

# 1. MỞ ĐẦU

Lợi nhuận của doanh nghiệp là khoản chênh lệch giữa doanh thu và chi phí mà doanh nghiệp bỏ ra đạt được doanh thu đó từ các hoạt động của doanh nghiệp đưa lại. Lợi nhuận là kết quả tài chính cuối cùng của các hoạt động sản xuất, kinh doanh, hoạt động tài chính, hoạt động khác đưa lại, là chỉ tiêu chất lượng để đánh giá hiệu quả kinh tế các hoạt động của doanh nghiệp.

Lợi nhuận có vai trò quan trọng tới sự tăng trưởng và phát triển của doanh nghiệp nói riêng và của toàn xã hội nói chung. Đây chính là động lực thúc đẩy doanh nghiệp năng động hơn nữa để khẳng định mình trong môi trường cạnh tranh gay gắt.

Bất kỳ một doanh nghiệp nào hoạt động trong cơ chế thị trường, điều đầu tiên mà họ quan tâm đó là lợi nhuận. Đây là một chỉ tiêu kinh tế tổng hợp phản ánh hiệu quả của quá trình kinh doanh, là yếu tố sống còn của doanh nghiệp. Doanh nghiệp chỉ tồn tại và phát triển khi nó tạo ra lợi nhuận, nếu doanh nghiệp hoạt động không có hiệu quả, thu không đủ bù đắp chi phí đã bỏ ra thì doanh nghiệp sẽ bị đào thải, đi đến phá sản. Đặc biệt trong điều kiện kinh tế thị trường có sự cạnh tranh diễn ra ngày càng gay gắt và khốc liệt vì vậy lợi nhuận là yếu tố cực kỳ quan trọng và quyết định đến sự tồn tại của doanh nghiệp.

Chính vì vậy, em đã chọn đề tài “Phân cụm khách hàng dựa vào lịch sử giao dịch” sử dụng phương pháp phân cụm Kmeans để xây dựng hệ thống hỗ trợ ra quyết định cho nhà quản lý. Hệ thống giúp nhà quản lý đưa ra những chính sách chăm sóc khách hàng phù hợp để tăng lợi nhuận cho doanh nghiệp.

Em xin gửi lời cảm ơn sâu sắc tới thầy *Lê Chí Ngọc* đã hướng dẫn tận tình để em có thể hoàn thiện được báo cáo này. Do điều kiện về thời gian và kiến thức, báo cáo không thể tránh khỏi những thiếu sót, em rất mong nhận được sự góp ý của thầy và các bạn. Em xin chân thành cảm ơn.

## 2.KHẢO SÁT HỆ THỐNG

### 2.1. Đặt vấn đề

#### 2.1.1. Nhu cầu hệ thống

Trong lĩnh vực bán lẻ, khách hàng có mối liên hệ chặt chẽ với lợi nhuận. Có thể nói khách hàng là yếu tố then chốt của doanh nghiệp, doanh nghiệp muốn có lợi nhuận và phát triển bền vững cần phải có khách hàng, cần duy trì, phát triển khách hàng và tìm mọi cách phải đáp ứng, thỏa mãn được yêu cầu của khách hàng. Tất cả các doanh nghiệp đều tìm cách giữ và thu hút thêm khách hàng của mình bằng nhiều hình thức.

Vậy làm thế nào để chăm sóc khách hàng hiệu quả ? Đây là bài toán lớn mà nhà quản lý phải giải quyết để giúp doanh nghiệp tồn tại và phát triển.

#### 2.1.2. Các khó khăn

##### **Không có được cái nhìn toàn diện về khách hàng:**

Việc không có được cái nhìn toàn diện về khách hàng là do vấn đề lưu trữ phân tán. File thông tin khách hàng được lưu riêng lẻ ở các máy nhân viên kinh doanh do đó dẫn tới việc phân tán thông tin, không có được cái nhìn tổng quan về số liệu khách hàng.

##### **Khó khăn trong việc khai thác dữ liệu khách hàng tiềm năng:**

Từ việc cập nhật dữ liệu khách hàng phân tán vì thế dẫn tới việc khai thác khách hàng sẽ khó khăn.

##### **Báo cáo thủ công:**

Báo cáo cũng là tác vụ thường gặp trong doanh nghiệp. Vấn đề mong muốn trong khâu này là phải nhanh chóng, chính xác và tức thời, tuy nhiên vì các doanh nghiệp đang lưu trữ dữ liệu phân tán nên việc làm báo cáo sẽ lâu, mất thời gian và không chính xác. Nếu với số lượng hồ sơ khách hàng nhiều thì càng khó khăn.

## 2.2. Giải pháp

### 2.2.1. Phân cụm khách hàng

Để giúp nhà quản lý hiểu khách hàng, một giải pháp hữu ích đó là phân cụm khách hàng dựa vào lịch sử giao dịch. Từ những đặc tính từng cụm, nhà quản lý có thể đưa ra những chính sách chăm sóc khách hàng phù hợp và hiệu quả. Từ những chính sách đó, doanh nghiệp có thể giữ chân khách hàng cũ và mở rộng tìm kiếm khách hàng mới, giúp tăng doanh thu và lợi nhuận.

### 2.2.2. Ưu điểm

- Ứng dụng công nghệ vào công tác quản lý.
- Áp dụng những tiến bộ khoa học kĩ thuật của trí tuệ nhân tạo.
- Báo cáo trực quan, sinh động.
- Dễ dàng đánh giá, ra quyết định nhanh chóng.

### 2.2.3. Phương pháp thực hiện

Sử dụng thuật toán Kmeans để phân cụm khách hàng.

Thuật toán Kmeans là thuật toán phân cụm dựa trên khoảng cách. Đây là phương pháp giải quyết lớp bài toán máy học học không giám sát. Kmeans được ứng dụng trong nhiều lĩnh vực như y học, kinh tế, môi trường, v.v. Kết quả của thuật toán có thể được sử dụng làm đầu vào cho nhiều bài toán máy học khác.

Ưu điểm của thuật toán:

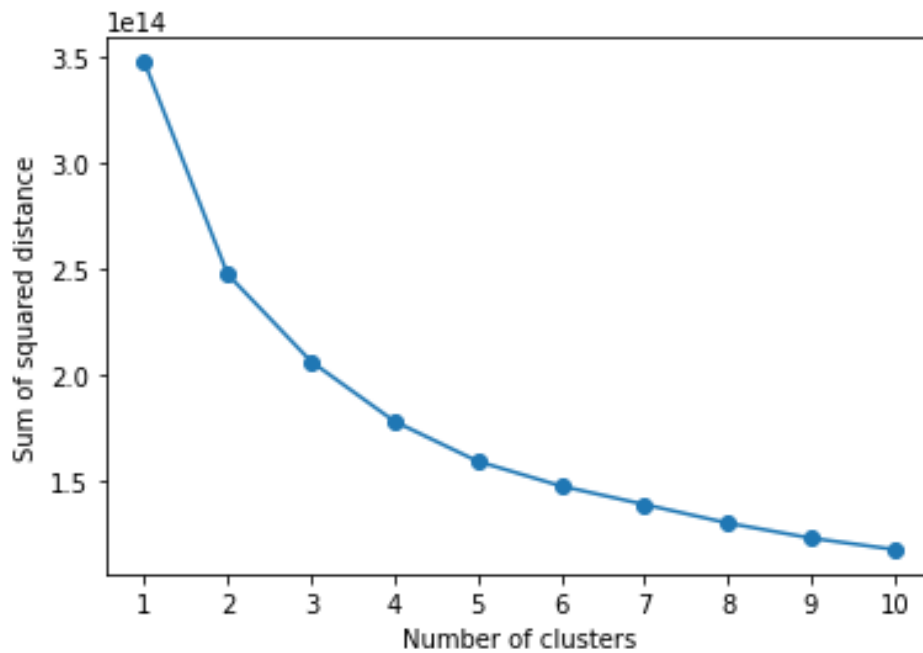
- Đơn giản
- Thực hiện nhanh

Một số hạn chế:

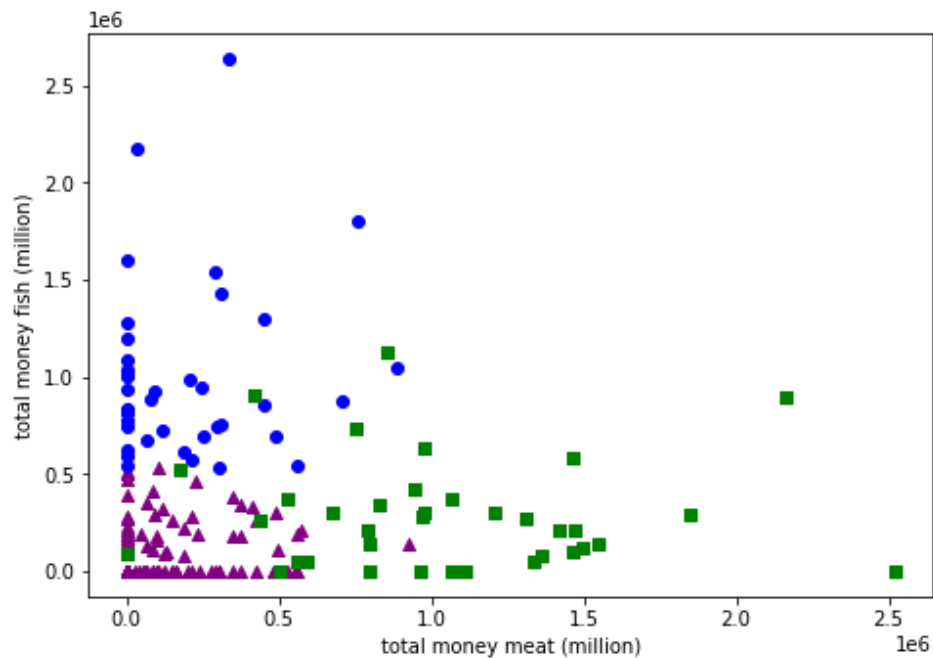
- Cần tính toán số lượng cụm
- Nhạy cảm với nhiễu
- Tập các phần tử trong cụm cần phải lõi
- Nghiệm tìm được là nghiệm địa phương

### 2.2.4. Kết quả

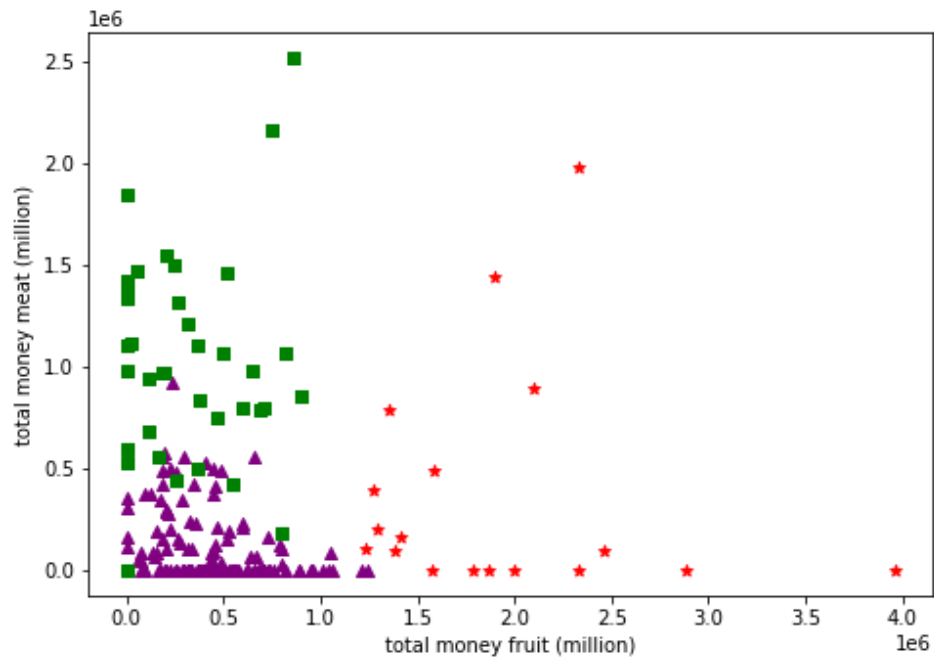
Một số kết quả thu được khi áp dụng thuật toán.



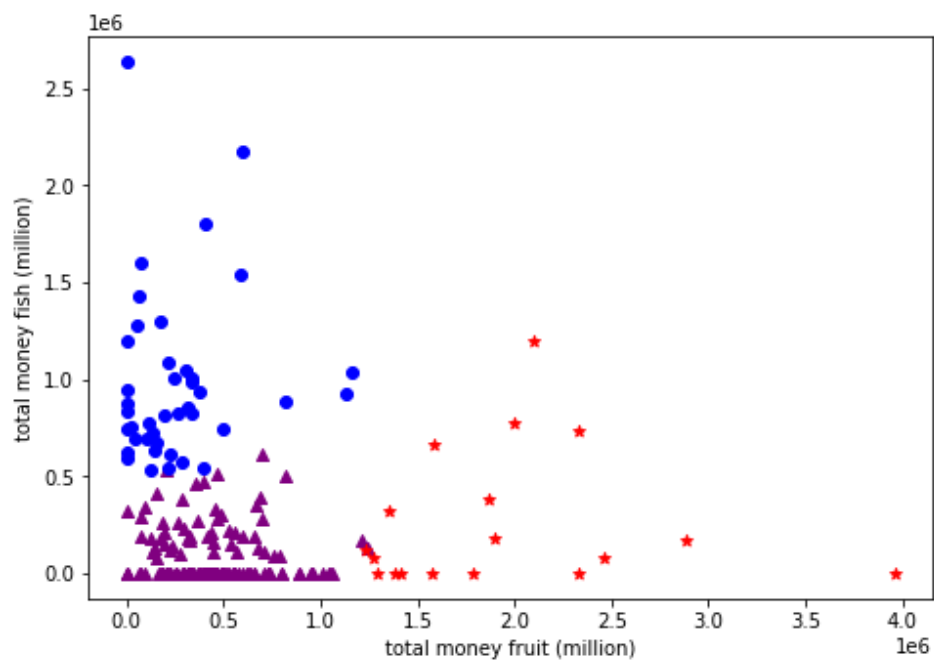
Biểu đồ đánh giá mối liên hệ giữa  
tổng bình phương khoảng cách và số lượng cụm



Biểu đồ phân bố dữ liệu của 3 trong 5 cụm  
theo nhóm hàng thịt và cá



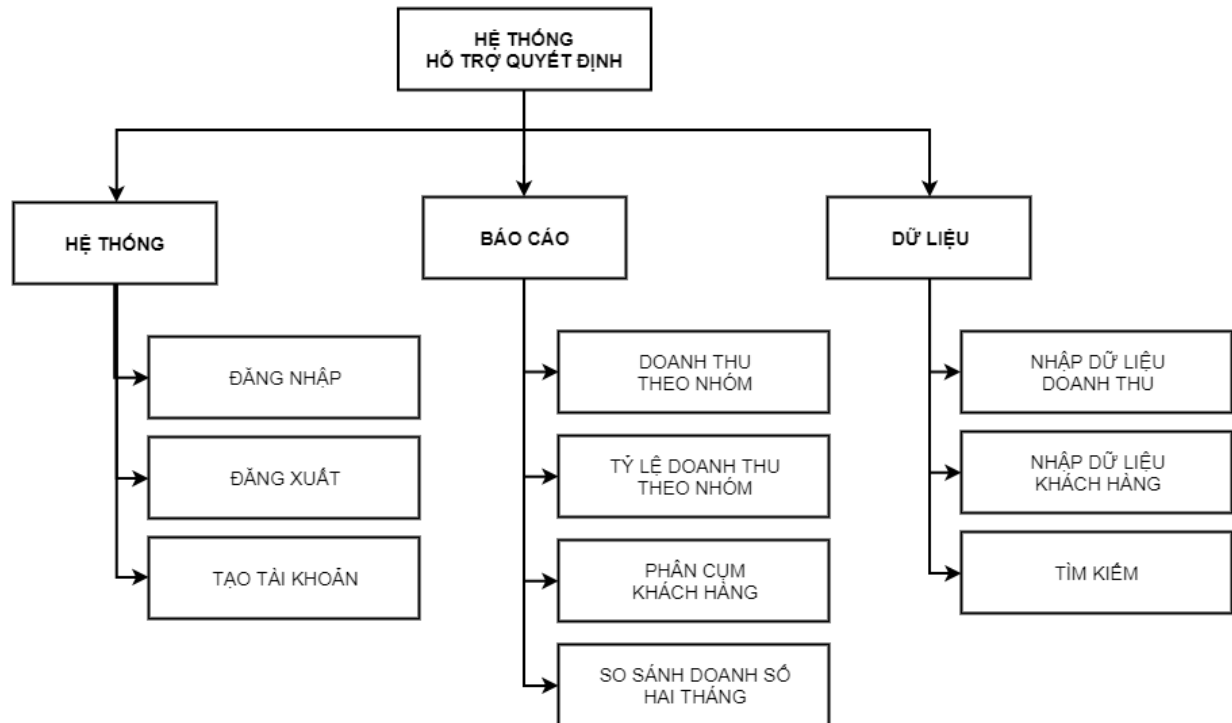
Biểu đồ phân bố dữ liệu của 3 trong 5 cụm  
theo nhóm hàng hoa quả và thịt



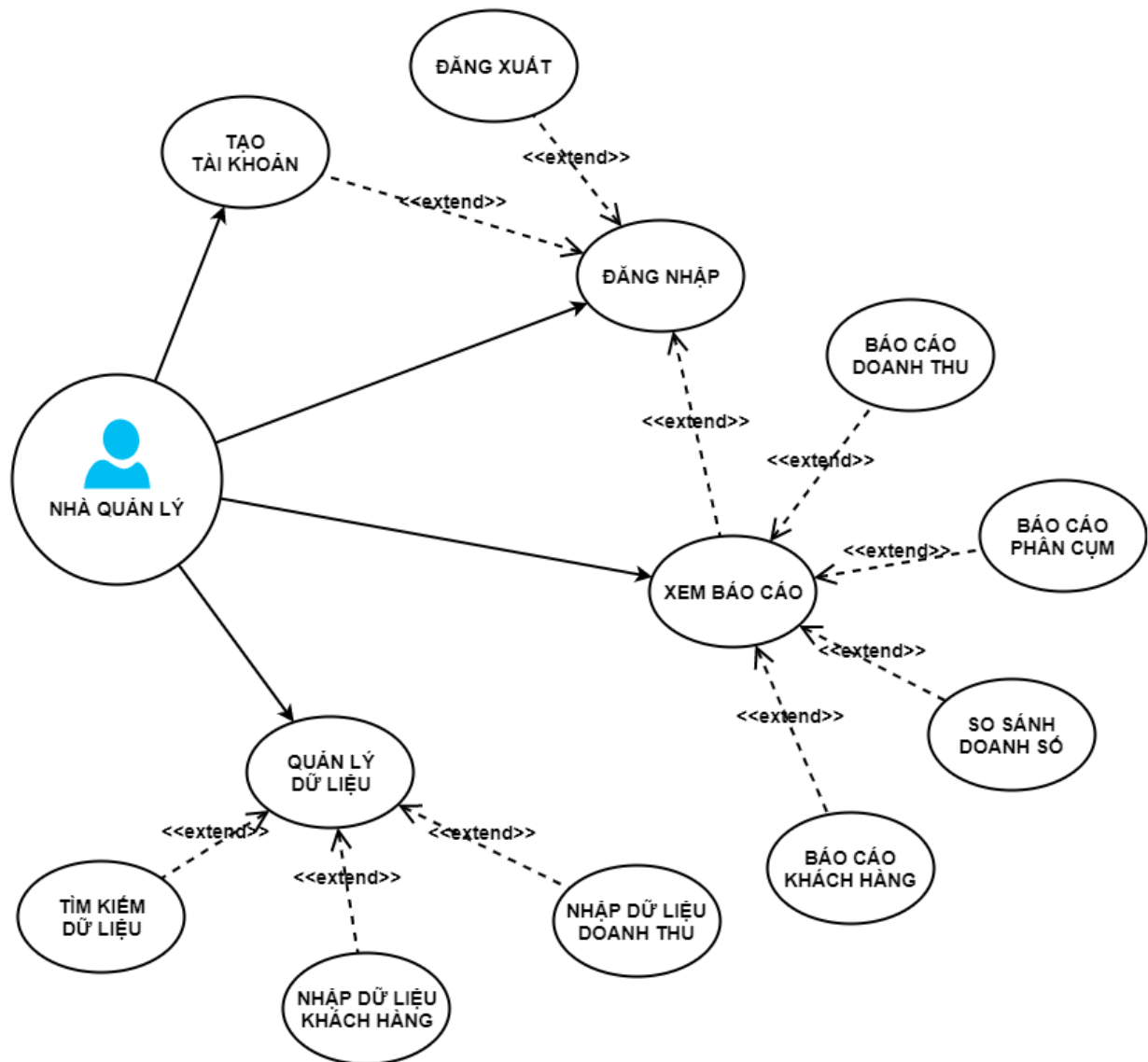
Biểu đồ phân bố dữ liệu của 3 trong 5 cụm  
theo nhóm hàng hoa quả và cá

### 3. PHÂN TÍCH HỆ THỐNG

#### 3.1. Sơ đồ phân cấp chức năng

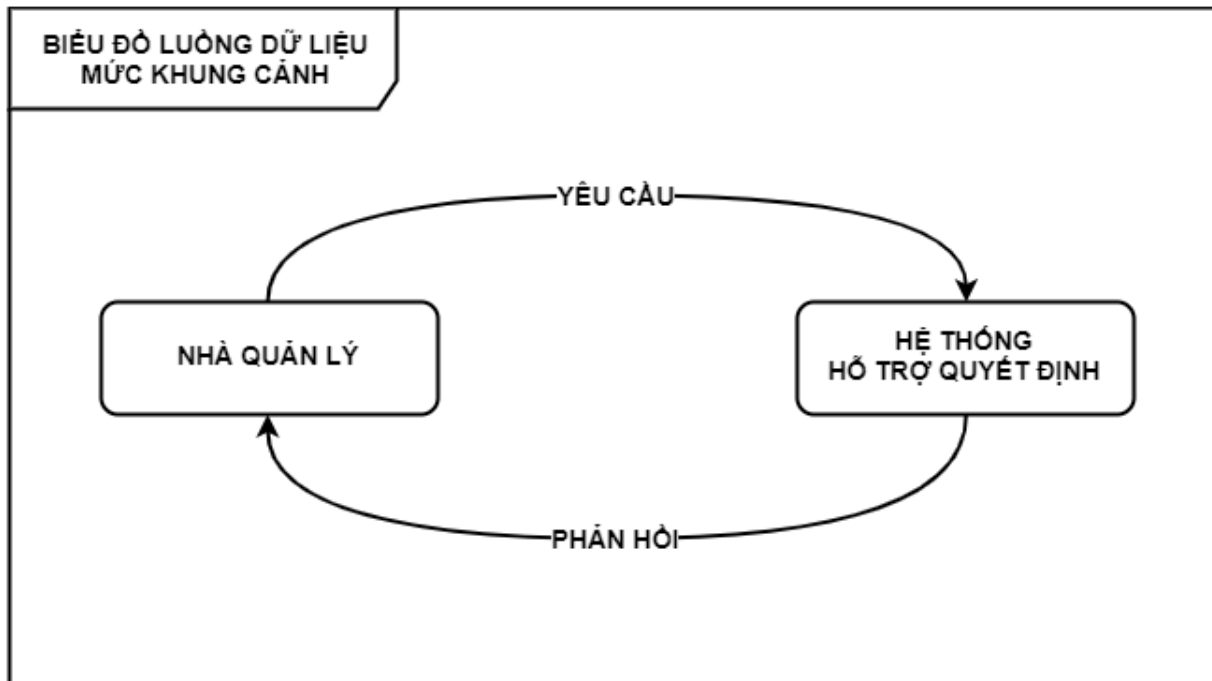


### 3.2. Biểu đồ usecase

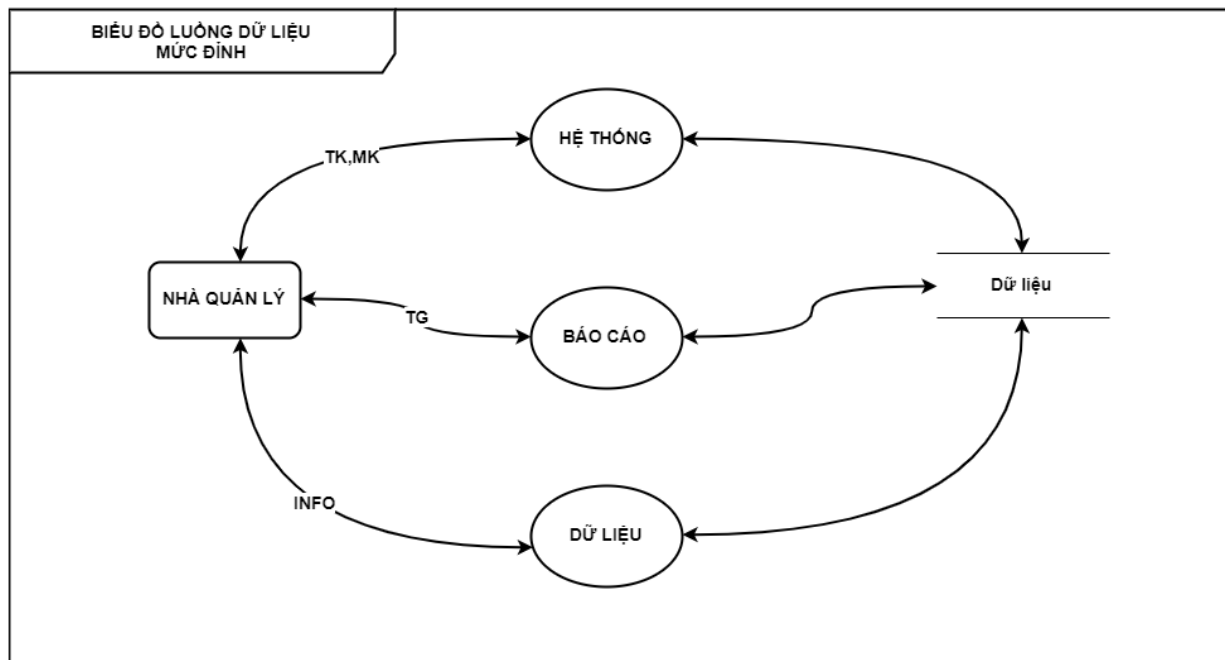


### 3.3. Biểu đồ luồng dữ liệu

#### 3.3.1. Mức khung cảnh

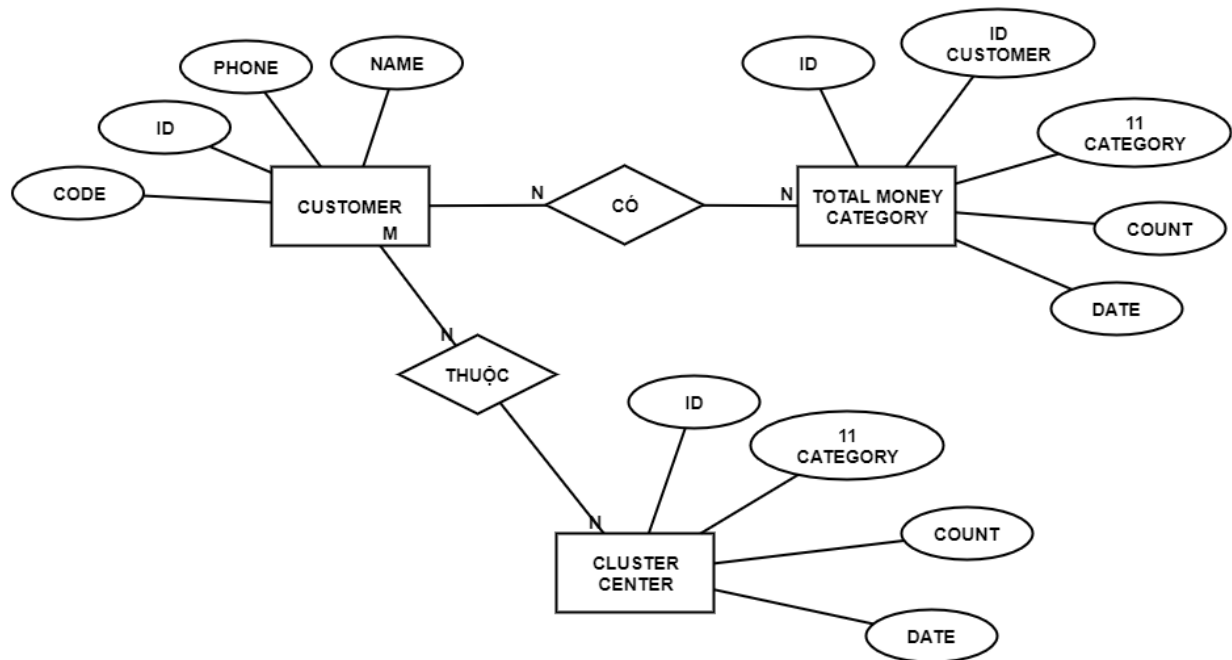


#### 3.3.2. Mức đỉnh



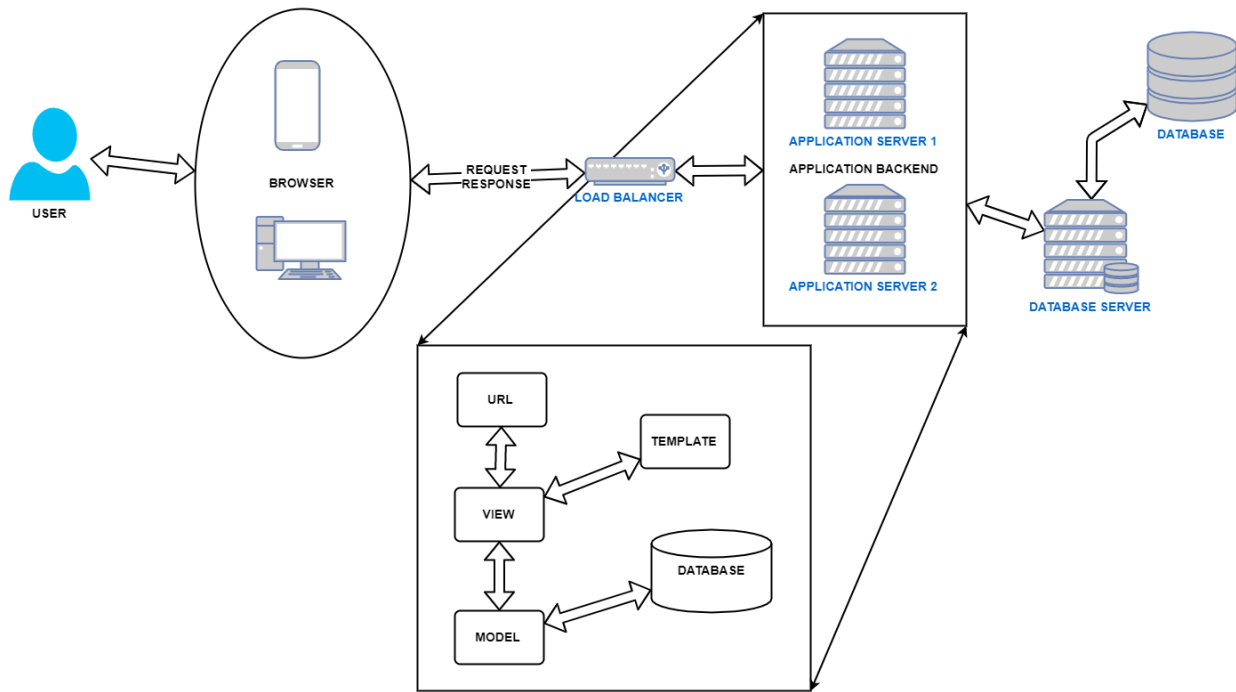


### 3.4. Sơ đồ thực thể lên kết

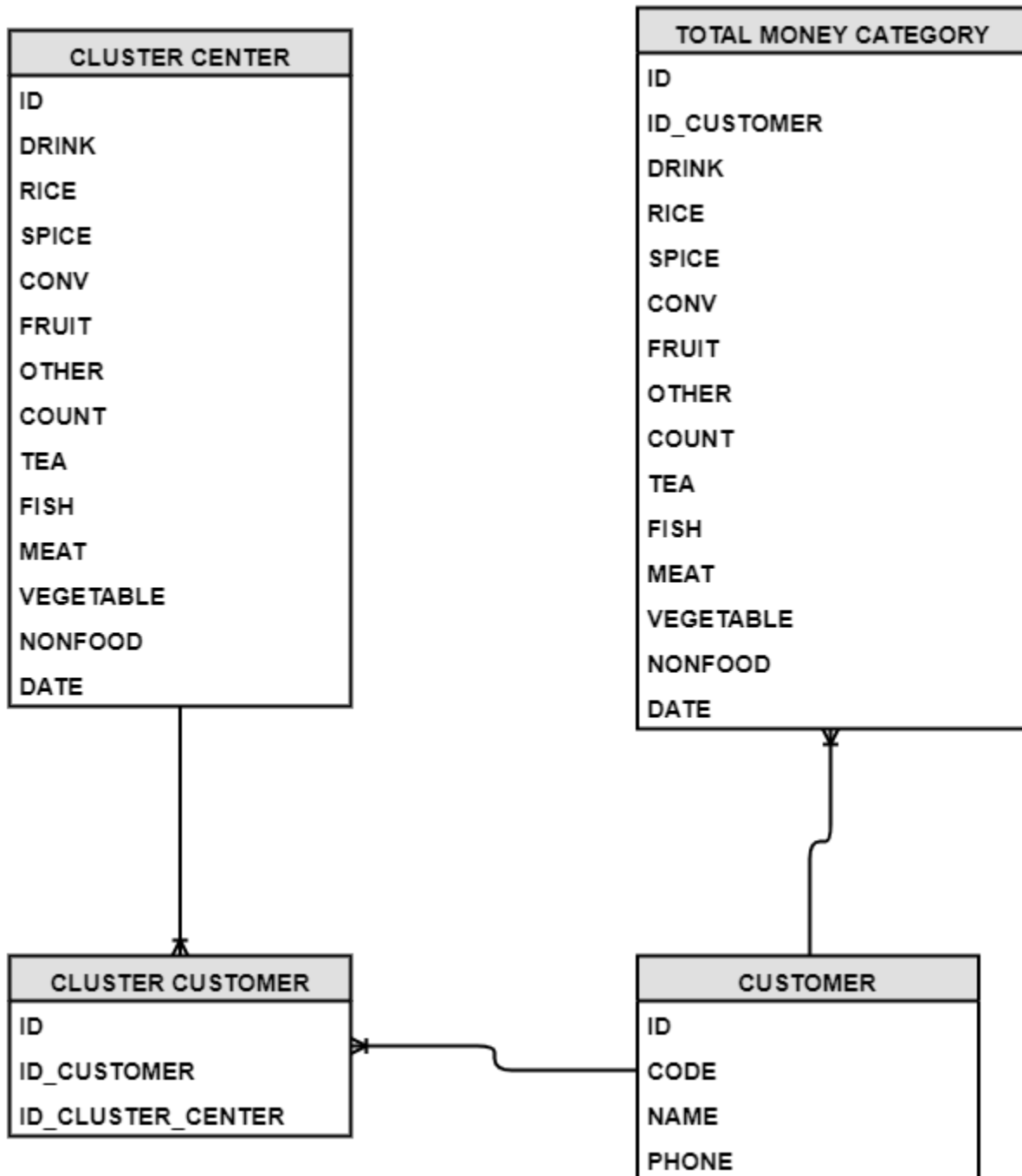


## 4. THIẾT KẾ HỆ THỐNG

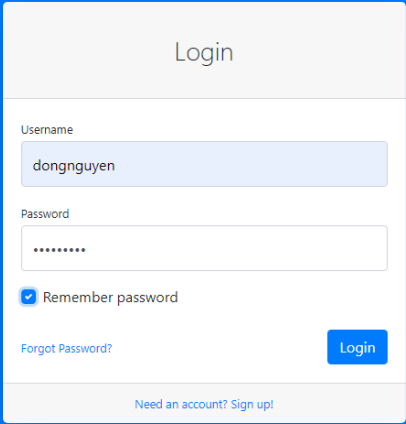
### 4.1. Sơ đồ kiến trúc hệ thống



## 4.2. Sơ đồ cơ sở dữ liệu



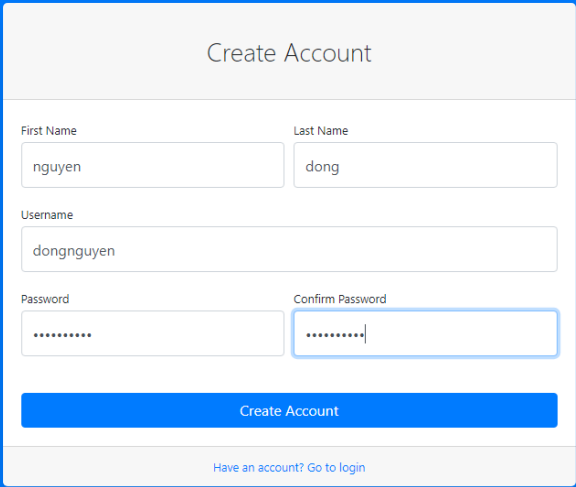
### 4.3. Giao diện người dùng



The image shows a login form centered on a solid blue background. The form is a white card with a light gray header containing the word "Login". Below the header, there are two input fields: "Username" with the text "dongnguyen" and "Password" with masked characters "\*\*\*\*\*". A checkbox labeled "Remember password" is checked. Below these is a link "Forgot Password?" and a blue "Login" button. At the bottom of the card is a link "Need an account? Sign up!".

Copyright © Dong Nguyen 2020

[Privacy Policy](#) - [Terms & Conditions](#)



The image shows a "Create Account" form centered on a solid blue background. The form is a white card with a light gray header containing the text "Create Account". Below the header, there are four input fields: "First Name" with "nguyen", "Last Name" with "dong", "Username" with "dongnguyen", and "Password" with "\*\*\*\*\*". There is also a "Confirm Password" field with "\*\*\*\*\*" and a cursor. A blue "Create Account" button is at the bottom of the form. Below the button is a link "Have an account? Go to login".

Copyright © Dong Nguyen 2020

[Privacy Policy](#) - [Terms & Conditions](#)

## Password Recovery

Enter your email address and we will send you a link to reset your password.

Email


[Return to login](#)

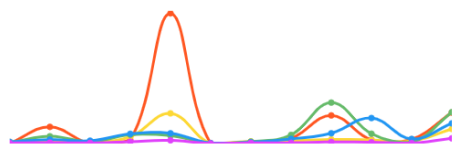
[Reset Password](#)


[Need an account? Sign up!](#)

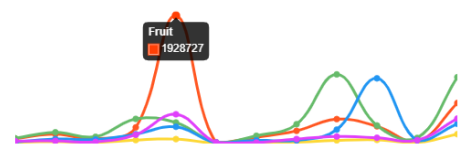
## Dashboard

Dashboard

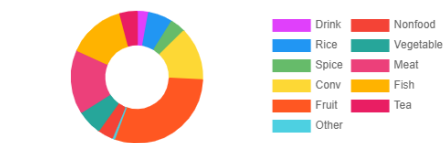
 Cluster Center Previous Month



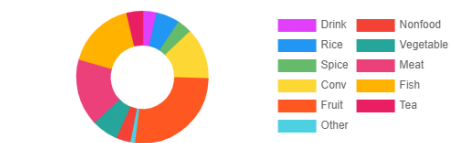
 Cluster Center Last Month




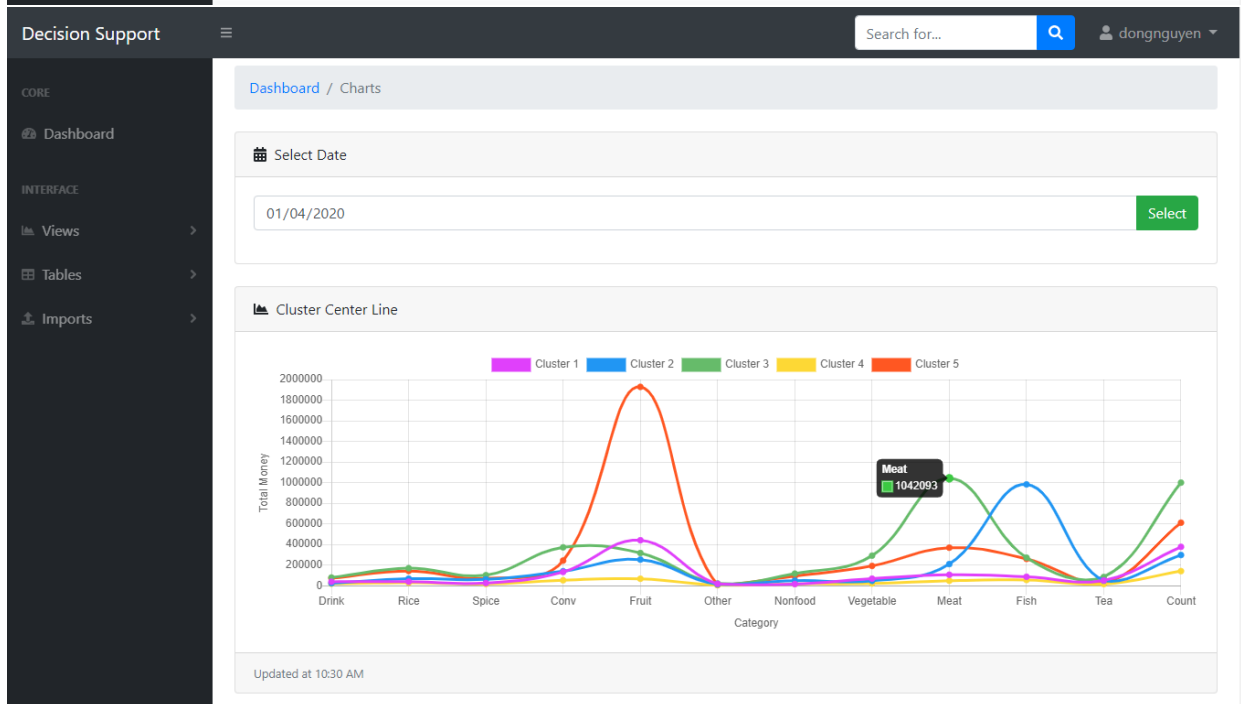
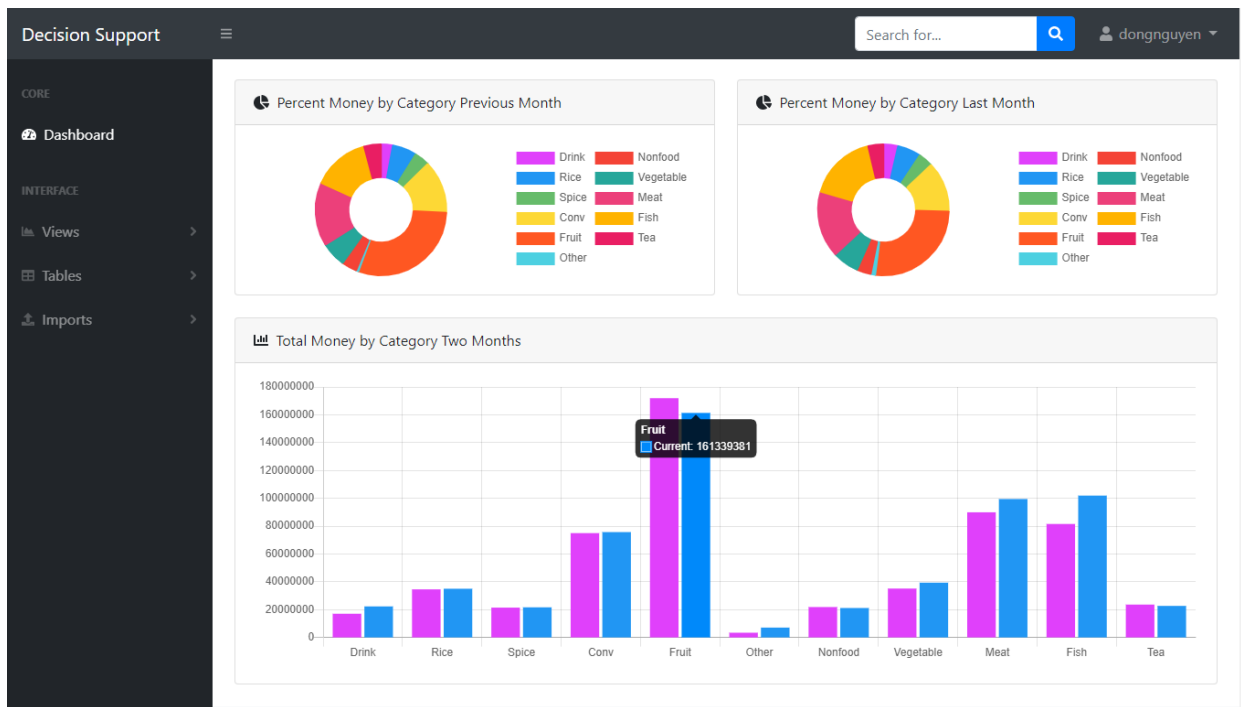
 Percent Money by Category Previous Month



 Percent Money by Category Last Month



 Total Money by Category Two Months



Decision Support

Search for...

dongnguyen

CORE

Dashboard

INTERFACE

Views

Tables

Customers

Total Money

Imports

Select Date

01/04/2020

Select

Total Money Table

Show 10 entries

Search: Anh Dinh Van

Customer	Drink	Rice	Spice	Conv	Fruit	Other	Nonfood	Vegetable	Meat	Fish	Tea
Anh Bui Dinh Van	772000	0	0	0	0	0	0	0	0	0	0
Anh Nguyen Dinh Van	390000	105000	80000	204300	51400	0	0	0	242880	386970	0
Anh Tran Dinh Van	0	0	0	0	370515	0	20000	0	0	0	0

Showing 1 to 3 of 3 entries (filtered from 857 total entries)

Previous1Next

Decision Support

Search for...

dongnguyen

CORE

Dashboard

INTERFACE

Views

Tables

Imports

Customers

Total Money

Imports Customer

Dashboard / Imports Customer

Imports Customer

Choose files

Browse

Import

Copyright © Dong Nguyen 2020

Privacy Policy · Terms & Conditions

## 4.4. Các công nghệ sử dụng

### 4.4.1. HTML

HTML là từ viết tắt của HyperText Markup Language (ngôn ngữ đánh dấu siêu văn bản) dùng mô tả cấu trúc của các trang Web và tạo ra các loại tài liệu có thể xem được trong trình duyệt.

HTML được tạo ra và phát triển bởi tổ chức W3C (World Wide Web Consortium). Hiện nay phiên bản mới nhất của HTML là HTML5 với nhiều tính năng ưu việt so với các phiên bản cũ.

### 4.4.2. CSS

CSS (Cascading Style Sheets) là một ngôn ngữ định dạng được sử dụng để mô tả trình bày các trang Web, bao gồm màu sắc, cách bố trí và phông chữ. CSS cho phép chúng hiển thị nội dung tương thích trên các loại thiết bị có kích thước màn hình khác nhau, chẳng hạn như màn hình lớn, màn hình nhỏ như điện thoại hay máy tính bản.

CSS là độc lập với HTML và có thể được sử dụng với bất kỳ ngôn ngữ đánh dấu nào xây dựng dựa trên XML. CSS tuân theo chuẩn chung do W3C quy định.

### 4.4.3. JavaScript

JavaScript là một ngôn ngữ nhỏ và nhẹ chạy trong môi trường máy chủ lưu trữ (ví dụ: trình duyệt web), JavaScript có thể được kết nối với các đối tượng của môi trường để cung cấp kiểm soát chương trình đối với chúng.

JavaScript cho phép thực hiện những hành động tạp trên các trang web như bản đồ tương tác v.v.

Javascript được hỗ trợ hầu như trên tất cả các trình duyệt như Firefox, Chrome, v.v. thậm chí các trình duyệt trên thiết bị di động cũng có hỗ trợ.

### 4.4.4. Bootstrap

Bootstrap là một framework cho phép thiết kế website reponsive nhanh hơn và dễ dàng hơn. Bootstrap bao gồm các HTML templates, CSS templates và Javascript tạo ra những thành phần



cơ bản có sẵn như: typography, forms, buttons, tables, navigation, modals, image carousels và nhiều thứ khác. Trong bootstrap có thêm các plugin Javascript giúp cho việc thiết kế reponsive dễ dàng hơn và nhanh chóng hơn.

#### 4.4.5. Django

Django là một framework cho phép xây dựng các trang web bằng ngôn ngữ bậc cao Python. Django xử lý nhiều vấn đề thường gặp trong việc phát triển web. Do đó, người lập trình có thể tập trung phát triển ứng dụng web nhanh chóng và dễ dàng.

#### 4.4.6. Python

Python là một ngôn ngữ lập trình thông dịch, hướng đối tượng và là một ngôn ngữ bậc cao, ngữ nghĩa động. Python hỗ trợ các module và gói, khuyến khích chương trình module hóa và tái sử dụng mã. Trình thông dịch Python và thư viện chuẩn mở rộng có sẵn dưới dạng mã nguồn hoặc dạng nhị phân miễn phí cho tất cả các nền tảng chính và có thể được phân phối tự do.

## 5. KẾT LUẬN

Bài báo cáo đã chỉ ra phương pháp phân cụm và thiết kế một hệ thống phân cụm khách hàng phục vụ cho nhu cầu của nhà quản lý, giúp quản lý một cách hiệu quả, trong đó chỉ ra phương pháp để phân cụm, các phân tích usecase, các biểu đồ, cũng như là cách thiết kế hệ thống. Tuy còn đơn giản và sử dụng với mục đích chính để minh họa, đánh giá đặc tính khách hàng nhưng hệ thống này có tiềm năng phát triển lớn để phục vụ những yêu cầu cao hơn của nhà quản lý.

Thông qua đây, ta cũng thấy rõ được những ưu điểm của hệ hỗ trợ quyết định trong quá trình quản lý: tự động hóa bằng các thuật toán máy học thông minh, giảm thiểu tối đa các quá trình thực hiện bằng tay, tăng hiệu quả quản lý. Tuy nhiên, cũng phải chú ý rằng hệ hỗ trợ quyết định chỉ có tác dụng hỗ trợ việc ra quyết định của nhà quản lý, chứ không được tự động ra quyết định thay cho nhà quản lý. Vai trò của con người trong một bộ máy công ty vẫn là tối quan trọng.