# LỜI CẢM ƠN

Lời đầu tiên, tôi xin được gửi lời cảm ơn chân thành tới các thầy cô giáo trong trường Đại học Công nghiệp Hà Nội cũng như các thầy cô trong khoa Công nghệ Thông tin đã truyền dạy cho tôi những kiến thức và kinh nghiệm quý báu trong quá trình học tập 4 năm tại trường. Trong quá trình học tập tại trường đã giúp tôi trưởng thành lên rất nhiều. Tôi đã có một môi trường học tập lành mạnh, trau dồi được nhiều kiến thức, cách học tập, cách sống, cách nghiên cứu và làm việc hiệu quả. Những điều đó góp phần rất lớn cho tôi trên con đường học tập và làm việc trong con đường đời tôi đang bước.

Tôi xin gửi lời cảm ơn tới **TS. Nguyễn Mạnh Cường**  – Giảng viên bộ môn, khoa Công nghệ Thông tin đã hết lòng hướng dẫn, chỉ bảo, giúp đỡ tôi trong quá trình làm đề tài thực tập tốt nghiệp này.

Tôi cũng muốn gửi lời cảm ơn tới tập thể lớp Kỹ thuật phần mềm 2-K8 đã tạo môi trường học tập, trao đổi kiến thức, tạo điều kiện phát triển các kỹ năng cho mỗi thành viên.

Tôi cũng xin gửi tới lời cảm ơn chân thành tới gia đình, bạn bè đã quan tâm, động viên, giúp đỡ tôi trong quá trình học tập, làm thực tập tốt nghiệp.

# LỜI NÓI ĐẦU

Cùng với sự phát triển của internet đã tạo nên sự bùng nổ thông tin. Đặc biệt cùng với sự tham gia của thương mại điện tử, số lượng thông tin trao đổi trên internet đang tăng theo cấp số nhân. Ngày nay, bạn có thể tìm bất cứ thông tin hoặc sản phẩm mình cần thông qua internet. Điều này khá thuận lợi cho khách hàng lựa chọn, tuy nhiên đó cũng chính là khó khăn. Với mỗi từ khóa được tìm kiếm, khách hàng vẫn phải lựa chọn trong danh sách hàng trăm thậm chí hàng nghìn sản phẩm có vẻ phù hợp. Do đó khách hàng cần có một sự trợ giúp nào đó để có thể tìm kiếm được sản phẩm phù hợp với mình nhất. Do đó hệ tư vấn ra đời. Ngày nay, hệ tư vấn không chỉ phát triển trong thương mại, nó còn được nghiên cứu và áp dụng trong các hệ giải trí như tư vấn phim, tư vấn âm nhạc, tư vấn sách….

Bài báo cáo này là kết quả sau một thời gian nghiên cứu tìm hiểu và tiến hành hiện thực một hệ tư vấn cho một website thương mại điện tử. Trong bài báo cáo, tôi sẽ trình bày các cách xây dựng hệ tư vấn, các thuật toán sử dụng để đưa ra các tư vấn, các thông tin được xây dựng. Tôi mong nhận được sự góp ý xây dựng và hoàn thiện bài báo cáo cũng như hệ thống tư vấn.

Tôi xin chân thành cảm ơn./

# TÓM TẮT NỘI DUNG

Nội dung của bài báo cáo tôi sẽ trình bày qua 3 phần tách biệt:

Phần một được giành để giới thiệu một số khái niệm cơ bản về các vấn đề tôi sẽ nghiên cứu và hiện thực. Đó là:

* Thương mại điện tử.
* Bán lẻ trực tuyến.
* Hệ tư vấn.
* Kỹ thuật được sử dụng.

Phần hai trình bày nghiên cứu khảo sát website EzzShop. Trong phần này, tôi sẽ trình bày cấu trúc hoạt động của hệ thống web site hiện tại và tất nhiên là chưa hề có sự kết hợp của bất kỳ hệ tư vấn nào. Các nội dung gồm:

* Cơ cấu hoạt động của hệ thống website.
* Các hoạt động của khách hàng.
* Cấu trúc của website.
* Các chức năng của hệ thống.
* Giao diện hệ thống.
* Yêu cầu đối với hệ tư vấn.

Phần ba trình bày chi tiết hơn về hệ tư vấn mà tôi sẽ hiện thực sau một thời gian được thầy Nguyễn Mạnh Cường hướng dẫn. Nội dung của phần này:

* Phương pháp xây dựng hệ tư vấn.
* Phân loại khách hàng.
* Phân loại các gợi ý/ tư vấn.
* Các thuật toán sử dụng để đưa ra các yêu cầu.

Trên đây là nội dung tóm tắt về các phần tôi sẽ trình bày trong bài báo cáo. Để chi tiết hơn về các nội dung đã nói, chúng ta sẽ đi vào chi tiết của nội dung trong từng phần riêng.

# **MỤC LỤC**

[LỜI CẢM ƠN 1](#_Toc481536868)

[LỜI NÓI ĐẦU 2](#_Toc481536869)

[TÓM TẮT NỘI DUNG 3](#_Toc481536870)

[MỤC LỤC 4](#_Toc481536871)

[DANH MỤC HÌNH ẢNH 7](#_Toc481536872)

[DANH MỤC BẢNG 7](#_Toc481536873)

[CHƯƠNG 1. MÔ TẢ BÀI TOÁN 8](#_Toc481536874)

[**1.1.** **Thương mại điện tử** 8](#_Toc481536875)

[**1.1.1.** **Khái niệm** 8](#_Toc481536876)

[**1.1.2.** **Phân loại** 8](#_Toc481536877)

[**1.1.3.** **Bán lẻ trực tuyến – E-tailing** 9](#_Toc481536878)

[**1.2.** **Hệ tư vấn – Recommendation System** 9](#_Toc481536879)

[**1.2.1.** **Khái niệm** 9](#_Toc481536880)

[**1.2.2.** **Kỹ thuật xây dựng hệ tư vấn** 9](#_Toc481536881)

[**1.2.3.** **Lọc cộng tác – Collaborative Filtering** 10](#_Toc481536882)

[CHƯƠNG 2. NGHIÊN CỨU, KHẢO SÁT HỆ THỐNG WEBSITE BÁN HÀNG EZZSHOP 11](#_Toc481536883)

[**2.1.** **Mở đầu** 11](#_Toc481536884)

[**2.2.** **Cấu trúc hoạt động của hệ thống thực** 11](#_Toc481536885)

[**2.2.1.** **Cơ cấu hoạt động** 11](#_Toc481536886)

[**2.2.2.** **Các hoạt động của khách hàng tại cửa hàng** 12](#_Toc481536887)

[**2.2.3.** **Các hoạt động định kỳ của cửa hàng** 12](#_Toc481536888)

[**2.3.** **Cấu trúc website EzzShop** 13](#_Toc481536889)

[**2.3.1.** **Giao diện chính của website** 13](#_Toc481536890)

[**2.3.2.** **Giao diện trang** 14](#_Toc481536891)

[**2.4.** **Các chức năng của hệ thống website EzzShop** 15](#_Toc481536892)

[**2.4.1.** **Các chức năng với người sử dụng là khách hàng** 15](#_Toc481536893)

[**2.4.2.** **Giao diện đối với người sử dụng là quản trị website/cửa hàng** 20](#_Toc481536894)

[**2.5.** **Yêu cầu đối với hệ tư vấn** 23](#_Toc481536895)

[**2.5.1.** **Vấn đề hiện tại** 23](#_Toc481536896)

[**2.5.2.** **Yêu cầu** 23](#_Toc481536897)

[CHƯƠNG 3. XÂY DỰNG HỆ TƯ VẤN 24](#_Toc481536898)

[**2.1.** **Phân loại loại khách hàng** 24](#_Toc481536899)

[**2.1.1.** **New client** 24](#_Toc481536900)

[**2.1.2.** **Tranditional client** 24](#_Toc481536901)

[**2.2.** **Phân loại các tư vấn gợi ý** 24](#_Toc481536902)

[**2.3.1.** **Purchase Items - Hàng từng được mua** 25](#_Toc481536903)

[**2.3.2.** **Not-Purchased Items – Hàng chưa từng được mua** 26](#_Toc481536904)

[**2.3.3.** **New Items – Hàng mới** 26](#_Toc481536905)

[**2.4.** **Cơ sở dữ liệu hiện tại** 27](#_Toc481536906)

[**2.4.1.** **Bảng thông tin về khách hàng** 27](#_Toc481536907)

[**2.4.2.** **Bảng thông tin hàng hóa** 28](#_Toc481536908)

[**2.4.3.** **Bảng thông tin về giao dịch** 30](#_Toc481536909)

[**2.4.4.** **Bảng thông tin chi tiết giao dịch** 31](#_Toc481536910)

[**2.5.** **Thuật toán gợi ý** 32](#_Toc481536911)

[**2.5.1.** **Ma trận Rate – Rate Matrix** 32](#_Toc481536912)

[**2.5.2.** **Best-Selling Items cho New clients – Các mặt hàng bán chạy nhất** 35](#_Toc481536913)

[**2.5.3.** **Best-Selling Items cho khác hàng truyền thống** 36](#_Toc481536914)

[**2.5.4.** **Purchased Items – Hàng từng được mua** 38](#_Toc481536915)

[**2.5.5.** **Not-Purchased Items – Hàng chưa từng được mua** 40](#_Toc481536916)

[**2.5.6.** **New Items – Hàng mới cho khách hàng truyền thống** 43](#_Toc481536917)

[**2.5.7.** **New Items – Hàng mới cho khách hàng mới** 44](#_Toc481536918)

[KẾT LUẬN 45](#_Toc481536919)

[**Những việc đã làm được** 45](#_Toc481536920)

[**Những việc chưa làm được** 45](#_Toc481536921)

[**Hướng phát triển** 45](#_Toc481536922)

# DANH MỤC HÌNH ẢNH

[Hình 1. Giao diện chính đối với khách hàng. 13](#_Toc481536923)

[Hình 2. Sản phẩm theo chủng loại 14](#_Toc481536924)

[Hình 3. Giao dilện Adlmin 14](#_Toc481536925)

[Hình 4. Trang chủ. 15](#_Toc481536926)

[Hình 5. Thông tin sản phẩm thuộc loại Jewels. 16](#_Toc481536927)

[Hình 6. Giao diện tìm kiếm hàng hóa 17](#_Toc481536928)

[Hình 7. Sản phẩm hiển thị theo loại sản phẩm. 17](#_Toc481536929)

[Hình 8. Thông tin các mặt hàng đã lựa chọn. 18](#_Toc481536930)

[Hình 9. Giao diện Checkout 19](#_Toc481536931)

[Hình 10. Danh sách hàng hóa 20](#_Toc481536932)

[Hình 11. Giao diện thêm mới hàng hóa 20](#_Toc481536933)

[Hình 12. Danh sách nhà cung cấp 21](#_Toc481536934)

[Hình 13. Thêm mới nhà cung cấp. 21](#_Toc481536935)

[Hình 14. Thống kê theo Nhà cung cấp 22](#_Toc481536936)

[Hình 15. Thống kê theo Sản phẩm 22](#_Toc481536937)

# DANH MỤC BẢNG

[Bảng 1 Ma trận Rate 35](#_Toc481535607)

# CHƯƠNG 1. MÔ TẢ BÀI TOÁN

## **Thương mại điện tử**

### **Khái niệm**

Thương mại điện tử, hay còn gọi là e-commerce, e-comm hay EC, là sự mua bán sản phẩm hay dịch vụ trên các hệ thống điện tử như Internet và các mạng máy tính. Thương mại điện tử dựa trên một số công nghệ như chuyển tiền điện tử, quản lý chuỗi dây chuyền cung ứng, tiếp thị Internet, quá trình giao dịch trực tuyến, trao đổi dữ liệu điện tử (EDI), các hệ thống quản lý hàng tồn kho, và các hệ thống tự động thu thập dữ liệu. Thương mại điện tử hiện đại thường sử dụng mạng World Wide Web là một điểm ít nhất phải có trong chu trình giao dịch, mặc dù nó có thể bao gồm một phạm vi lớn hơn về mặt công nghệ như email, các thiết bị di động cũng như điện thoại.

Thương mại điện tử thông thường được xem ở các khía cạnh của kinh doanh điện tử (e-business). Nó cũng bao gồm việc trao đổi dữ liệu tạo điều kiện thuận lợi cho các nguồn tài chính và các khía cạnh thanh toán của việc giao dịch kinh doanh.

### **Phân loại**

E-commerce có thể được chia ra thành:

* Edo-tailing (bán lẻ trực tuyến) hoặc "cửa hàng ảo" trên trang web với các danh mục trực tuyến, đôi khi được gom thành các "trung tâm mua sắm ảo".
* Trao đổi dữ liệu điện tử (EDI), trao đổi dữ liệu giữa Doanh nghiệp với Doanh nghiệp
* Email, fax và cách sử dụng chúng như là phương tiện cho việc tiếp cận và thiếp lập mối quan hệ với khách hàng (ví dụ như bản tin - newsletters)
* Việc mua và bán giữa Doanh nghiệp với Doanh nghiệp
* Bảo mật các giao dịch kinh doanh

Và trong tất cả các phân loại thương mại điện đó, chúng ta sẽ tiến hành xây dựng một hệ thống hỗ trợ cho E-tailing – Bán lẻ trực tuyến. Và đây có lẽ cũng là thể hiện nổi bật nhất của khái niệm thương mại điện tử - E Commerce.

### **Bán lẻ trực tuyến – E-tailing**

Kinh doanh trên mạng là mua bán hàng hoá và dịch vụ trên Internet. Thương mại điện tử có thể được chia thành: E-tailing “cửa hàng ảo” trên các trang web với danh mục sản phẩm trực tuyến, đôi khi tập hợp thành một “trung tâm mua sắm ảo”. E-tailing là một nơi mua sắm bán lẻ trực tiếp, liên tục 24 giờ và có khả năng tương tác và cung cấp thông tin về sản phẩm và đặt hàng với khách hàng tiềm năng.

Các hệ thống bán hàng trực tuyến đã tạo nhiều điều kiện thuận lợi để người mua có thể tiếp cận nhiều mặt hàng cùng lúc. Tuy nhiên, việc trình bày và trang trí quá nhiều các mặt hàng trên trang web đã gây ra không ít khó khăn cho người mua. Họ khó có thể chọn ra cho mình một sản phẩm ưng ý nhất. Để khách hàng có thể đến và mua được một sản phẩm ưng ý thì một lời tư vấn, một sự trợ giúp là rất quan trọng. Trong phương thức bán hàng truyền thống những lời tư vấn như thế từ một người bán hàng sẽ tạo ra một lợi thế rất lớn cho cửa hàng. Do đó, để phương thức bán hàng qua mạng thực sự phát triển thì bên cạnh các lợi thế vốn có của mình việc có thêm một “người trợ giúp” là hết sức cần thiết.

## **Hệ tư vấn – Recommendation System**

### **Khái niệm**

Hệ tư vấn(Recommender system) – là một thành phần trong hệ thống thông tin. Mục đích của nó là hỗ trợ người dùng tìm kiếm được đúng thông tin cần thiết.

Một hệ thống gợi ý tốt có thể đóng vai trò như một người trung gian hỗ trợ khách hàng đưa ra các quyết định mua hàng đúng đắn. Bằng cách xác định mục đích và nhu cầu của khách hàng, hệ thống có thể đưa ra một tập các gợi ý giúp cho người mua dễ dàng chọn lựa sản phẩm yêu thích hơn. Qua đó hiệu suất của việc mua bán hàng trực tuyến được tăng cao một cách đáng kể.

### **Kỹ thuật xây dựng hệ tư vấn**

Một hệ tư vấn có thể được cài đặt theo nhiều kỹ thuật khác nhau. Tuy nhiên, hai kỹ thuật lọc thông tin được sử dụng nhiều nhất đó là lọc dựa trên nội dung và lọc cộng tác.

Lọc dựa trên nội dung – Content Based Filtering (CBF) – là phép lọc dựa trên việc so sánh nội dung các item(sản phẩm), nếu người dùng nào sử dụng các item có nội dung gần giống với nội dung của sản phẩm mới này thì sản phẩm mới này được tư vấn cho người dùng đó.

Lọc cộng tác – Collaborative Filtering (CF) là một kỹ thuật mạnh và nó đã được áp dụng khá thành công trong nhiều hệ tư vấn. Thực chất, lọc cộng tác là một hình thức tư vấn tự động bằng cách dựa trên sự tương tự giữa những người dùng hoặc giữa những sản phẩm trong hệ thống và đưa ra dự đoán sự quan tâm của người dùng tới một sản phẩm, hoặc đưa ra gợi ý một sản phẩm mới cho người dùng nào đó.

Tóm lại, việc nghiên cứu xây dựng các hệ tư vấn hiện nay là cần thiết và khá triển vọng. Đã có khá nhiều những áp dụng hệ tư vấn vào trong các hệ thống như eBay, Amazon, Youtube, …

Kỹ thuật sẽ được sử dụng trong bài báo cáo này chính là Lọc cộng tác (CF). Phần sau sẽ nêu chi tiết hơn về lọc cộng tác và các thuật toán, kỹ thuật được sử dụng.

### **Lọc cộng tác – Collaborative Filtering**

Trong cuộc sống hằng ngày, mọi người thường tin vào những lời giới thiệu từ những người khác thông qua lời nói, thư từ văn bản, các nguồn tin thu được trên các phương tiện thông tin đại chúng. Về bản chất, lọc cộng tác cũng chính là hình thức tư vấn như trên, tuy nhiên, việc này được thực hiện tự động bởi các máy tính. Giả thuyết của lọc cộng tác là: “Nếu người dùng u và u' đánh giá cho n sản phẩm tương tự nhau, hoặc có hành vi tương tự nhau (như: xem, mua, nghe…) thì họ sẽ có các đánh giá tương tự nhau đối với các sản phẩm khác”.

Trong bài báo cáo này, tôi sẽ giới thiệu kỹ thuật lọc cộng tác trong hệ thống gợi ý, cài đặt và kiểm nghiệm kỹ thuật này. Tôi sẽ tiến hành nghiên cứu một hệ thống bán hàng trực tuyến, từ đó tích hợp kỹ thuật gợi ý vào hệ thống để tư vấn cho khách hàng những sản phẩm họ có thể yêu thích nhất. Việc tích hợp thành công này sẽ mở ra một hướng mới cho các hệ thống bán hàng trực tuyến hiện nay - đa phần chưa cá nhân hóa cho từng người dùng. Tuy vậy, việc cài đặt một hệ thống thực sự có tích hợp giải thuật gợi ý vào trong đó thì vẫn còn là lĩnh vực mới mẻ.

# CHƯƠNG 2. NGHIÊN CỨU, KHẢO SÁT HỆ THỐNG WEBSITE BÁN HÀNG EZZSHOP

## **Mở đầu**

Hệ thống cần xây dựng nhằm phục vụ cho các đối tượng là khách hàng thân thiết, vãng lai, khách hàng mới của cửa hàng.

Hệ thống được xây dựng nhằm mục đích:

* là nơi phục vụ nhu cầu mua sắm chăn ga gối đệm của khách hàng.
* là nơi tham khảo thông tin về giá cả, mẫu mã, chủng loại của hàng hóa mà khách hàng quan tâm.
* là nơi cửa hàng đưa ra những thông tin về khuyến mãi, thông tin cửa hàng, các chế độ đãi ngộ của cửa hàng, …

## **Cấu trúc hoạt động của hệ thống thực**

### **Cơ cấu hoạt động**

Khách hàng biết đến cửa hàng qua các kênh thông tin chung như:

* Truyền miệng.
* Quen biết, người thân.
* Nhận diện bằng mắt thường qua vị trí của cửa hàng.
* Quảng cáo, tờ rơi.

Việc mua bán trao đổi được thực hiện ngay tại cửa hàng. Nhân viên hoặc chủ cửa hàng phải trực tiếp tư vấn khách hàng để đảm bảo khách hàng tìm được món đồ mình thích.

### **Các hoạt động của khách hàng tại cửa hàng**

* Khách hàng biết đến cửa hàng qua các kênh thông tin bên trên.
* Khách hàng quan tâm đến một mặt hàng, cho rằng cửa hàng có sản phẩm.
* Khách hàng tới cửa hàng tham khảo thêm thông tin (nếu có hàng).
* Khách hàng mua hàng nếu cảm thấy phù hợp.
* Cửa hàng thực hiện chuyển hàng về cho khách hàng khi đã thanh toán thành công.

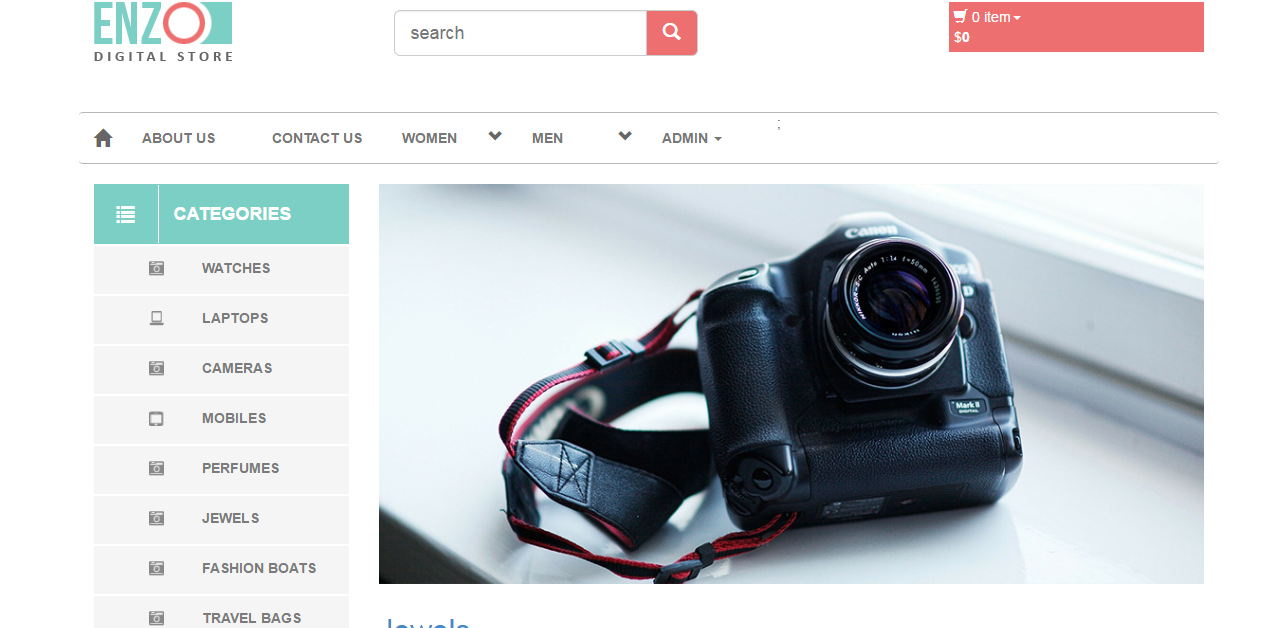
### **Các hoạt động định kỳ của cửa hàng**

* Kiểm kê hàng hóa.
* Nhập hàng mới.
* Bổ sung thương hiệu mới.

## **Cấu trúc website EzzShop**

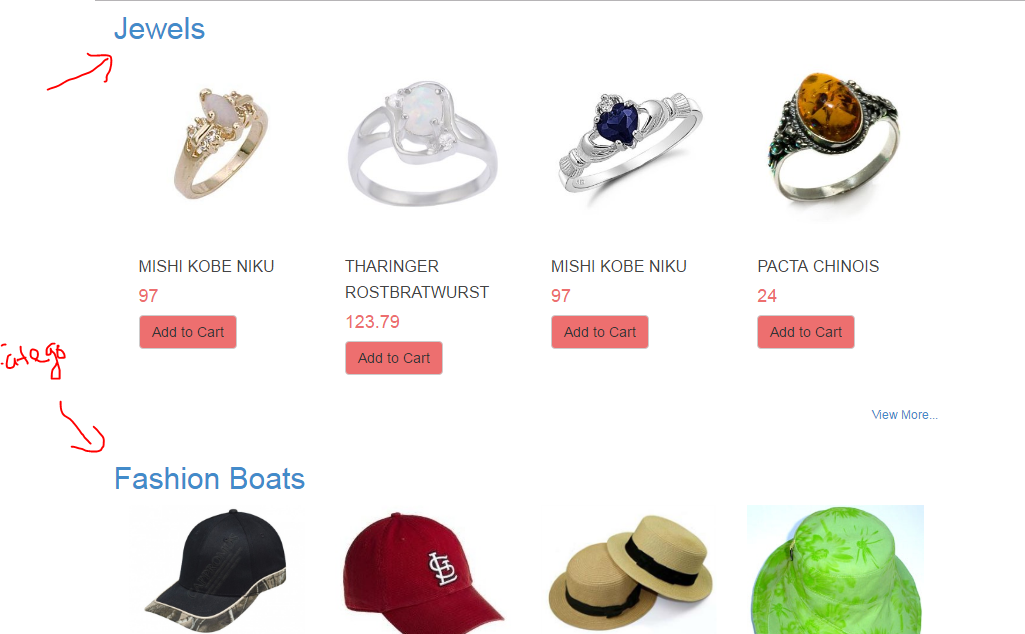
Cấu trúc website gồm 2 phần:

### **Giao diện chính của website**



Hình 1. Giao diện chính đối với khách hàng.

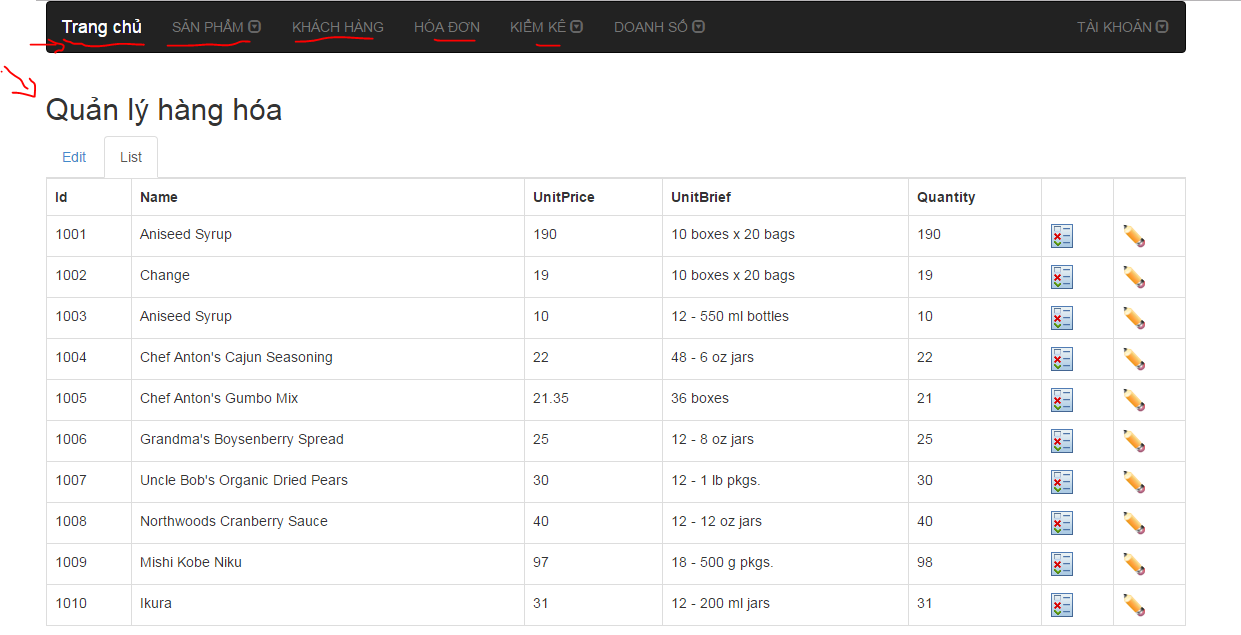
Phần phía bên tay trái là danh mục các loại sản phẩm mà cửa hàng kinh doanh. Khách hàng có thể lựa chọn loại hàng và hệ thống sẽ hiển thị thông tin chi tiết phía bên tay phải:



Hình 2. Sản phẩm theo chủng loại

### **Giao diện trang**

Cung cấp các chức năng cho người quản trị website/cửa hàng nhằm quản lý các thông tin về sản phẩm chủng loại, thống kê, báo cáo.



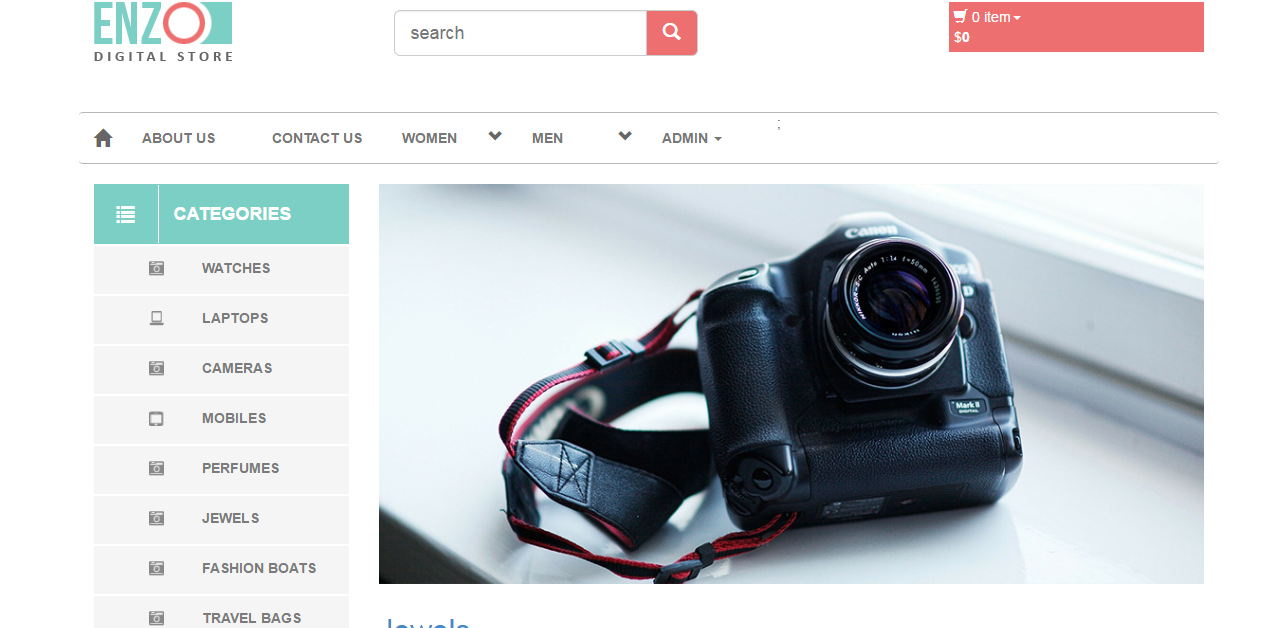
Hình 3. Giao dilện Adlmin

## **Các chức năng của hệ thống website EzzShop**

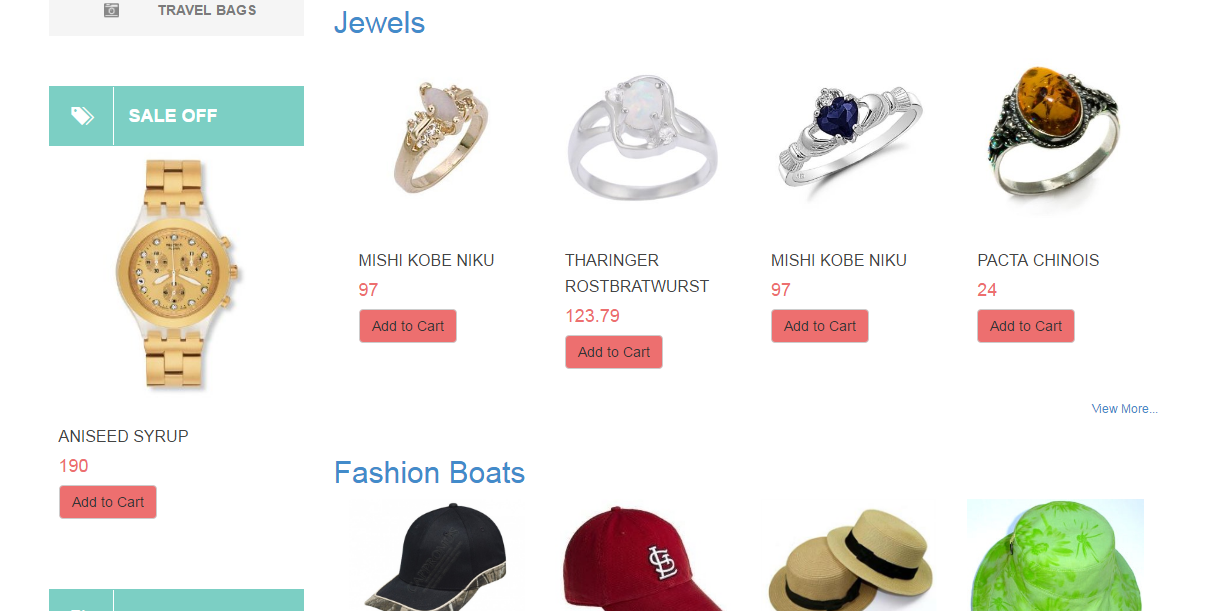
### **Các chức năng với người sử dụng là khách hàng**

#### Xem thông tin hàng hóa

Khách hàng có thể tham khảo thông tin hàng hóa ngay tại trang chủ của website:



Hình 4. Trang chủ.

Trang web hiển thị các sản phẩm theo từng danh mục loại hàng hóa mà cửa hàng đang có. Khách hàng sẽ lướt qua toàn bộ các sản phẩm và tìm hiểu thêm thông tin về giá, chất lượng, nguồn gốc, …

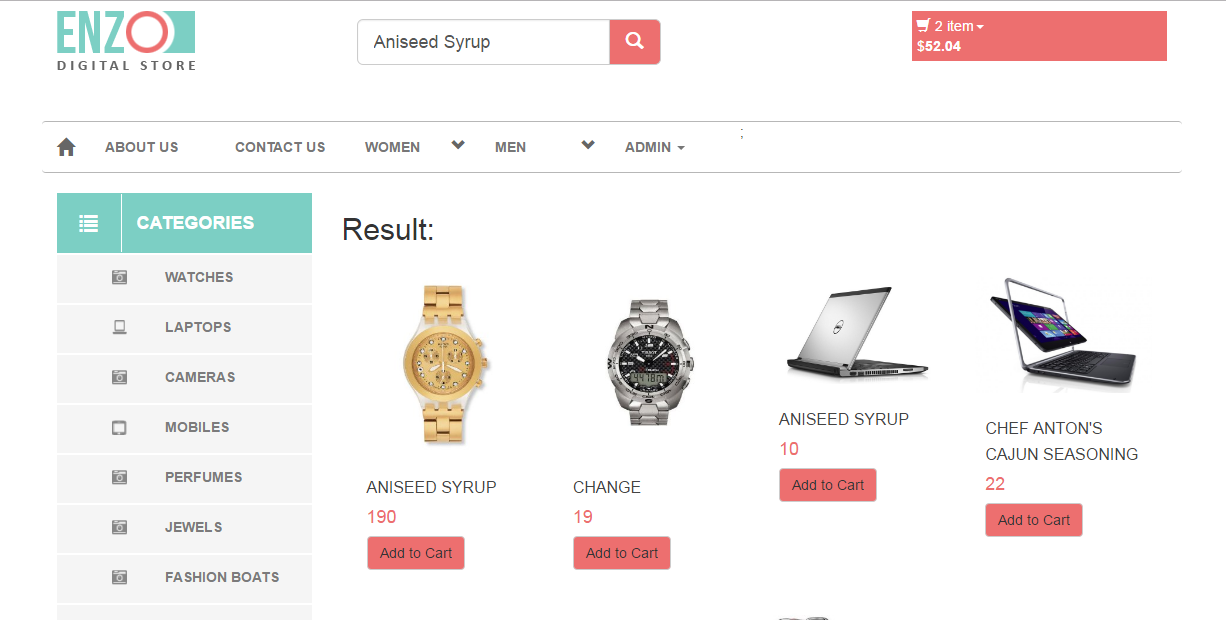
Hình 5. Thông tin sản phẩm thuộc loại Jewels.

Thông tin về hàng hóa chi tiết về hàng hóa được hiển thị làm sao cho ngắn gọn, xúc tích, đầy đủ.

#### Tìm kiếm hàng hóa quan tâm

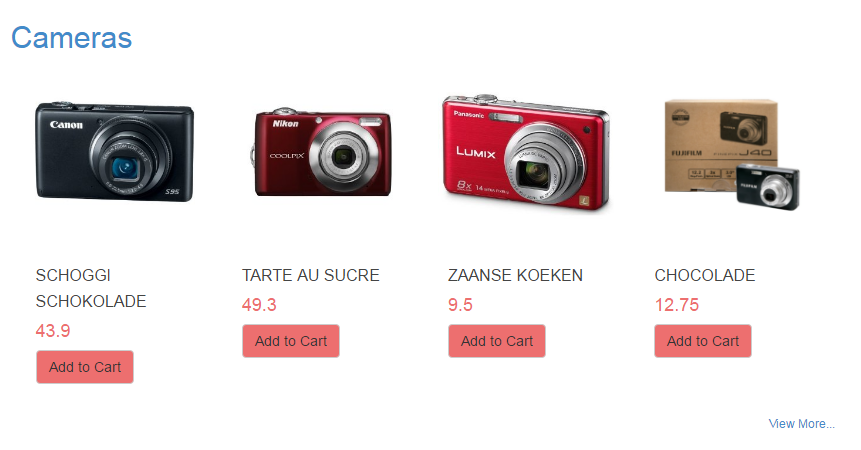
Khách hàng cần tìm chính xác thứ hàng hóa mà khách hàng mong muốn. Hỗ trợ tìm kiếm theo tên sản phẩm, loại sản phẩm, danh mục, chất liệu, …

Với khách hàng tìm kiếm sản phẩm theo tên sản phẩm hoặc tên thương hiệu, tên loại hàng hóa thì có thể tìm kiếm bằng cách gõ thông tin vào mục tìm kiếm :



Hình 6. Giao diện tìm kiếm hàng hóa

Ngoài ra có thể lựa chọn hàng hóa dựa theo các sản phẩm mà hệ thống liệt kê trên màn hình.

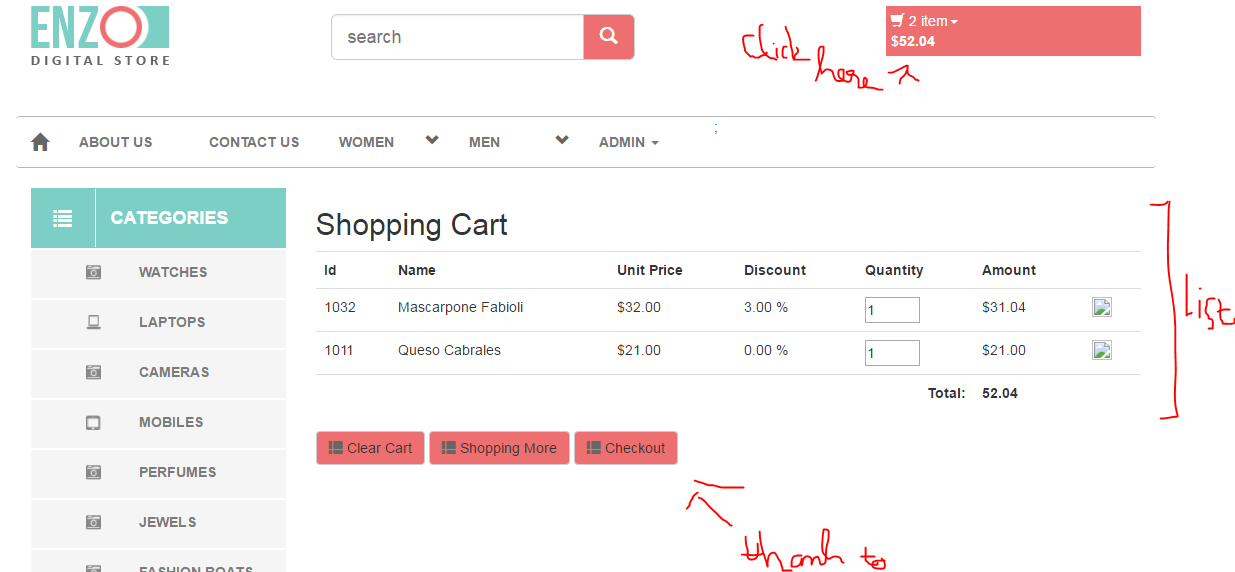


Hình 7. Sản phẩm hiển thị theo loại sản phẩm.

#### Đặt hàng

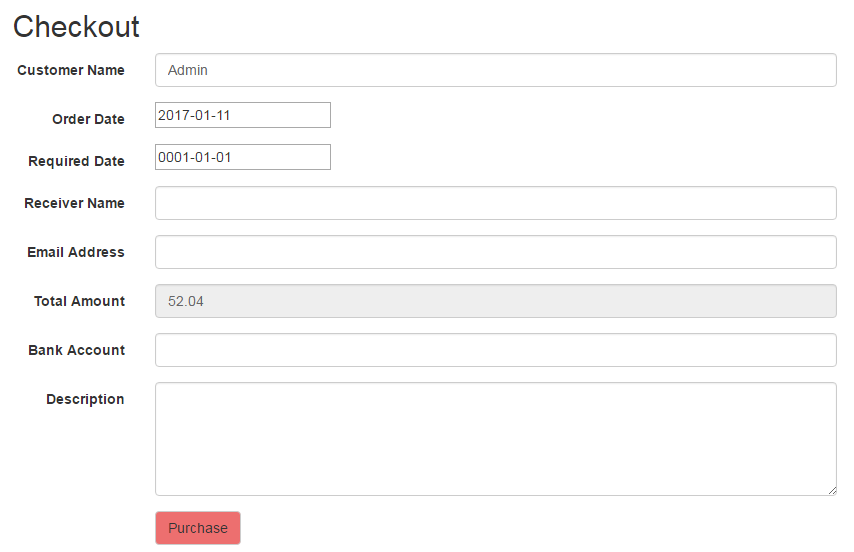
Sau khi tham khảo thông tin về giá cả cũng như về chất lượng, nếu khách hàng muốn mua hàng thì website cần hỗ trợ việc đặt hàng và phương thức, cách thức thanh toán.

Xem thông tin các mặt hàng đã lựa chọn:



Hình 8. Thông tin các mặt hàng đã lựa chọn.

Sau khi lựa chọn được mặt hàng ưng ý hoàn toàn có thể chọn *Checkout* để thanh toán:



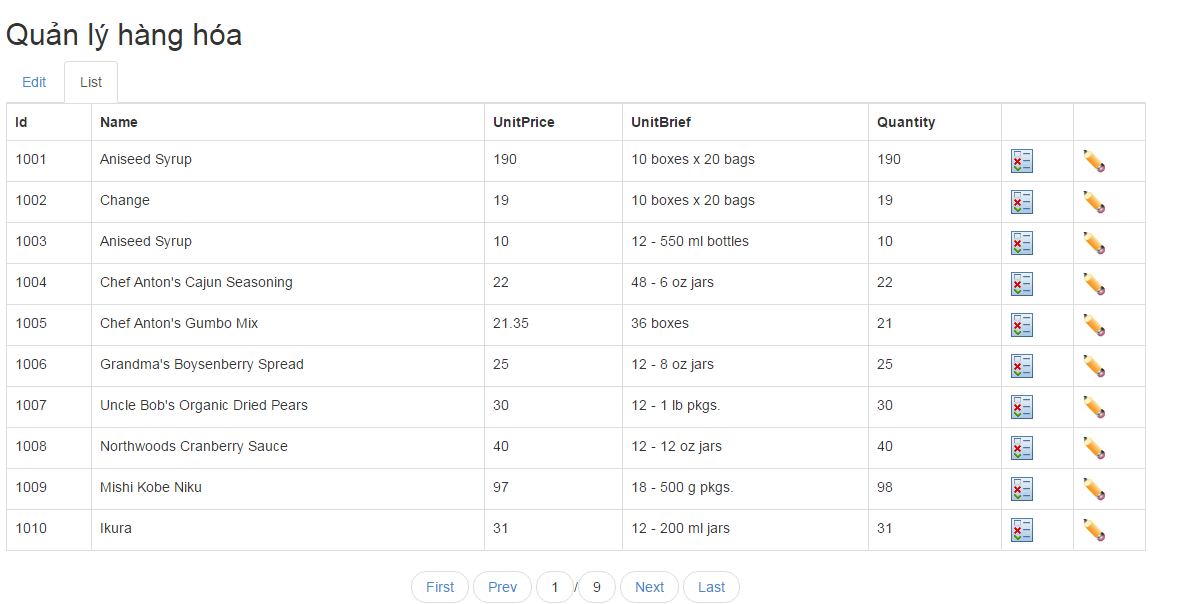
Hình 9. Giao diện Checkout

hoặc *Clear* để hủy toàn bộ và tiếp tục tham quan cửa hàng.

### **Giao diện đối với người sử dụng là quản trị website/cửa hàng**

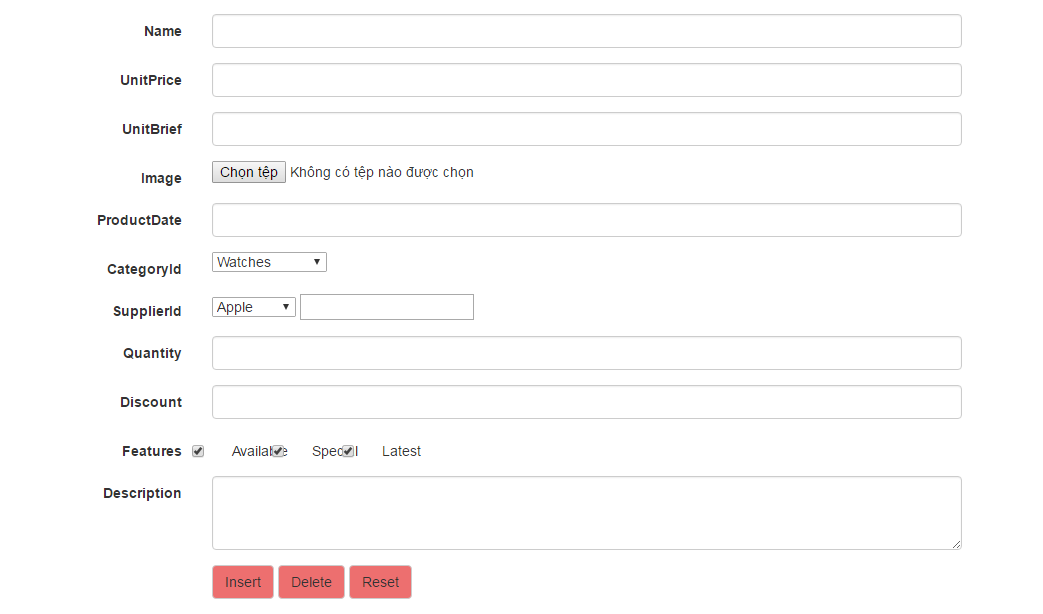
#### Quản lý hàng hóa

Thống kê danh sách hàng hóa có trong kho:



Hình 10. Danh sách hàng hóa

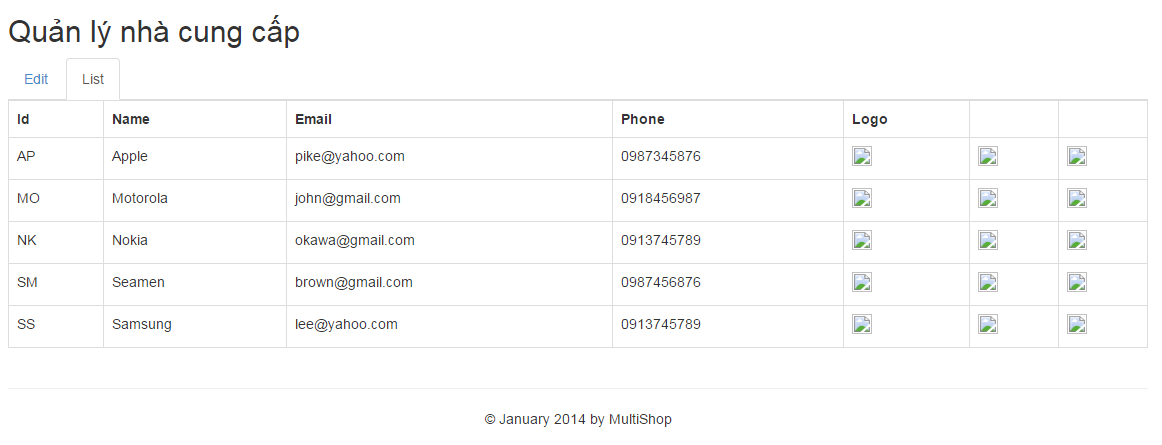
Hệ thống cũng hỗ trợ việc thêm mới hàng hóa vào kho:



Hình 11. Giao diện thêm mới hàng hóa

#### Thống kê nhà cung cấp

Xem thông tin của các nhà cung cấp qua giao diện:



Hình 12. Danh sách nhà cung cấp

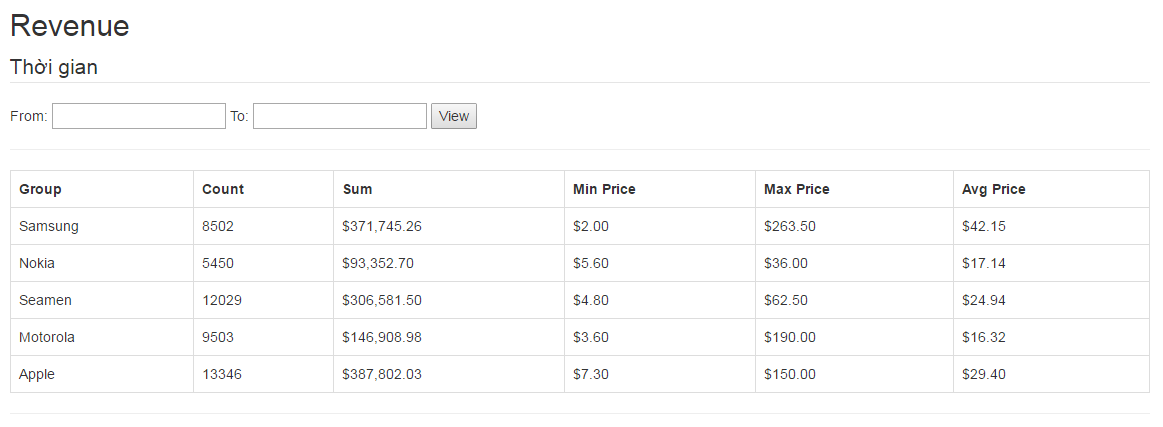
Việc thêm mới nhà cung cấp cũng được hỗ trợ trên hệ thống :



Hình 13. Thêm mới nhà cung cấp.

#### Thống kê doanh số

Việc thống kê doanh số theo Sản phẩm và theo Nhà cung cấp.



Hình 14. Thống kê theo Nhà cung cấp

Và theo Sản phẩm:



Hình 15. Thống kê theo Sản phẩm

## **Yêu cầu đối với hệ tư vấn**

### **Vấn đề hiện tại**

Hệ thống hiện tại của website chưa hề có một thuật toán tư vấn lựa chọn hàng hóa nào cả. Kết quả lựa chọn hàng hóa là như nhau đối với mọi loại hàng hóa, mọi khách hàng.

Để nâng cao doanh số và doanh thu cũng như đáp ứng tốt nhu cầu của khách hàng, hệ thống hiện tại cần một hệ thống hỗ trợ ra quyết định. Hệ gợi ý này sẽ giúp việc hiển thị hàng hóa phù hợp với các loại khách hàng của cửa hàng.

### **Yêu cầu**

Hệ hộ trợ ra quyết định cần:

* Tính toán và dự đoán khả năng người dùng sẽ thích sản phẩm hay thông tin nào đó. Từ đó đưa ra những gợi ý về sản phẩm hay thông tin có thể phù hợp với người dùng nhất.
* Đưa ra các quyết định phù hợp với mỗi loại khách hàng của hệ thống(New clients và Tranditional clients).
* Hàng hóa đưa ra cần có hiệu suất đạt kết quả tích cực cao. Bán được đúng hàng cho đúng loại khách hàng.

# CHƯƠNG 3. XÂY DỰNG HỆ TƯ VẤN

## **Phân loại loại khách hàng**

### **New client**

Khách hàng mới được hiểu đơn giản là những khách hàng chưa từng mua mặt hàng nào của cửa hàng. Với nhóm khách hàng này, người dùng có thể đã có hoặc chưa từng có tài khoản trong hệ thống. Vì chưa từng có số liệu về các lần mua hàng, nên việc tính toán để đưa ra các gợi ý thường rất chung chung.

### **Tranditional client**

Ngược lại với khách hàng mới – new clients, nhóm khách hàng truyền thống gồm các khách hàng đã từng thực hiện ít nhất 1 giao dịch trên hệ thống. Các khách hàng trong nhóm này đã có thông tin giao dịch nên việc tính toán để đưa ra gợi ý sẽ được dựa vào thông tin của toàn nhóm hiện tại.

## **Phân loại các tư vấn gợi ý**

* Có 4 loại gợi ý cho các loại khách hàng:
* R1: Best-Selling Items
* R2: Not-Purchased Items
* R3: Purchased Items
* R4: New Items
  1. **Best-Selling items – Bán chạy nhất**
* Là các mặt hàng được gợi ý do hệ thống đánh giá là đang được lựa chọn nhiều nhất trên nhóm khách hàng tương ứng trong thời gian gần đây.
* Các mặt hàng gợi ý được tính dựa trên mặt hàng có số lượt mua nhiều nhất trong số đó có lượng mua nhiều nhất trong thời gian gần nhất.
* Phù hợp với loại khách hàng New clients do hệ thống chưa hề có thông tin gì thêm để đưa ra các recommend R2, 3. Purchased Items

### **Purchase Items - Hàng từng được mua**

* Là các mặt hàng được hệ thống đánh giá là có khả năng cao nhất khách hàng sẽ mua lại sản phẩm đó.
* Phù hợp với đối tượng là các khách hàng đã có thông tin trong hệ thống.
* Các dữ liệu sẽ sử dụng:
  + Transactional Data
  + User Demographic Data
  + Item DemoGraphic Data
* Các gợi ý được tính dựa trên loại hàng:
  + Mua gần đây nhất.
  + Tổng số lượng mua nhiều nhất
  + Số lượt mua nhiều nhất

### **Not-Purchased Items – Hàng chưa từng được mua**

* Là các mặt hàng hiện có nhưng chưa bao giờ được khách hàng lựa chọn. Hệ thống sẽ gợi ý mặt hàng nào sẽ có khả năng cao nhất đáp ứng yêu cầu của khách hàng.
* Phù hợp với đối tượng là các khách hàng đã có thông tin trong hệ thống.
* Các dữ liệu sẽ sử dụng:
  + Transactional Data
  + User Demographic Data
  + Item DemoGraphic Data

### **New Items – Hàng mới**

* Hàng mới nhất, hàng vừa nhập dựa theo ngày sản xuất , ngày nhập.
* Hàng được gợi ý là hàng chưa từng được khách hàng mua.
* Thông tin sử dụng: Items Demographic Data
* Các mặt hàng được gợi ý sẽ là các mặt hàng mới nhất hiện có trong kho.

## **Cơ sở dữ liệu hiện tại**

Trên đây là danh sách cấu trúc các bảng dữ liệu là đầu vào cho hệ tư vấn, theo thứ tự:

* Bảng **Customers** – Khách hàng
* Bảng **Products** – Hàng hóa
* Bảng **Orders** – Giao dịch
* Bảng **OrderDetails** – Chi tiết giao dịch

### **Bảng thông tin về khách hàng**

[dbo].[Customers](

//Mã khách hàng

[Id] [nvarchar](20) NOT NULL

//Mật khẩu

[Password] [nvarchar](50) NOT NULL,

//Họ tên đầy đủ

[Fullname] [nvarchar](50) NOT NULL,

//Email

[Email] [nvarchar](50) NOT NULL,

//Ảnh đại diện, mặc định là Photo.gif

[Photo] [nvarchar](50) NOT NULL

CONSTRAINT [DF\_Customers\_Photo] DEFAULT (N'Photo.gif')

//Trạng thái kích hoạt

[Activated] [bit] NOT NULL CONSTRAINT [DF\_Customers\_Active] DEFAULT ((0)),

//Khóa chính

CONSTRAINT [PK\_Customers] PRIMARY KEY CLUSTERED

### **Bảng thông tin hàng hóa**

[dbo].[Products](

//Khóa chính

[Id] [int] IDENTITY(1,1) NOT NULL,

//Tên hàng hóa

[Name] [nvarchar](40) NOT NULL,

//Mô tả ngắn

[UnitBrief] [nvarchar](50) NOT NULL,

//Giá cả với mặc định là 0

[UnitPrice] [float] NOT NULL CONSTRAINT [DF\_Products\_UnitPrice] DEFAULT ((0)),

//Ảnh đại diện cho hàng, mặc định là photo.gif

[Image] [nvarchar](50) NOT NULL CONSTRAINT [DF\_Products\_Image] DEFAULT (N'Product.gif'),

//Ngày sản xuất, mặc định là ngày hiện tại

[ProductDate] [datetime] NOT NULL CONSTRAINT [DF\_Products\_ProductDate] DEFAULT (getdate()),

//Trạng thái của hàng hóa, mặc định là có

[Available] [bit] NOT NULL CONSTRAINT [DF\_Products\_Available] DEFAULT ((1)),

//Mô tả chi tiết

[Description] [nvarchar](2000) NULL,

//Mã loại hàng

[CategoryId] [int] NOT NULL,

//Mã nhà sản xuất, mặc định là chưa rõ

[SupplierId] [nvarchar](50) NOT NULL CONSTRAINT [DF\_Products\_SupplierId] DEFAULT (N'Chưa rõ'),

//Số lượn hàng, mặc định là 0

[Quantity] [int] NOT NULL DEFAULT ((0)),

//Giảm giá, mặc định là 0%

[Discount] [float] NOT NULL CONSTRAINT [DF\_Products\_Discount] DEFAULT (0),

//Khuyến mãi được biệt/ hàng hiếm/…

//Mặc định là hàng bình thường

[Special] [bit] NOT NULL CONSTRAINT [DF\_Products\_Special] DEFAULT ((0)),

//Hàng mới nhất hay không, mặc định là không

[Latest] [bit] NOT NULL CONSTRAINT [DF\_Products\_New] DEFAULT ((0)),

//Lượt xem

[Views] [int] NOT NULL CONSTRAINT [DF\_Products\_Views] DEFAULT ((0)),

// Khóa chính

CONSTRAINT [PK\_Products] PRIMARY KEY CLUSTERED

### **Bảng thông tin về giao dịch**

[dbo].[Orders](

//Mã giao dịch

[Id] [int] IDENTITY(1,1) NOT NULL,

//Mã khách hàng

[CustomerId] [nvarchar](20) NOT NULL,

//Ngày đặt hàng, mặc định là ngày hiện tại

[OrderDate] [datetime] NOT NULL CONSTRAINT [DF\_Orders\_OrderDate] DEFAULT (getdate()),

//Ngày giao hàng yêu cầu

[RequireDate] [datetime] NOT NULL,

//Người nhận

[Receiver] [nvarchar](50) NOT NULL,

//Địa chỉ

[Address] [nvarchar](60) NOT NULL,

//Mô tả thêm

[Description] [nvarchar](1000) NULL,

//Số tiền, mặc định là 0

[Amount] [float] NOT NULL CONSTRAINT [DF\_\_Orders\_\_Amount\_\_381A47C8] DEFAULT ((0)),

//Khóa chính

CONSTRAINT [PK\_Orders] PRIMARY KEY CLUSTERED

### **Bảng thông tin chi tiết giao dịch**

[dbo].[OrderDetails](

//Mã chi tiết, tự động tăng

[Id] [int] IDENTITY(1,1) NOT NULL,

//Mã giao dịch

[OrderId] [int] NOT NULL,

//Mã hàng hóa

[ProductId] [int] NOT NULL,

// Giá tiền hàng hóa, mặc định là 0

[UnitPrice] [float] NOT NULL CONSTRAINT [DF\_Order\_Details\_UnitPrice] DEFAULT ((0)),

//Số lượng hàng, mặc định là 1

[Quantity] [int] NOT NULL CONSTRAINT [DF\_Order\_Details\_Quantity] DEFAULT ((1)),

//Giảm giá, mặc định là 0%

[Discount] [float] NOT NULL CONSTRAINT [DF\_OrderDetails\_Discount] DEFAULT ((0)),

//Khóa chính

CONSTRAINT [PK\_OrderDetails] PRIMARY KEY

## **Thuật toán gợi ý**

Trong hệ thống thuật toán gợi ý dưới đây, chúng ta biểu diễn các đánh giá của người dùng cho các sản phẩm qua ma trận gồm một tập người dùng U và tập sản phẩm I. Tập I có thể rất lớn (trăm ngàn tới hàng triệu…). Chúng ta dùng các ký tự để phân biệt người dùng và các sản phẩm: *u, u'* (đại diện cho người dùng) *- i, i'* (đại diện cho các sản phẩm). Ký hiệu rui để chỉ mức độ thích của người dùng *u* cho một sản phẩm *i* nào đó, chẳng hạn giá trị này trong khoảng từ 1 (không thích) đến 5 (thích nhất) đối với dữ liệu bán hàng.

Từ dữ liệu giao dịch trong hệ thống, chúng ta sẽ tính hành tính ra ma trận Rate của hệ thống.

### **Ma trận Rate – Rate Matrix**

Ma trận Rate là ma trận hai chiều chứa các giá trị ưu tiên của hệ thống. Mỗi dòng trong ma trận đại diện cho tập các giá trị ưu tiên của khách hàng U với toàn bộ sản phẩm của cửa hàng. Mỗi cột thể hiện tập các giá trị ưu tiên của một mặt hàng – sản phẩm I, trên toàn bộ khách hàng của hệ thống.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *I*  U | *I0* | *I1* | *I2* | *I3* | *…* | *In* |
| *U0* | **13** | ? | **3** | **2** | … | **4** |
| *U1* | **12** | **15** | ? | **3** | **…** | ? |
| *U2* | ? | **12** | **5** | ? | … | **1** |
| *U3* | **50** | ? | **4** | **2** | **…** | ? |
| *U4* | ? | **51** | ? | ? | … | **5** |
| *….* | … | … | … | … | … | … |
| *Um* | ? | **2** | **5** | ? | **…** | **2** |

Hình trên là ví dụ minh họa về một ma trận rate. Trong ma trận ta thấy được số độ ưu tiên của từng khách hàng trên từng mặt hàng. Ví dụ, khách hàng U0, có đã từng mua I1,I2,I3. U0 không hề có thông tin giao dịch với các mặt hàng: I1,… . Mặt hàng I0 có lịch sử giao dịch với U0,1,3 nhưng lại không hề có lần nào được mua bởi U2,4.

Dựa vào mảng ma trận Rate ta có thể thấy được thông tin về độ yêu của 1 khách hàng bất kỳ trên toàn bộ các mặt hàng của cửa hàng. Tương tự, chúng ta có thể thấy được rằng một mặt hàng trong hệ thống đã từng được khách hàng mua hay chưa hoặc có được ưa chuộng hay không?

Tóm lại, ta gọi tập R là tập các giá trị đánh giá của người dùng trên các sản phẩm. Dữ liệu được biểu diễn trong một ma trận 2 chiều, trong đó mỗi dòng là một user, mỗi cột là một item và mỗi ô là đánh giá của user trên item tương ứng). Với mỗi người dùng u ∈ U, cần tìm mục i ∈ I (u chưa/xem/mua i) sao cho hàm *fui* đo độ phù hợp (xếp hạng/đánh giá) *rui* của người dùng u có giá trị trên i.

Kỹ thuật xây dựng ma trận:

* Ma trận được tính toán dựa trên lịch sử giao dịch của khách hàng.
* Các thông tin được sử dụng là số lượt mua và số lượng mua.
* Thuật toán:
  + Input: Thông tin giao dịch của hệ thống.
  + Output: Mảng ma trận Rate của hệ thống - R.
  + B1: Nhập thông tin đã giao dịch của khách hàng với từng mặt hàng tương ứng.
  + B2: Giá trị Rate được tính bằng công thức: Rui = Sui + Kui.
  + Trong đó:
    - Sui: là số lượng mua mặt hàng i của khách hàng u.
    - Kui: Là số lượt mua mặt hàng i của khách hàng u.
    - Rui: Là giá trị rate tại hàng u cột i trong ma trận Rate.
    - : là các hệ số, .

Sau quá trình xây dựng ma trận Rate, ta được kết quả là một mảng 2 chiều chứa các giá trị độ ưu tiên của từng mặt hàng tương ứng với từng khách hàng. Ví dụ:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *I*  U | *I0* | *I1* | *I2* | *I3* | *…* | *In* |
| *U0* | **1** | ? | **3** | **2** | … | **4** |
| *U1* | **1** | **5** | ? | **3** | **…** | ? |
| *U2* | ? | **2** | **5** | ? | … | **1** |
| *U3* | **5** | ? | **4** | **2** | **…** | ? |
| *U4* | ? | **5** | ? | ? | … | **5** |
| *….* | … | … | … | … | … | … |
| *Um* | ? | **2** | **5** | ? | **…** | **2** |

Bảng 1. Ma trận Rate

Với khách hàng Uo thuộc tập U, mặt hàng I0 có độ ưu tiên (yêu thích) là 1. tương ứng là I2,3 lần lượt là 3,2; mặt hàng I1 khách hàng chưa từng có lịch sử giao nên giá trị ưu tiên bằng 0. Nhìn vào đó ta cũng có thể đoán được, khách hàng này khá thích món đó I2 vì nó có độ ưu tiên lớn nhất.

Mặt khác, nhìn vào ma trận ta cũng có thể thấy được nhưng khách hàng nào có chung sở thích, từ đó, ta có thể gom nhóm các khách hàng để từ đó phân nhóm thành các nhóm khách hàng có chung sở thích.

### **Best-Selling Items cho New clients – Các mặt hàng bán chạy nhất**

Loại gợi ý này phục vụ cho các nhóm khách hàng có đặc điểm đó là:

* Khách hàng chưa hề có thông tin cá nhân trong hệ thống, khách vãng lai.
* Khách hàng đã có thông tin cơ bản nhưng chưa có thông tin giao dịch.

Trong phần này ta sẽ xây dựng thuật toán cho loại khách hàng chưa từng mua một mặt hàng trong hệ thống, chưa từng có thông tin giao dịch và chưa hề có thông tin của khách hàng. Việc tính toán sẽ được thực hiện trên toàn bộ hệ thống. Tất cả các giao dịch đã có trong hệ thống sẽ được sử dụng để tính ra mặt hàng nào đang bán chạy nhất và thu lại doanh thu lớn nhất cho cửa hàng.

Thuật toán được sử dụng như sau:

* Input: Thông tin giao dịch của toàn bộ hàng hóa trong hệ thống.
* Output: Các mặt hàng có số lượng bán nhiều nhất trong toàn hệ thống.
* B1: Nhập số lượng bán của các mặt hàng trong hệ thống Sij.
* B2: Tính tổng số lượng hàng đã bán theo từng loại hàng hóa trên toàn bộ hệ thống Ti.
* B3: Tính tổng tất cả số lượng của tất cả hàng hóa ta được thu được K.
* B4: Lấy tổng của từng loại hàng Ti chia cho tổng số lượng K ta thu được một hệ số xi(0≤x≤1).
* B5:Sắp xếp các hệ số ta được một mảng các giá trị giảm dần M.
* B6:Dựa vào mảng ta lấy được các mặt hàng cần gợi ý là các mặt hàng có chỉ số cao nhất.
* Ghi chú:
  + Ti: Tổng số lượng bán ra của mặt hàng I;
  + Si: Số lượng bán ra của mặt hàng i ở lượt thứ j.
  + K: Tổng toàn phần số lượng bán ra của hệ thống.
  + xi: Chỉ số của mặt hàng trên toàn hệ thống.
  + M: Mảng các chỉ số của các mặt hàng đã bán.

### **Best-Selling Items cho khác hàng truyền thống**

Với khách hàng đã có thông tin giao dịch trong hệ thống và ít nhất cũng đã mua 1 mặt hàng. Các mặt hàng cùng nhóm với mặt hàng đã mua sẽ được ưu tiên hơn để gợi ý đề xuất. Các gợi ý cũng sẽ được tính toán dựa trên nhóm tương ứng của khách hàng sau khi phân nhóm.

Thuật toán sẽ được sử dụng cho thuật toán này:

* B1: Nhập số lượng bán của các mặt hàng được bán trong nhóm Nz : Sij.
* B2: Tính tổng số lượng hàng đã bán theo từng loại hàng hóa trên toàn bộ nhóm Nz: Ti.
* B3: Tính tổng tất cả số lượng của tất cả hàng hóa ta được thu được Kz.
* B4: Lấy tổng của từng loại hàng Ti chia cho tổng số lượng Kz ta thu được một hệ số xi(0≤x≤1).
* B5: Duyệt qua tất cả nhóm mặt hàng mà khách hàng đã từng mua trong hệ thống. Nếu mặt hàng này cùng nhóm với một trong các nhóm đã mua thì độ ưu tiên được tăng thêm 10%.
* B6: Sắp xếp các hệ số ta được một mảng các giá trị của nhóm giảm dần Mz.
* B7:Dựa vào mảng ta lấy được các mặt hàng cần gợi ý cho khách hàng mới đã có tài khoản là các mặt hàng tương ứng có chỉ số cao nhất trong mảng.
* Ghi chú:
  + Ti: Tổng số lượng bán ra của mặt hàng I.
  + Sij: Số lượng bán ra của mặt hàng i ở lượt thứ j.
  + Kz: Tổng toàn phần số lượng bán ra của nhóm Nz.
  + xi: Chỉ số của mặt hàng trên toàn nhóm. Với mặt hàng thuộc nhóm liên quan tới mặt hàng đã mua: xi = xi \* 10%.
  + Mz: Mảng các chỉ số của các mặt hàng đã bán.
  + Nz: Nhóm khác hàng z, u .

Thuật toán đối với nhóm khách hàng này có khác với thuật toán phần 2.5.2. Đó là việc tính toán dựa trên lượng thông tin trong nhóm khách hàng tương ứng. Tiếp đến là các mặt hàng cùng nhóm với nhóm các mặt hàng đã được mua sẽ được ưu tiên hơn. Cụ thể trong thuật toán này, chúng ta đã cho ưu tiên hơn 10% so với giá trị gốc.

Trên đây là hai thuật toán được sử dụng để xuất ra gợi ý các mặt hàng bán chạy nhất cho hai nhóm khách hàng là khách hàng mới và khách hàng truyền thống. Phần tiếp theo chúng ta sẽ đi vào xây dựng 2 thuật toán dành riêng cho nhóm khách hàng truyền thống.

### **Purchased Items – Hàng từng được mua**

Với loại gợi ý này, các thông tin sẽ được sử dụng là các thông tin giao dịch đã có của khách hàng U trong khoảng thời gian 1 năm gần đây nhất. Dựa vào lịch sử giao dịch mua hàng gần nhất, chúng ta sẽ tính được tốc độ trung bình sử dụng mặt hàng theo ngày (AVG = Tổng số lượng mua trong 365 ngày chia cho 365, đơn vị items/day). Từ đó tính được trung bình số lượng hàng hóa sử dụng trong 1 ngày.

AVG = (item/day)

Danh sách gợi ý sẽ phụ thuộc vào độ chênh lệch DCL. Độ chênh lệch DCL là độ chênh lệch giữa số lượng sử dụng thực tế (SLTT) và số lượng sử dụng theo lý thuyết (SLLT). Số lượng thực tế - SLTT, là số lượng mặt hàng mà khách hàng đã mua gần đây nhất của mặt hàng I. Giả sử, cách đây 20 ngày khách hàng u’ mua một mặt hàng i’, số lượng mua là 20 sản phẩm. Vậy SLTT là 20. Số lượng lý thuyết – SLLT, là số luợng mặt hàng sẽ được dùng được tính dựa trên giá trị trung bình AVG. SLLT sẽ được tính bằng khoảng thời gian gần đầy nhất (tính theo ngày) mua hàng nhân với số lượng hàng dùng trên một ngày (AVG). Trong trường hợp trên, chúng ta sẽ tính được SLLT = 20 \*AVG. Dưới đây là công thức tính DCL:

**DCL = ABS(SLTT – SLLT)**

Việc đánh giá DCL để đưa ra gợi ý được trình bày như sau:

Nếu SLTT lớn hơn SLLT, tức là hiện tại sản phẩm vẫn đang còn thừa. Số chênh lệch càng lớn thì thừa càng nhiều và khách hàng chưa cần mặt hàng đó. Số chênh lệch càng nhỏ thì sản phẩm khách hàng đã sử dụng gần hết và sắp có nhu cầu mua thêm.

Nếu SLTT nhỏ hơn SLLT. Tức là sản phẩm đã không còn được khách hàng sử dụng hoặc vừa hết. Nếu độ chênh lệch càng lớn, sản phẩm đã không còn được dùng trong một khoảng thời gian dài và khách hàng đa phần không yêu thích và muốn sử dụng mặt hàng này nữa. Độ chênh lệch càng nhỏ, sản phẩm vừa mới được dùng hết, có thể giai đoạn này khách hàng đang rất cần mua sản phẩm để sử dụng. Sự yêu thích của khách hàng với mặt hàng khá cao.

Tóm lại, nếu DCL càng lớn thì độ yêu thích của khách hàng với mặt hàng càng nhỏ. Ngược lại, DCL càng nhỏ thì khách hàng đang rất cần mặt hàng này. Và dưới góc độ số học thì độ chênh lệch này càng gần 0 thì khách hàng càng cần mặt hàng này.

Thuật toán:

* Input: Thông tin giao dịch của khách hàng tương ứng trong vòng thời gian bằng 1 năm gần đây nhất.
* Output: Các mặt hàng để ra gợi ý cho khách hàng.
* B1: Tính tổng số lượng mặt hàng Ii trong 365 ngày gần nhất: Ti.
* B2: Tính trung bình tốc độ sử dụng mặt hàng trong 1 năm: Avgi = Ti / 365 (hàng/ngày)
* B3:Xác định số lượng hàng Ii được mua cách D ngày gần đây nhất: SLTTi
* B4: Dựa vào trung bình sử dụng ta tính được số lượng hàng sử dụng lý thuyết trong D ngày : SDLTi = Ai \* D.
* B5: Tính toán độ chênh lệch giữa SLTTi và SLLTi: DCLi = SLLTi - SLTTi.
* B6: Độ ưu tiên càng tiến gần về 0 thì độ ưu tiên càng cao.
* B7: Các mặt hàng có độ ưu tiên cao nhất sẽ là các mặt hàng được gợi ý cho khách hàng Ii.
* Chú thích:
  + Ii: Mặt hàng thứ i được mua trong hệ thống.
  + Ti: Tổng số lượng đã bán ra của mặt hàng Ii.
  + Ai: Tốc độ sử dụng trung bình mặt hàng i trong vòng 365 ngày gần nhất.
  + SLTTi: Số lượng đã mua thực tế.
  + SLLTi: Số lượng sử dụng theo lý thuyết.
  + DCLi : Độ chênh lệch giữa số lượng mua thực tế và số lượng lý thuyết.

### **Not-Purchased Items – Hàng chưa từng được mua**

Với loại gợi ý này, các thông tin được sử dụng sẽ vẫn là các thông tin giao dịch trong hệ thống, và đặc biệt đó là ma trận Rate trong phần **2.5.1.** Dựa vào ma trận Rate, thuật toán này cần điền các ô còn để trống (giá trị là ?) trong **Bảng 1-** các sản phẩm chưa từng có lịch sử giao dịch với khách hàng hiện tại. Để dễ dàng tính toán, chúng ta sẽ coi các mặt hàng trong ma trận là các vetor vô hướng. Tọa độ của chúng là giá trị rate tương ứng đối với từng khách hàng u (u). Vector p:

pi = ( pi1, pi2, pi3, p4, …, pin )

Từ đó, ta sẽ được danh sách các vector tương ứng với các sản phẩm đã mua trong nhóm khách hàng tương ứng (nếu có). Theo quy ước đó, chúng ta hoàn toàn có thể tính khoảng cách giữa các item chưa mua và các item đã mua dựa theo khoảng cách giữa các vector. Các khoảng cách sẽ được tính toán dựa vào công thức tính khoảng cách d giữa 2 mặt hàng pi  và qj là công thức khoảng cách khá nổi tiếng *Euclid.*

Công thức khoảng cách dpq giữa 2 vector p = ( p1, p2, p3, p4, .. pn ) và q = (q1, q2, q3, q4, .., qn ) được tính như sau:

**dpq =**

Từ danh sách khoảng cách đó, chúng ta sẽ lần lượt tính từng khoảng cách từ các vector sản phẩm đã mua tới các vector sản phẩm chưa mua. Từ danh sách khoảng cách đó, mặt hàng chưa mua nào có tổng khoảng cách từ các mặt hàng đã mua tới nhỏ nhất thì độ tương đồng thích của khách hàng đối với mặt hàng chưa mua đó là lớn nhất và được ưu tiên cao nhất. Mặt hàng có khoảng cách càng gần thì càng được ưu tiên gợi ý.

Thuật toán chi tiết như sau:

* Input : Ma trận Rate như được tính trong phần **2.5.1**.
* Output: Ma trận Rate mới sau khi đã điền đầy đủ các ô còn trống.
* Thuật toán:
  + B1: Lấy thông tin về mặt hàng đã mua qua ma trận Rate đã có, ta được tập Iđã mua (Idm).
  + B2: Lấy thông tin về mặt hàng chưa mua qua ma trận Rate đã có, ta được tập Cchưa mua (Ccm).
  + B3: Với từng khách hàng, chúng ta sẽ tính toán khoảng cách từ các mặt hàng đã mua tới các mặt hàng chưa mua: dI->C.
  + B4: Tính tổng khoảng cách từ tất cả mặt hàng đã mua trong Idm tới từng mặt hàng chưa mua trong Ccm, ta thu được tổng Sic tương ứng với mặt hàng chưa mua.
  + B5: Giá trị tổng khoảng cách càng nhỏ thì độ tương đồng giữa hai mặt hàng càng cao. Sắp xếp tăng dần các khoảng cách ta sẽ thu được thứ tự các mặt hàng ưu tiên.

Đã mua

Chưa mua

d1->4

d2->4

d1->5

d3->4

d3->5

d2->5

Hình 2 Thuật toán hàng chưa mua.

* + - Theo ví dụ trên ta có:

S4 = d1->4 + d2->4 + d3->4.

S5 = d1->5 + d2->5 + d3->5.

* + B6: Các mặt hàng được gợi theo theo thứ tư như mảng đã sắp xếp.
  + Chú thích:
    - di->j: khoảng cách giữa mặt hàng đã mua i tới mặt hàng chưa mua j của khách hàng u.
    - Sk: tổng khoảng cách từ các mặt hàng đã mua tới mặt hàng chưa mua k.

### **New Items – Hàng mới cho khách hàng truyền thống**

Việc tính toán hoàn toàn có thể dựa trên các thông tin giao dịch đã có vì nhóm khách hàng này chắc chắn từng có thông tin giao dịch trong hệ thống. Dữ liệu sẽ thu nhỏ lại theo nhóm khách hàng tương ứng và các nhóm mặt hàng có các mặt hàng từng mua.

Thuật toán sẽ thực hiện lọc ra danh sách các mặt hàng mới theo ngày sản xuất trong hệ thống. Mặt hàng trong danh sách nếu có cùng nhóm với một trong số nhóm các mặt hàng đã mua thì sẽ được cộng thêm điểm ưu tiên. Danh sách gợi ý cuối cùng là danh sách điểm ưu tiên của từng mặt hàng mới và được sắp xếp giảm dần. Danh sách này được sử dụng để xuất ra gợi ý cho khách hàng.

Thuật toán chi tiết như sau:

* Input : Ma trận Rate đã có như phần **2.5.1**.
* Output: Các mặt hàng cần đưa ra để gợi ý.
* B1: Lấy danh sách các mặt hàng đã được bán cho khách hàng u (Có trong ma trận rate) ta được tập Iu.
* B2: Lấy danh sách các nhóm mặt hàng tương ứng với các mặt hàng trong Iu.
* B3: Với nhóm mặt hàng tương ứng với các loại hàng trong Iu, các mặt hàng được gợi ý tương ứng là các mặt hàng mới nhập trong 1 tháng gần đây và chưa từng được bán.
* B4: Nếu chỉ số rate của mặt hàng i trong nhóm Iu  đối với khách hàng u càng cao thì nhóm mặt hàngtương ứng với i sẽ có độ ưu tiên gợi ý càng cao.
* B5: Các mặt hàng mới phù hợp sẽ được gợi ý qua 4 bước trên.
* Chú thích:
  + i: Mặt hàng i đã mua.
  + u: Khách hàng u đã mua mặt hàng i.
  + Iu: Nhóm mặt hàng tương ứng, sao cho i Iu.

### **New Items – Hàng mới cho khách hàng mới**

Vì khách hàng chưa từng có thông tin trong hệ thống nên mặt hàng được gợi ý sẽ chỉ cần dựa trên các thông tin về hàng hóa và giao dịch trong hệ thống. Các mặt hàng được gợi ý là các mặt hàng mới nhất và chưa từng được mua trên toàn bộ hệ thống, không hề có phân nhóm.

* Thuật toán:
  + Input: Thông tin về hàng hóa và giao dịch.
  + Output: Các mặt hàng mới nhất trên toàn hệ thống.
  + B1: Lấy danh sách các mặt hàng mới nhập trong thời gian gần đây.
  + B2: Loại trừ khỏi danh sách các mặt hàng đã được bán ra.
  + B3: Danh sách thu được là danh sách các mặt hàng mới nhất chưa từng được bán cho bất kỳ khách hàng nào.
  + B4: Hàng được gợi ý là các mặt hàng có trong danh sách này.

# KẾT LUẬN

## **Những việc đã làm được**

Sau quá trình nghiên cứu và hiện thực một hệ tư vấn cho trực tiếp một website thương mại điện tử. Em thấy được độ quan trọng của hệ tư vấn trong hệ thống. Hệ tư vấn tốt sẽ hỗ trợ việc gợi ý đề xuất gần đúng mặt hàng mà khách hàng đang cần tìm, mặt hàng đang quan tâm.

Trong thời gian tám tuần của quá trình nghiên cứu, tôi đã có thể biết được một chiến lược trong thương mại điện tử. Website thương mại điện tử không chỉ có hiển thị sản phẩm là có thể bán được hàng mà cần phải có sự hỗ trợ thêm từ nhiều hệ thống phụ trợ để hệ thống có được hiệu năng cao hơn.

Hiệu năng của thuật toán phụ thuộc rất nhiều vào thuật toán sử dụng để xây dựng hệ tư vấn. Trong bài báo cáo, tôi đã có cơ hội được thầy Nguyễn Mạnh Cường hướng dẫn xây dựng 7 thuật toán gợi ý. Thuật toán được sử dụng còn rất sơ sài và cần được ph át triển thêm. Việc phát triển tập trung vào việc khai phá thông tin và kết nối chúng.

## **Những việc chưa làm được**

* Các thuật toán còn rất sơ sài, đơn giản.
* Kết quả đưa ra gợi ý chưa được sát với yêu cầu của khách hàng.
* Sức sáng tạo còn hạn chế.
* Khả năng phát triển thuật toán chưa cao.

## **Hướng phát triển**

* Tập trung phát triển thuật toán, cố gắng khai phá thêm dữ liệu đã có để xây dựng hệ thống thuật toán chi tiết và hiệu năng sử dụng cao.
* Hoàn thành hệ tư vấn để tiến hành bảo vệ trong thời gian sắp tới.

Một lần nữa, tôi xin chân thành cảm ơn thầy Tiến sĩ Nguyễn Mạnh Cường đã hướng dẫn, giúp đỡ tôi hoàn thành tốt quá trình làm thực tập của tôi. Tôi chúc thầy mạnh khỏe và thành công hơn nữa trong công việc và cuộc sống.

**TÀI LIỆU THAM KHẢO**

* Dịch vụ tư vấn thông minh cho các hệ thống phân phối hàng hóa.ppt – TS. Nguyễn Mạnh Cường, Ha Noi University of Industry.
* Một số hướng tiếp cận trong xây dựng hệ thống gợi ý theo ngữ cảnh – Nguyen Thai-Nghe, Can Tho University([link](https://www.researchgate.net/publication/292281409_MOT_TIEP_CAN_TRONG_XAY_DUNG_HE_THONG_GOI_Y_THEO_NGU_CANH)).
* Hệ thống gợi ý sản phẩm trong bán hàng trực tuyến sử dụng kỹ thuật lọc cộng tác - Nguyễn Hùng Dũng & Nguyễn Thái Nghe, Can Tho University([link](http://sj.ctu.edu.vn/ql/docgia/download/baibao-7145/05-CNTT_NGUYEN%20HUNG%20DUNG(36-51).pdf)).
* Trang blog [nttuyen.wordpress.com](https://nttuyen.wordpress.com/2009/06/24/recommender-system-overview/).