type

Choose...

Material

Choose...

Send

Solutions

Marketing

Analytics

Commerce

Support

Pricing

Documentation

Guides

Company

About

Blog

Jobs

Legal

Claim

Privacy

Terms

Subscribe to our newsletter

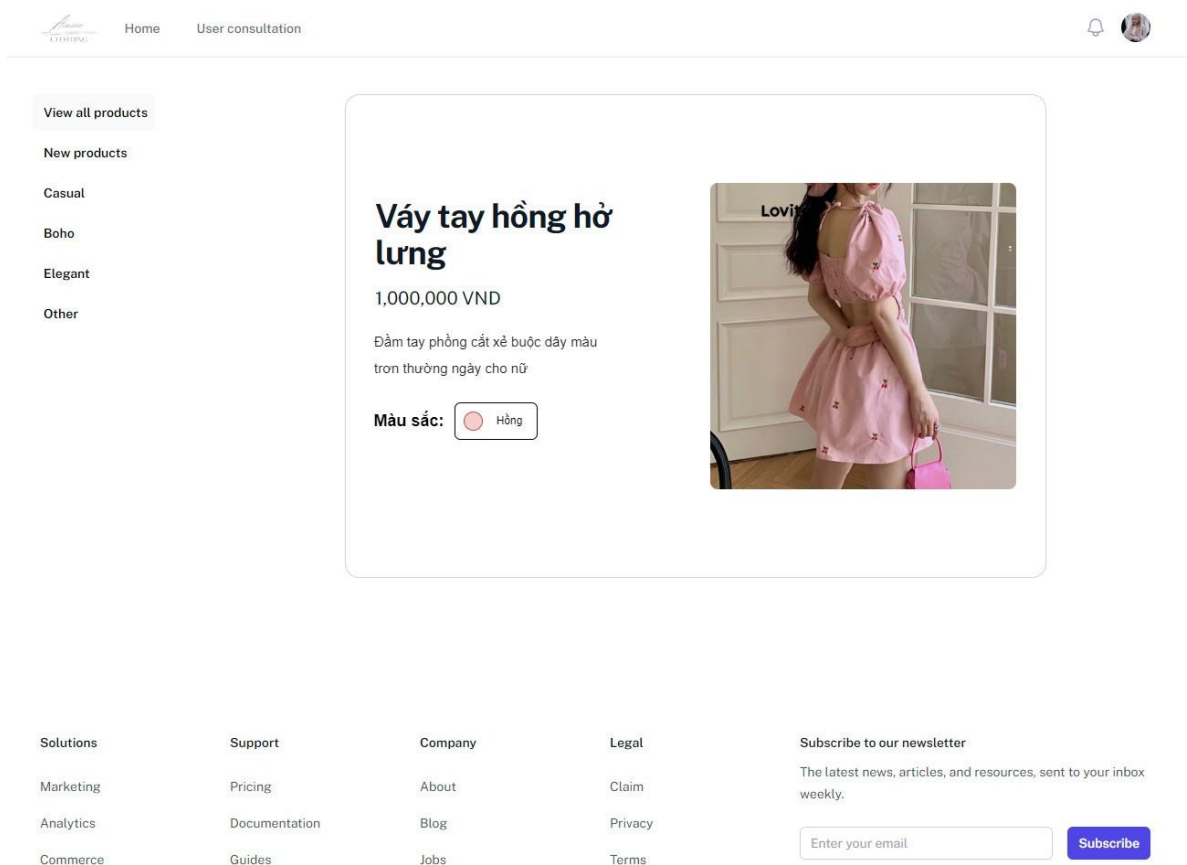
The latest news, articles, and resources, sent to your inbox weekly.

Enter your email

Subscribe

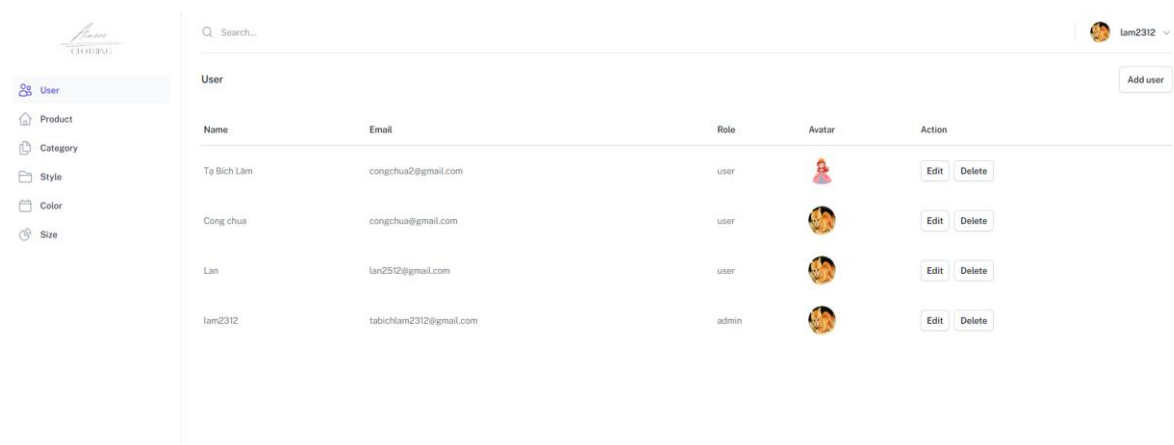
Hình 3.2. Giao diện trang gợi ý khách hàng

Mô tả: Trang gợi ý khách hàng được thiết kế với mục tiêu giúp khách hàng tìm sản phẩm phù hợp với nhu cầu và sở thích. Khi truy cập vào, khách hàng sẽ được yêu cầu lựa chọn cung cấp một số thông tin cơ bản về phong cách, màu sắc, kích cỡ, thể loại và thể loại theo tiêu chí mà họ đang tìm kiếm. Dựa trên những thông tin này, hệ thống sẽ sử dụng các thuật toán phân tích dữ liệu để đưa ra danh sách các sản phẩm được đề xuất, phù hợp với các yêu cầu của khách hàng.



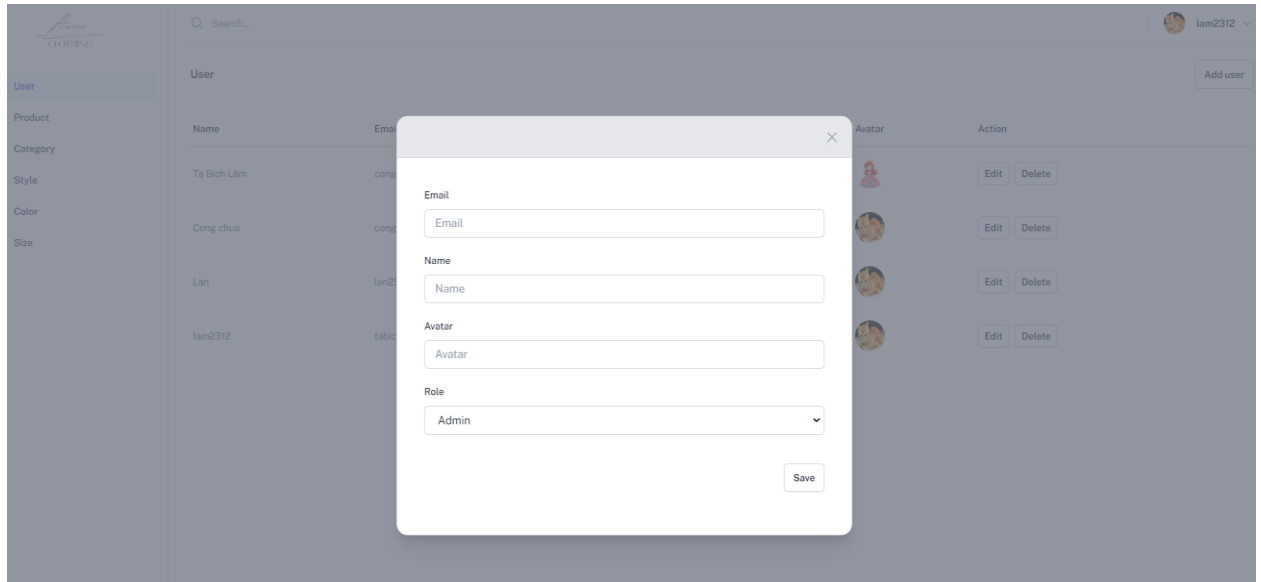
Hình 3.3. Giao diện trang mô tả sản phẩm

Mô tả: Trang giao diện mô tả sản phẩm cung cấp thông tin chi tiết về các sản phẩm. Đây là nơi mà khách hàng có thể tìm hiểu về sản phẩm một cách toàn diện trước khi quyết định mua hàng. Từ việc hiển thị hình ảnh chất lượng cao đến việc cung cấp thông tin chi tiết về giá cả, mô tả sản phẩm.



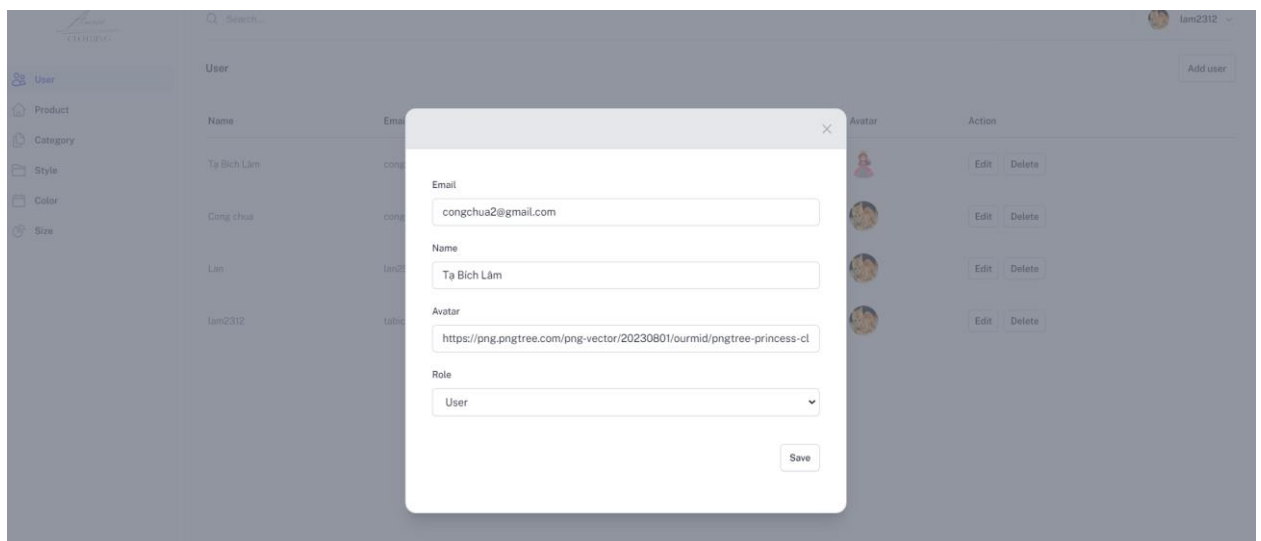
Hình 3.4. Giao diện trang admin

Mô tả: Giao diện trang admin là nơi quản lý và điều khiển toàn bộ nội dung, sản phẩm và hoạt động của trang web từ phía quản lý. Đây là nơi quan trọng để quản lý thêm/sửa/xóa sản phẩm, quản lý người dùng.



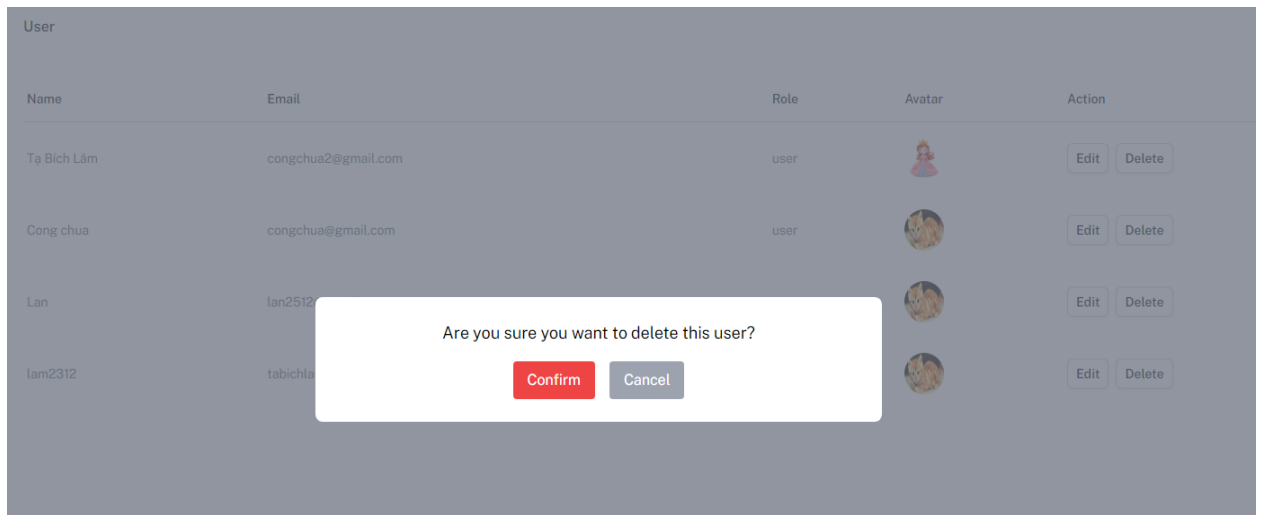
Hình 3.5. Giao diện thêm người dùng

Mô tả: Giao diện thêm người dùng là nơi cho phép quản lý thêm mới người dùng vào hệ thống của trang web. Qua giao diện này, quản trị viên có thể cung cấp thông tin cần thiết và quản lý quyền truy cập của người dùng.



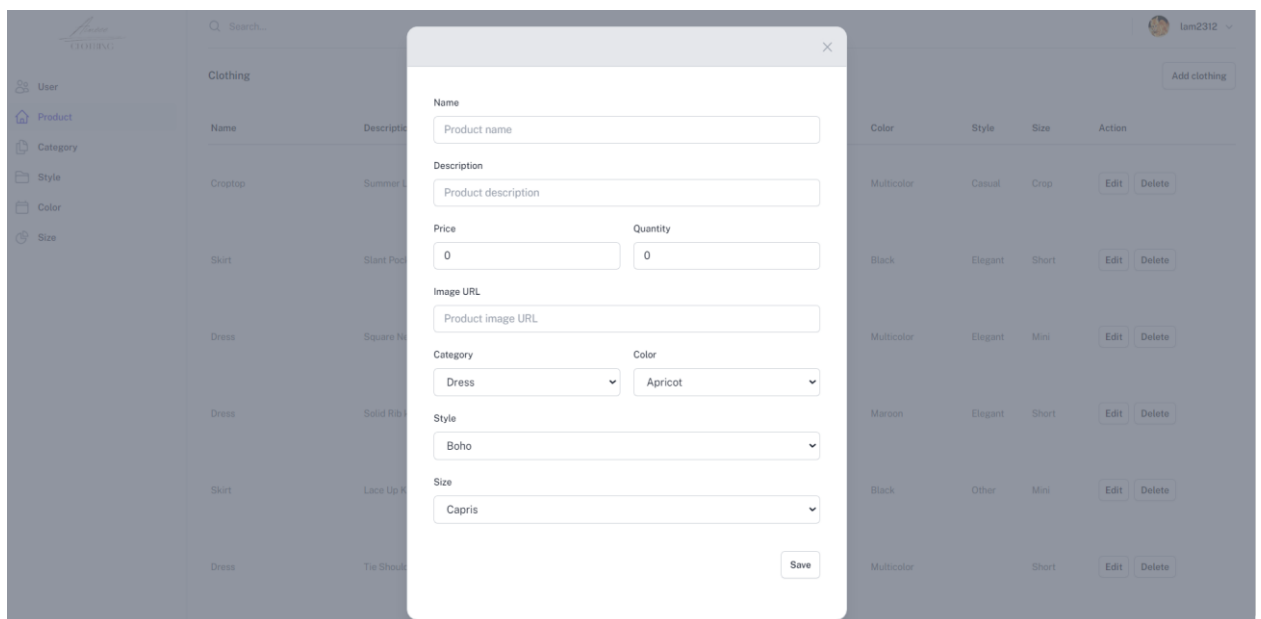
Hình 3.6. Giao diện sửa người dùng

Mô tả: Giao diện sửa người dùng là nơi cho phép quản lý chỉnh sửa thông tin và của người dùng trong hệ thống của trang web. Qua giao diện này, quản lý có thể cập nhật thông tin cá nhân, vai trò của người dùng.



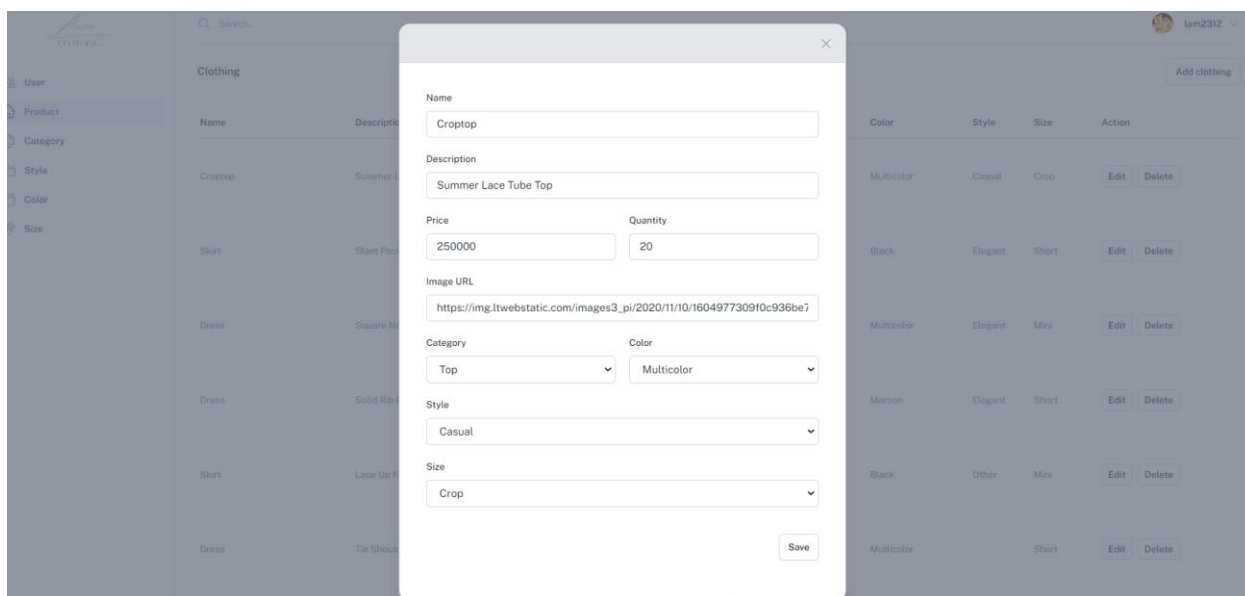
Hình 3.7. Giao diện xóa

Mô tả: Giao diện xóa là nơi cho phép quản lý xóa một người dùng khỏi hệ thống của trang web. Qua giao diện này, quản lý có thể xác nhận việc xóa người dùng và thông báo cho họ về hành động này nếu cần thiết.



Hình 3.8. Giao diện thêm sản phẩm

Mô tả: Giao diện thêm sản phẩm là nơi cho phép quản lý thêm mới các sản phẩm vào cửa hàng. Qua giao diện này, quản lý có thể cung cấp thông tin chi tiết về sản phẩm, bao gồm hình ảnh, mô tả, giá cả và thông tin khác.



The screenshot displays a web application interface for managing clothing items. A modal form is open in the center, allowing the user to add a new product. The form contains the following fields:

- Name:** Croptop
- Description:** Summer Lace Tube Top
- Price:** 250000
- Quantity:** 20
- Image URL:** https://img.ltwebstatic.com/images3_pi/2020/11/10/1604977309f0c936be7
- Category:** Top
- Color:** Multicolor
- Style:** Casual
- Size:** Crop

A 'Save' button is located at the bottom right of the modal. In the background, a table lists existing clothing items with columns for Color, Style, Size, and Action (Edit/Delete).

Hình 3.9. Giao diện sửa sản phẩm

Mô tả: Giao diện sửa sản phẩm là nơi cho phép quản lý chỉnh sửa thông tin chi tiết về một sản phẩm đã tồn tại trong cửa hàng trực tuyến. Qua giao diện này, quản lý có thể cập nhật hình ảnh, mô tả, giá cả và thông tin khác của sản phẩm.

3.2. Sử dụng mô hình học máy để xây dựng chức năng tự động gợi ý sản phẩm

3.2.1. Mô tả chức năng tự động gợi ý sản phẩm

Dữ liệu được lấy trong Dataset Women Clothes, dữ liệu cung cấp là một bảng tính chứa thông tin về các mặt hàng quần áo khác nhau. Mỗi hàng đại diện cho một trang phục cụ thể, và các cột chứa các thuộc tính khác nhau của trang phục.

Bảng 3.2. Thuộc tính và mô tả

| Thuộc Tính | Mô Tả Chi Tiết |
|--------------|---|
| Details | Các chi tiết đặc biệt hoặc phụ kiện đi kèm của sản phẩm, thường là các yếu tố làm nổi bật và tạo điểm nhấn cho sản phẩm, chẳng hạn như viền bèo (lettuce trim), gân (rib-knit). Waistline |
| Waistline | Vị trí của vòng eo sản phẩm, cho biết sản phẩm có vòng eo cao (high waist), tự nhiên (natural), hoặc thấp (low waist). Fabric |
| Fabric | Đặc điểm của chất liệu vải, chẳng hạn như có độ co giãn nhẹ (slight stretch), co giãn cao (high stretch), hoặc không co giãn (non-stretch). |
| Sleeve Type | Loại tay áo của sản phẩm, ví dụ: tay áo bình thường (regular sleeve), tay áo phồng (bishop sleeve). |
| Name | Tên gọi của sản phẩm, thường là tên thương hiệu kèm theo mô tả ngắn về sản phẩm. |
| Fit Type | Kiểu dáng và cách sản phẩm ôm sát vào cơ thể, như ôm sát (slim fit), vừa vặn bình thường (regular fit), rộng rãi (loose fit). |
| Style | Phong cách của sản phẩm, ví dụ: giản dị (casual), thanh lịch (elegant), thể thao (sporty), boho, quyến rũ (glamorous). |
| url | Đường dẫn URL đến trang web chứa sản phẩm. |
| Pattern Type | Loại họa tiết hoặc mẫu trên sản phẩm, ví dụ: nhuộm tie dye, kẻ ô (plaid), trơn (plain). |
| Material | Chất liệu chính của sản phẩm, ví dụ: cotton, polyester, PU leather. |
| Belt | Thông tin về việc sản phẩm có đi kèm dây lưng hay không. |
| Body | Loại sản phẩm, ví dụ: áo (top), quần (bottom), váy (dress). |

| Thuộc Tính | Mô Tả Chi Tiết |
|------------------|--|
| Color | Màu sắc chính của sản phẩm. |
| Placket Type | Loại khóa hoặc nút áo của sản phẩm, ví dụ: dạng kéo qua đầu (pullovers), cài nút (button fly). |
| Gender | Giới tính mà sản phẩm hướng đến, thường là nữ (F), nam (M), hoặc không phân biệt giới tính (unisex). |
| Hem Shaped | Hình dạng của đường gấu áo hoặc váy, ví dụ: bút chì (pencil), loe (flared). |
| Composition | Thành phần chất liệu của sản phẩm, bao gồm tỷ lệ các loại sợi như spandex, cotton, polyester. |
| Sheer | Độ trong suốt của sản phẩm, ví dụ: không trong suốt (no), hơi trong suốt (slightly sheer). |
| Image | Đường dẫn URL đến hình ảnh của sản phẩm. |
| Sleeve Length | Độ dài của tay áo, ví dụ: tay dài (long sleeve), tay ngắn (short sleeve), không tay (sleeveless). |
| Type | Loại chung của sản phẩm, ví dụ: áo (top), quần (bottom), váy (dress). |
| Arabian Clothing | Chỉ định xem sản phẩm có thuộc loại trang phục Ả Rập hay không. |
| Brand | Thương hiệu sản xuất sản phẩm. |
| Length | Độ dài của sản phẩm, ví dụ: crop (ngắn trên eo), mini (ngắn trên đầu gối), midi (dài đến bắp chân), long (dài đến mắt cá). |
| Chest pad | Sản phẩm có đệm ngực hay không. |

| Thuộc Tính | Mô Tả Chi Tiết |
|-------------------|---|
| Neckline | Kiểu dáng cổ áo, ví dụ: cổ tròn (round neck), cổ vuông (square neck), cổ cao (stand collar). |
| Closure Type | Loại khóa hoặc cách cài áo, ví dụ: khóa kéo (zipper fly), cài nút (button front). |
| Care Instructions | Hướng dẫn cách giặt và bảo quản sản phẩm. |
| Season | Mùa mà sản phẩm phù hợp để mặc, ví dụ: xuân (spring), hè (summer), thu (fall), đông (winter). |

Hiện thị kết quả: Các sản phẩm được gợi ý sẽ được hiển thị trên trang web dựa trên độ phù hợp với người dùng và mức độ ưa thích của họ.

3.2.2. Thống kê các trường dữ liệu trong Dataset Women Clothes

Dữ liệu bao gồm thông tin về bảy thuộc tính trong dataset. Số lượng dữ liệu thu thập cho từng thuộc tính là như sau:

Bảng 3.3: Thống kê các trường dữ liệu ban đầu

| Kích thước dữ liệu | Thuộc tính |
|--------------------------------|---|
| Gồm 29 trường và 14859 dữ liệu | Details, Waist, Line, Fabric, Sleeve Type, Name, Fit Type, Style, url, Pattern Type, Material, Belt, Body, Color, Placket, Type, Gender, Hem, Shaped, Composition, Sheer, Image, Sleeve, Length, Type Arabian Clothing, Brand, Length, Chest pad, Neckline, Closure Type Care, Instructions, Season |

Bảng 3.4: Dữ liệu sau khi làm sạch

| Kích thước dữ liệu | Thuộc tính |
|-------------------------------|--|
| Gồm 6 trường và 11859 dữ liệu | Style, Material, Length, Type, Summer, Color |

Thuộc tính: Đây là các đặc điểm của quần áo được ghi nhận trong dataset, bao gồm 'Style', 'Material', 'Color', 'Type', 'Length', và 'Season'.

Số lượng dữ liệu: Số lượng mẫu hoặc bản ghi cho từng thuộc tính trong dataset là 11859.

Số trường: Số lượng các thuộc tính khác nhau được ghi nhận là 6.

=> Dataset này có tổng cộng 11859 mẫu và mỗi mẫu chứa thông tin về 6 thuộc tính khác nhau của quần áo nữ.

3.2.3. Bộ dữ liệu sử dụng trong mô hình

- Tên bộ dữ liệu: Women's clothes
- Nguồn gốc :

Bộ dữ liệu này được lấy từ Kaggle, một nền tảng nổi tiếng cung cấp nhiều bộ dữ liệu cho các dự án học máy và phân tích dữ liệu. Bộ dữ liệu được chia sẻ bởi một người dùng trên Kaggle và có sẵn cho các mục đích nghiên cứu và phân tích. Link tới bộ dữ liệu trên Kaggle:

<https://www.kaggle.com/datasets/justjess678/womens-clothes/data> [12]

- Số trường: Bộ dữ liệu này có tổng cộng 29 trường.
- Số dữ liệu: Bộ dữ liệu bao gồm 14369.

3.2.4. Phân chia dữ liệu huấn luyện

Phân chia dữ liệu huấn luyện và kiểm tra là một bước quan trọng trong quy trình phát triển mô hình học máy. Giúp đánh giá hiệu suất của mô hình trên dữ liệu chưa từng thấy và đảm bảo rằng mô hình không bị overfit vào dữ liệu huấn luyện. Tỷ lệ phân chia phổ biến là 80:20, tức là 80% dữ liệu sẽ được sử dụng để huấn luyện mô hình và 20% còn lại sẽ được sử dụng để kiểm tra mô hình.

Lý do phân chia dữ liệu: đánh giá mô hình phân chia dữ liệu giúp đánh giá hiệu suất của mô hình trên dữ liệu mới, từ đó có thể xác định được mô hình có bị overfit hoặc underfit hay không và kiểm tra tính tổng quát việc kiểm tra mô hình trên dữ liệu chưa từng thấy giúp đảm bảo rằng mô hình có thể tổng quát hóa tốt và hoạt động tốt trên dữ liệu thực tế.

Sau khi dữ liệu đã được chia thành các tập huấn luyện và kiểm tra, SVM, KNN, Decision Tree, được huấn luyện trên tập huấn luyện và sau đó được sử dụng để gợi ý quần áo cho cửa hàng tư vấn. Quá trình sẽ diễn ra như sau kết quả của việc dự đoán này được sử dụng để đánh giá hiệu suất của mô hình trên dữ liệu mới, không nhìn thấy trước đó, thông qua các độ đo như độ chính xác.

3.2.5 Phân tích bộ dữ liệu

Tiền xử lý dữ liệu là một bước quan trọng và không thể thiếu trong quá trình phân tích dữ liệu và xây dựng mô hình học máy. Mục tiêu của tiền xử lý dữ liệu là biến đổi dữ liệu thô thành dạng sạch, nhất quán và dễ sử dụng cho các bước phân tích và mô hình hóa tiếp theo.

Làm sạch dữ liệu (Data Cleaning):

- Loại bỏ các thuộc tính không cần thiết: Trong tập dữ liệu ban đầu có nhiều thuộc tính, nhưng không phải tất cả đều cần thiết cho mục đích của ứng dụng. Việc loại bỏ các thuộc tính không cần thiết sẽ giúp tập trung vào những thông tin quan trọng và giảm thiểu độ phức tạp của dữ liệu.

Ví dụ: Chỉ giữ lại các thuộc tính ‘Style’, ‘Material’, ‘Color’, ‘Type’, ‘Length’, ‘Season’.

- Xử lý giá trị thiếu (Missing Values): Giá trị thiếu là một vấn đề phổ biến trong các tập dữ liệu và cần được xử lý trước khi tiến hành phân tích. Việc này giúp đảm bảo tính toàn vẹn và chất lượng của dữ liệu.

Xác định giá trị thiếu: Sử dụng các kỹ thuật để xác định vị trí và số lượng giá trị thiếu trong tập dữ liệu.

Xử lý giá trị thiếu:

Loại bỏ giá trị thiếu: Loại bỏ các hàng hoặc cột chứa nhiều giá trị thiếu hơn một ngưỡng xác định trước (ví dụ: nếu hơn 30% giá trị trong một cột là thiếu, cột đó có thể bị loại bỏ).

Điền giá trị thiếu: Điền các giá trị thiếu bằng cách sử dụng giá trị trung bình, trung vị, hoặc giá trị phổ biến nhất của cột đó. Đối với các thuộc tính phân loại, có thể điền bằng giá trị chế độ (mode).

Loại bỏ các giá trị ngoại lai (Outliers): Giá trị ngoại lai có thể ảnh hưởng tiêu cực đến phân tích dữ liệu và mô hình dự báo. Do đó, cần phải xác định và xử lý các giá trị này.

Xác định giá trị ngoại lai: Sử dụng các phương pháp thống kê (như IQR - Interquartile Range, z-score) để xác định các giá trị nằm ngoài khoảng giá trị bình thường.

Xử lý giá trị ngoại lai: Loại bỏ hoặc điều chỉnh các giá trị này để giảm thiểu ảnh hưởng của chúng.

- Chuẩn hóa dữ liệu (Normalization): Chuẩn hóa dữ liệu giúp đưa các thuộc tính về cùng một thang đo, tạo điều kiện thuận lợi cho quá trình phân tích và mô hình hóa.

Phương pháp chuẩn hóa: Sử dụng các kỹ thuật như Min-Max Scaling hoặc Z-score Scaling để chuẩn hóa các giá trị.

Đảm bảo tính nhất quán và hợp lệ của dữ liệu: Kiểm tra và đảm bảo rằng dữ liệu không chứa các giá trị không hợp lệ hoặc không nhất quán.

Kiểm tra tính hợp lệ: Xác định và loại bỏ các giá trị không hợp lệ (ví dụ: giá trị âm trong các thuộc tính không thể có giá trị âm).

Đảm bảo tính nhất quán: Đảm bảo rằng các thuộc tính phân loại có các giá trị nhất quán và đồng nhất (ví dụ: không có lỗi chính tả trong các giá trị phân loại)

Chuyển đổi dữ liệu (Data Transformation)

- Chuẩn hóa (Normalization): Chuẩn hóa dữ liệu giúp đưa các thuộc tính về cùng một thang đo, tạo điều kiện thuận lợi cho quá trình phân tích và mô hình hóa. Có hai phương pháp chính:
 - Min-Max Scaling: Chuyển đổi các giá trị về khoảng từ 0 đến 1.
 - Z-score Scaling: Chuyển đổi các giá trị sao cho dữ liệu có trung bình là 0 và độ lệch chuẩn là 1.

Chuẩn hóa giúp dữ liệu trở nên đồng nhất, dễ so sánh và phù hợp cho các mô hình học máy nhạy cảm với khoảng cách như KNN hoặc SVM.

- Mã hóa (Encoding): Mã hóa là quá trình chuyển đổi các thuộc tính phân loại thành các giá trị số để thuận tiện cho các mô hình học máy. Có hai phương pháp chính:

One-hot Encoding: Sử dụng khi thuộc tính phân loại có nhiều giá trị khác nhau. Mỗi giá trị sẽ được biểu diễn bằng một cột riêng với giá trị 0 hoặc 1. Ví dụ, thuộc tính Color có các giá trị Red, Green, Blue sẽ được chuyển thành ba cột: Color_Red, Color_Green, Color_Blue.

Label Encoding: Sử dụng khi thuộc tính phân loại có giá trị thứ tự hoặc chỉ có hai giá trị. Mỗi giá trị sẽ được mã hóa thành một số nguyên duy nhất. Ví dụ, thuộc tính Gender có hai giá trị F và M sẽ được mã hóa thành 0 và 1.

Quy ước về số (Numerical Convention): Chuyển đổi các thuộc tính phân loại thành các giá trị số tương ứng với các danh mục để dễ dàng sử dụng trong các mô hình học máy.

Quy ước số :

- Season: Spring = 1, Summer = 2, Winter = 4.
- Tiền xử lý thuộc tính văn bản (Text Preprocessing): Các thuộc tính như URL và Image không cần chuyển đổi, giữ nguyên định dạng văn bản vì chúng được sử dụng để tham chiếu đến hình ảnh và trang sản phẩm.

Các bước chi tiết cho từng thuộc tính:

Bảng 3.5: Bước chi tiết cho từng thuộc tính

| Thuộc tính | Chuyển đổi |
|--------------|---|
| Style | Mã hóa One-hot Encoding: Chuyển đổi từng giá trị của thuộc tính 'Style' thành một cột riêng biệt với giá trị 1 đến 4. |
| URL và Image | Giữ nguyên: Không cần thay đổi. Dùng để tham chiếu tới trang sản phẩm và hình ảnh sản phẩm. |
| Material | Mã hóa One-hot Encoding: Chuyển đổi từng giá trị của thuộc tính |

| Thuộc tính | Chuyển đổi |
|------------|---|
| | ‘Material’ thành một cột riêng biệt với giá trị 1 đến 27. |
| Color | Mã hóa One-hot Encoding: Chuyển đổi từng giá trị của thuộc tính ‘Color’ thành một cột riêng biệt với giá trị 1 đến 57. |
| Gender | Giữ nguyên: Không cần thay đổi. |
| Type | Mã hóa One-hot Encoding: Chuyển đổi từng giá trị của thuộc tính ‘Type’ thành một cột riêng biệt với giá trị 1 hoặc 3. |
| Length | Mã hóa One-hot Encoding: Chuyển đổi từng giá trị của thuộc tính ‘Length’ thành một cột riêng biệt với giá trị 1 hoặc 10. |
| Season | Mã hóa One-hot Encoding: Tách các mùa trong danh sách thành các cột riêng biệt, mỗi cột biểu diễn một mùa với giá trị 1, 2, 4, 12, 13, 34, 123, 1234. |

Xử Lý Dữ Liệu Bị Thiếu (Handling Missing Data)

Dữ liệu bị thiếu là một vấn đề phổ biến trong bất kỳ bộ dữ liệu nào. Xử lý dữ liệu bị thiếu là một bước quan trọng trong quá trình tiền xử lý dữ liệu để đảm bảo tính toàn vẹn và chất lượng của dữ liệu. Việc xử lý dữ liệu bị thiếu có thể bao gồm:

Loại bỏ cột: Loại bỏ các cột có nhiều giá trị bị thiếu.

⇒ Ví dụ: Nếu một hàng có hơn 2 thuộc tính bị thừa, có thể quyết định loại bỏ.

Loại bỏ cột: Khi một cột có nhiều giá trị thừa, có thể loại bỏ cột đó nếu nó không quan trọng.

⇒ Ví dụ: Nếu một cột có hơn 2 giá trị bị thừa, có thể quyết định loại bỏ cột đó, đặc biệt khi cột này không quan trọng cho phân tích hoặc mô hình hóa.

Điền giá trị thay thế: Điền các giá trị bị thiếu bằng cách sử dụng các giá trị thay thế có thể giữ lại nhiều dữ liệu hơn so với việc loại bỏ hàng/cột.

Lưu Trữ Dữ Liệu Đã Tiền Xử Lý

Sau khi hoàn thành các bước tiền xử lý dữ liệu, việc lưu trữ dữ liệu sạch vào file hoặc cơ sở dữ liệu là bước cuối cùng. Bước này giúp bảo toàn công sức đã bỏ ra để làm sạch và chuẩn bị dữ liệu, đồng thời tạo điều kiện thuận lợi cho các bước phân tích và mô hình hóa tiếp theo.

- Lý do lưu trữ dữ liệu đã tiền xử lý

Bảo toàn công sức: Quá trình tiền xử lý dữ liệu thường tốn nhiều thời gian và công sức. Việc lưu trữ dữ liệu giúp tránh việc phải thực hiện lại các bước này.

Tính nhất quán: Lưu trữ dữ liệu tiền xử lý đảm bảo tính nhất quán cho các lần phân tích sau này, giúp tránh sai sót do dữ liệu thay đổi.

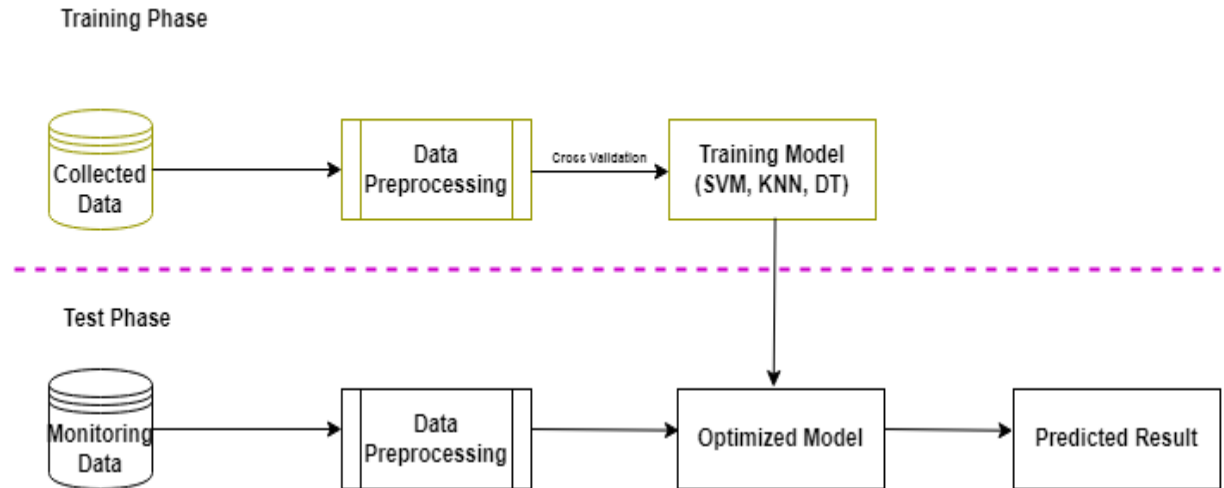
Thuận tiện cho phân tích: Dữ liệu đã tiền xử lý sẵn sàng để sử dụng, giúp các nhà phân tích và nhà khoa học dữ liệu tiết kiệm thời gian và tập trung vào việc phân tích và xây dựng mô hình.

- Lưu trữ vào file: Excel (Hữu ích cho việc chia sẻ dữ liệu với người dùng không chuyên về lập trình)

CHƯƠNG 4 TỐI ƯU HÓA MÔ HÌNH KỸ THUẬT HỌC MÁY

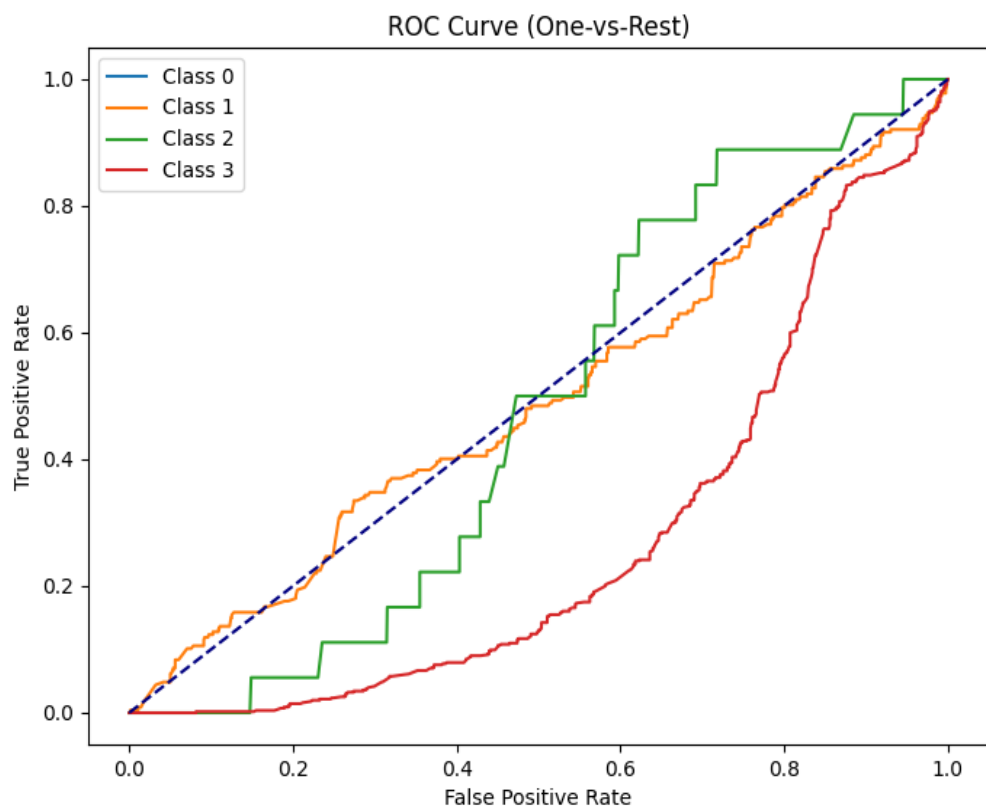
4.1. Lựa chọn mô hình học máy

Giới thiệu mô hình

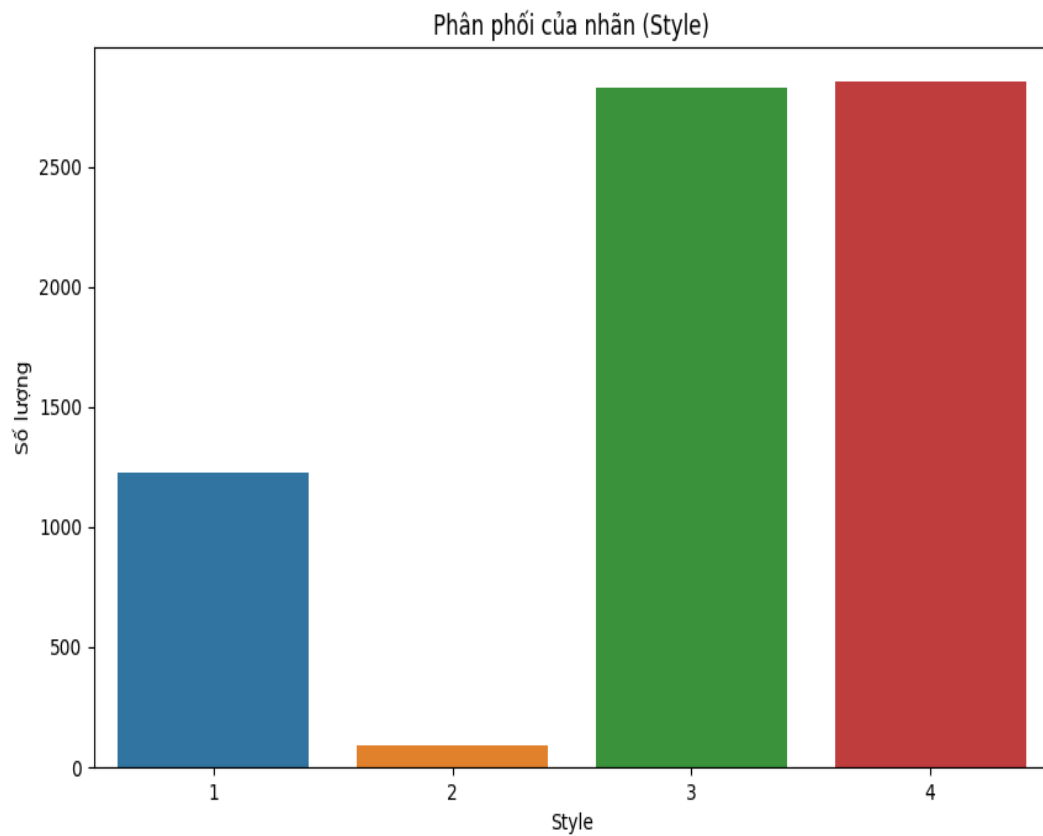


Hình 4.1. Mô hình Training và Test

- Mô hình SVM:



Hình 4.2. Sơ đồ ROC SVM



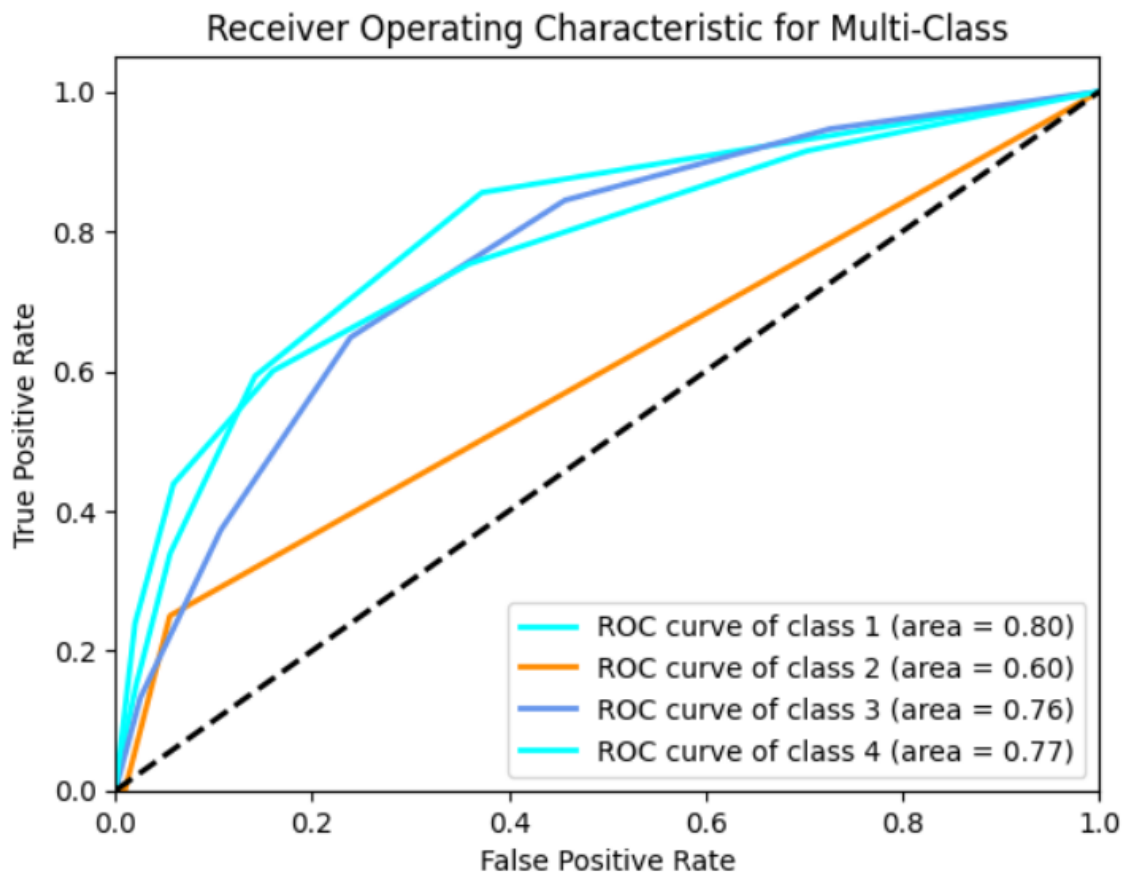
Hình 4.3. Phân phối nhãn style

| | Style | Material | Color | Type | Length | Season |
|---|-------|----------|-------|------|--------|--------|
| 0 | 3 | 23 | 6 | 1 | 10 | 123 |
| 1 | 3 | 16 | 38 | 2 | 8 | 2 |
| 2 | 3 | 4 | 33 | 2 | 10 | 2 |
| 3 | 4 | 17 | 6 | 1 | 8 | 1234 |
| 4 | 1 | 16 | 38 | 2 | 10 | 2 |

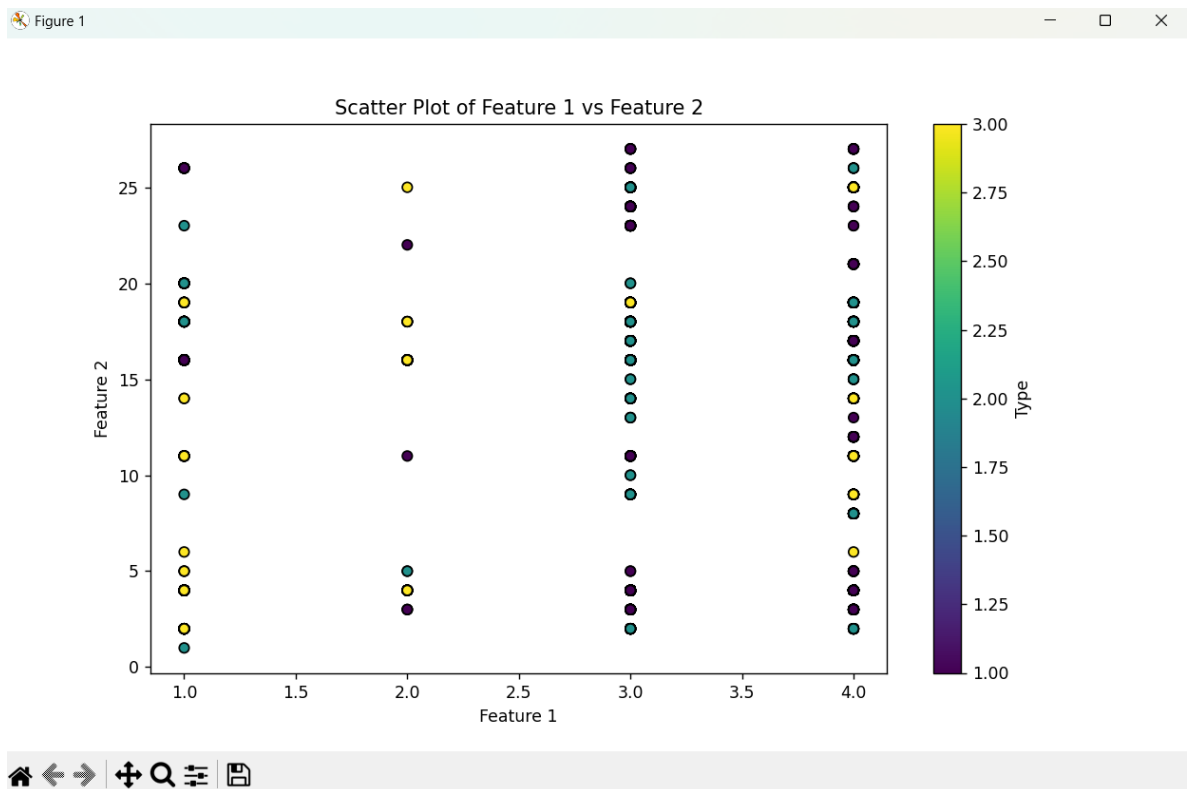
Tỉ lệ chính xác của mô hình: 0.68

Hình 4.4. Kết quả chạy SVM

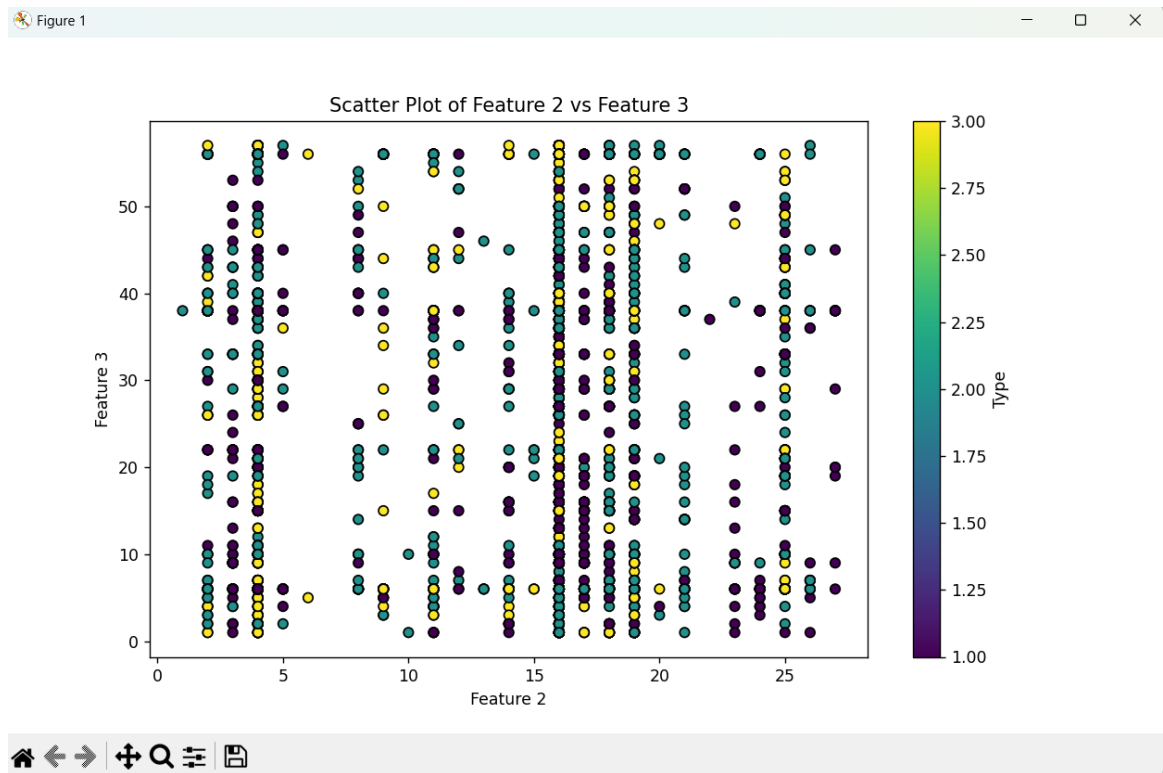
– Mô hình KNN:



Hình 4.5. Sơ đồ ROC

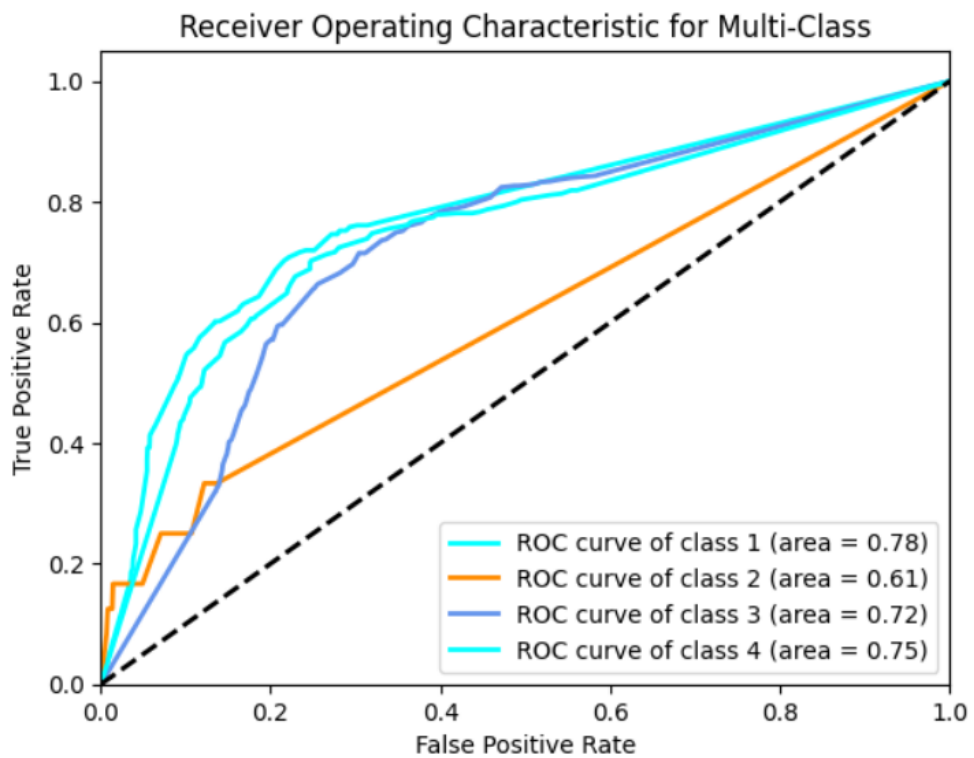


Hình 4.6. Hình Scatter Piot of Fearute 1 vs Fearute 2

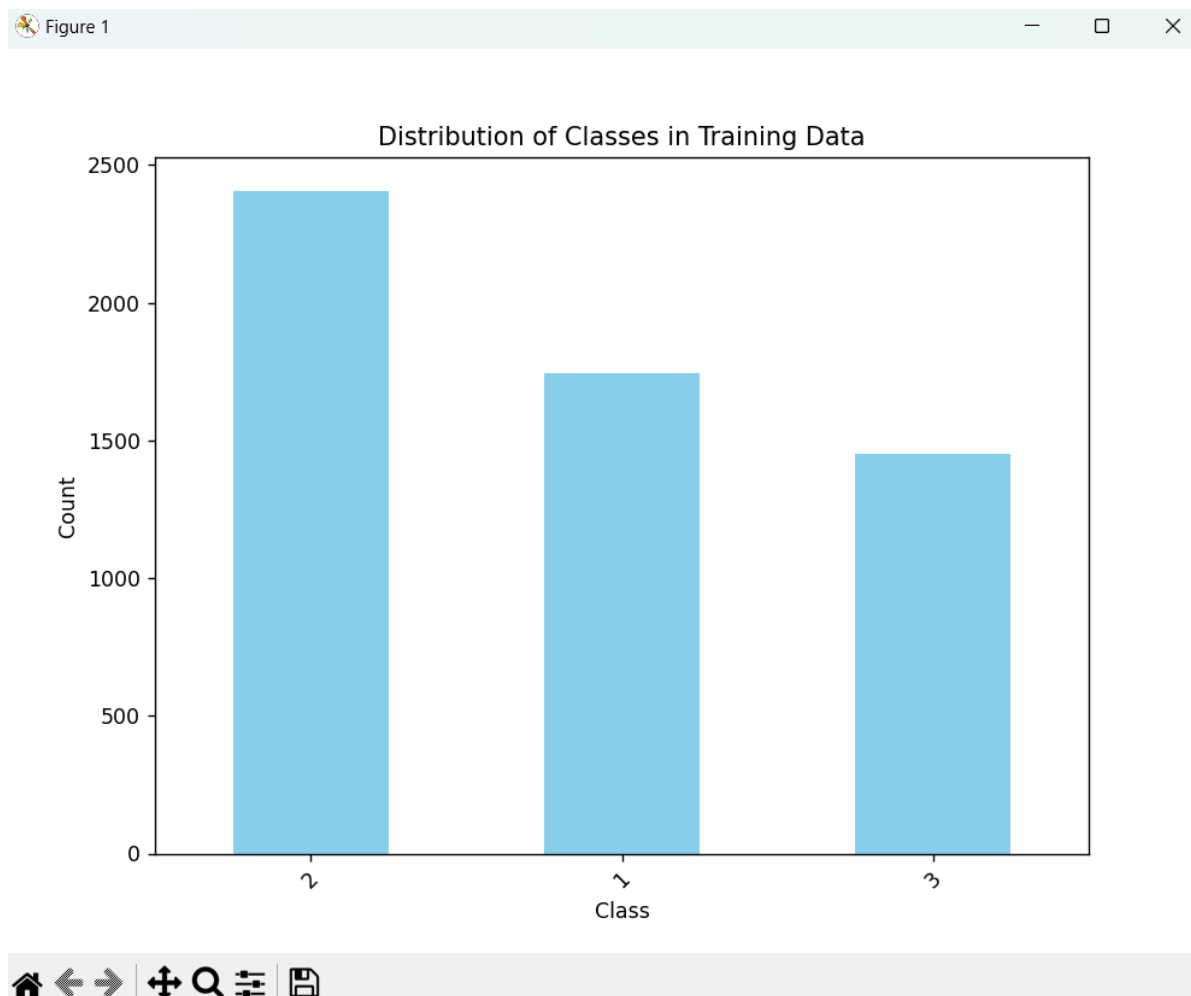


Hình 4.7. Hình Scatter Plot of Feature 2 vs Feature 3

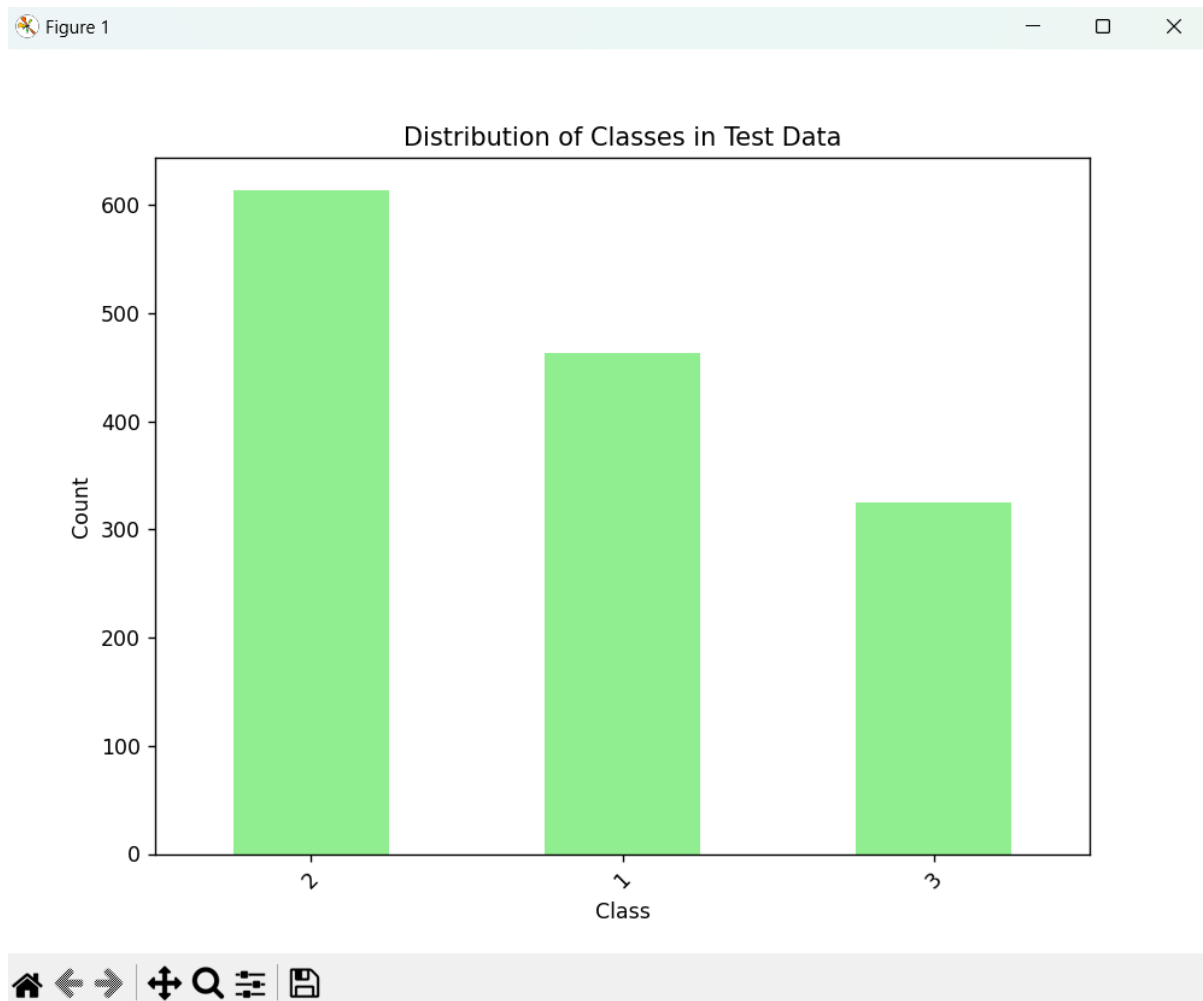
– Mô hình Decision Tree



Hình 4.8. Sơ đồ ROC Decision Tree



Hình 4.9. Hình Distribution of Classes in Training Data



Hình 4.10. Hình Distribution of Classes in Test Data

4.2. Mô hình học máy phù hợp

- So sánh SVM, KNN và Decision Tree

Bảng 4.1: So sánh SVM, KNN và Decision Tree

| | SVM | KNN | Descesion Tree |
|----------|------|------|----------------|
| Accuracy | 0.68 | 0.84 | 0.82 |
| F1 Score | 0.85 | 0,75 | 0.80 |
| Run time | 20 | 5 | 7 |

Lựa chọn mô hình học máy

Để lựa chọn mô hình học máy phù hợp, cần so sánh các chỉ số quan trọng của các mô hình khác nhau. Dựa trên các tiêu chí về độ chính xác, F1 Score và thời gian chạy, ta có thể thấy rằng mô hình SVM có độ chính xác là 0.68 và F1 Score cao nhất là 0.85, cho thấy SVM có sự cân bằng tốt giữa Precision và Recall.

Tuy nhiên, thời gian chạy của SVM là 10, dài nhất trong ba mô hình. Mô hình KNN có độ chính xác thấp nhất là 0.80 và F1 Score thấp nhất là 0.75, nhưng lại có thời gian chạy ngắn nhất với 5. Trong khi đó, mô hình Decision Tree có độ chính xác 0.82 và F1 Score 0.80, cao hơn KNN nhưng thấp hơn SVM, và thời gian chạy là 7, nằm giữa KNN và SVM.

Nếu độ chính xác và F1 Score là yếu tố quan trọng nhất, SVM là lựa chọn tốt nhất vì nó có các chỉ số cao nhất, ưu tiên độ chính xác và F1 Score em lựa chọn SVM.

4.3. Đánh giá và kết quả

Về trang web của em đã cung cấp nhiều chức năng tiện ích cho người dùng và quản lý. Người dùng có thể đăng ký tài khoản bằng cách điền vào biểu mẫu, đăng nhập bằng email và mật khẩu để truy cập trang web chính. Quản lý có thể thực hiện nhiều thao tác như thêm, sửa đổi, xóa sản phẩm, màu sắc và kích cỡ. Người dùng có thể xem chi tiết sản phẩm, bao gồm hình ảnh, mô tả, giá cả, màu sắc, kích cỡ và số lượng còn lại. Hệ thống đã tích hợp chức năng chatbox tư vấn, giúp người dùng dễ dàng chọn lọc những sản phẩm mà họ muốn.

Em đã sử dụng hệ thống ứng dụng các mô hình học máy để xây dựng chức năng chatbox, giúp cung cấp hỗ trợ người dùng một cách tự động. Nhờ vào việc sử dụng mô hình SVM, chatbox có thể phân loại và xử lý các yêu cầu tư vấn với độ chính xác lên đến 0,68, cải thiện hiệu quả tương tác và hỗ trợ trực tuyến cho người dùng. Thông qua đó, người sử dụng có thể nhận được phản hồi nhanh chóng và chính xác liên quan đến sản phẩm. Chatbox không chỉ giúp tiết kiệm thời gian mà còn nâng cao trải nghiệm người dùng bằng cách cung cấp thông tin.

Khi phát triển website tự động gợi ý và tư vấn khách hàng cho cửa hàng quần áo, em đã có một số hạn chế liên quan đến thời gian và kỹ thuật. Về mặt thời gian, quá trình phát triển hệ thống phức tạp này đòi hỏi nhiều thời gian để hoàn thiện, cần nâng cấp khả năng tích hợp hệ thống tự động gợi ý với các hệ thống hiện có như hệ thống thanh toán, cũng như thời gian để đào tạo các mô hình học máy với dữ liệu thực tế. Về mặt kỹ thuật, tích hợp hệ thống tự động gợi ý với các phần khác của website để phát triển như quản lý kho và thanh toán. Khả năng mở rộng hệ thống để đáp ứng nhu cầu ngày càng tăng của người dùng cũng là một yếu tố quan trọng, đòi hỏi thiết kế linh hoạt và dễ mở rộng.

Trong tương lai, em có thể nâng cấp website tự động gợi ý và tư vấn khách hàng cho cửa hàng quần áo nhiều chức năng để đáp ứng nhu cầu. Đầu tiên các chatbox sẽ thông minh hơn sẽ hỗ trợ khách hàng ngay lập tức, đồng thời bổ sung chức năng chat video hoặc thoại với nhân viên tư vấn để tăng cường tương tác cá nhân. Tiếp theo, nâng cấp hệ thống đánh giá và phản hồi thông minh sẽ cho phép khách hàng dễ dàng đánh giá sản phẩm và dịch vụ, đồng thời sử dụng AI để phân tích phản hồi và cải thiện chất lượng dịch vụ. Mở rộng chương trình khách hàng thân thiết sẽ được nâng cấp với các ưu đãi cá nhân hóa và khuyến mãi độc quyền. Mở thêm hệ thống thanh toán sẽ hỗ trợ nhiều phương thức thanh toán mới. Cuối cùng, việc sử dụng AI để phân tích dữ liệu lớn và dự báo các xu hướng thời trang mới sẽ giúp cửa hàng cung cấp các sản phẩm phù hợp với xu hướng và nhu cầu thị trường.

KẾT LUẬN

Em đã bắt đầu bằng việc phân tích kỹ lưỡng yêu cầu và nhu cầu của khách hàng khi mua sắm trực tuyến. Điều này đã thúc đẩy em phát triển một ứng dụng có khả năng gợi ý sản phẩm và tư vấn cá nhân hóa dựa trên sở thích và yêu cầu của từng khách hàng.

Trong quá trình thiết kế hệ thống, em đã chú trọng vào việc xây dựng một giao diện người dùng thân thiện và dễ sử dụng. Đồng thời, em đã sử dụng các công nghệ và công cụ hiện đại nhất để đảm bảo tính ổn định và hiệu suất của hệ thống, từ việc sử dụng ngôn ngữ lập trình phổ biến như Python và Javascript.

Quy trình phát triển của em đã được thực hiện một cách có tổ chức và hiệu quả, bao gồm các giai đoạn từ lập kế hoạch, phát triển, kiểm thử đến triển khai. Bằng cách này, đã đảm bảo rằng ứng dụng được phát triển đúng hạn và đáp ứng được các yêu cầu.

Với những thành tựu này, trong tương lai có thể đề xuất các hướng phát triển tiếp theo cho ứng dụng, bao gồm việc tích hợp thêm tính năng mới như hệ thống đánh giá và nhận xét từ người dùng, mở rộng quy mô để phục vụ nhiều đối tượng khách hàng hơn, và liên tục cải thiện tính năng gợi ý và tư vấn.

TÀI LIỆU THAM KHẢO

TIẾNG VIỆT

- [1]. H. C. Trung, “Công nghệ Máy học (Machine Learning) là gì?,” 25/8/2023. [Trực tuyến]. Available: <https://www.ibm.com/cloud/learn/machine-learning>

WEBSITE

- [2]. <https://stackoverflow.com/>
- [3]. https://ptithcm.edu.vn/wpcontent/uploads/2023/07/2020_HTTT_LeHoangBao_TTLV.pdf
- [4]. <https://stringee.com/vi/blog/post/reactJS-la-gi>
- [5]. <https://vnetwork.vn/news/what-is-an-api-and-how-to-secure-it-effectively/>
- [6]. <https://fptshop.com.vn/tin-tuc/danh-gia/nestjs-la-gi-184643>
- [7]. <https://media.neliti.com/media/publications/451728-applying-decision-tree-technique-in-data-71044134.pdf>
- [8]. <https://trituenhantao.io/kien-thuc/decision-tree/>
- [9]. <https://viblo.asia/p/gioi-thieu-ve-support-vector-machine-svm-6J3ZgPVEImB>
- [10]. <https://vietnambiz.vn/thuat-toan-k-lang-gieng-gan-nhat-k-nearest-neighbor-knn-la-gi-2020022911113334.htm>
- [11]. [https://jgac.vn/journal/article/view/344#:~:text=Random%20forest%20%C3%A0%20m%E1%BB%99t%20ph%C6%B0%C6%A1ng,quy%E1%BA%BFt%20%C4%91%E1%BB%8Bnh%20\(Decision%20tree\).](https://jgac.vn/journal/article/view/344#:~:text=Random%20forest%20%C3%A0%20m%E1%BB%99t%20ph%C6%B0%C6%A1ng,quy%E1%BA%BFt%20%C4%91%E1%BB%8Bnh%20(Decision%20tree).)
- [12]. <https://www.kaggle.com/datasets/justjess678/womens-clothes/data>