

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CẦN THƠ KHOA CÔNG NGHỆ THÔNG TIN &
TRUYỀN THÔNG**



KHAI KHOÁNG DỮ LIỆU

Đề tài

**XÂY DỰNG WEBSITE
GOM NHÓM DỮ LIỆU**

**Giảng viên hướng dẫn:
Thầy: Lưu Tiến Đạo**

**Sinh viên thực hiện:
Trần Nam Dương B1609765
Nguyễn Hải Anh B1609759**

Khóa : 42

Cần Thơ, 7/6/2020

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC CẦN THƠ KHOA CÔNG NGHỆ THÔNG TIN &
TRUYỀN THÔNG**



KHAI KHOÁNG DỮ LIỆU

Đề tài

**XÂY DỰNG WEBSITE
GOM NHÓM DỮ LIỆU**

**Giảng viên hướng dẫn:
Thầy: Lưu Tiến Đạo**

**Sinh viên thực hiện:
Trần Nam Dương B1609765
Nguyễn Hải Anh B1609759**

Khóa : 42

Cần Thơ, 7/6/2022

This image shows a full page of white paper with horizontal dashed lines, typical of primary school writing paper. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

LỜI CẢM ƠN

Lời đầu tiên, chúng em xin trân trọng cảm ơn và bày tỏ lòng biết ơn sâu sắc nhất tới thầy **Lưu Tiến Đạo – Giảng viên khoa công nghệ thông tin Trường Đại học Cần Thơ, giáo viên hướng dẫn bộ môn Khai Khoáng Dữ Liệu** đã nhiệt tình hướng dẫn, chỉ bảo.

Chúng em xin chân thành cảm ơn các thầy cô giáo đang giảng dạy tại khoa công nghệ thông tin **Trường Đại học Cần Thơ** đã nhiệt tình ủng hộ, cung cấp tài liệu và đưa ra những ý kiến đóng góp quý báu!

Cuối cùng, tôi xin dành lời cảm ơn chân thành tới bạn bè đã động viên, khuyến khích và tạo điều kiện cho chúng tôi hoàn thành tốt đề tài của mình.

Xin chân thành cảm ơn!

Cần Thơ, ngày 07 tháng 06 năm 2020

Người viết

Trần Nam Dương

Nguyễn Hải Anh

MỤC LỤC

I.	PHẦN GIỚI THIỆU	6
1.	Đặt vấn đề	6
2.	Lịch sử giải quyết	6
3.	Mục tiêu đề tài	6
4.	Đối tượng và phạm vi nghiên cứu	7
5.	Ý nghĩa của đề tài	7
II.	PHẦN NỘI DUNG	8
1.	Nguyên lý hoạt động	8
2.	Hướng dẫn sử dụng	9
2.1	Truy cập website.....	9
2.2	Chọn tập dữ liệu	9
2.3	Chọn cột	10
2.4	Hiển thị sơ đồ Elbow	11
2.5	Hiển thị kết quả	12
III.	PHẦN TỰ ĐÁNH GIÁ VÀ HƯỚNG PHÁT TRIỂN	13
1.	Những điểm đã làm được.....	13
2.	Hướng phát triển.....	13
	TÀI LIỆU THAM KHẢO	14

II. PHẦN GIỚI THIỆU

1. Đặt vấn đề

Hiện nay-thế kỉ 21, khi những ứng dụng công nghệ thông tin và việc tin học hóa được xem là một trong những yếu tố mang tính quyết định trong hoạt động của các chính phủ, tổ chức, cũng như các công ty, nó đóng vai trò hết sức quan trọng, có thể tạo ra những bước đột phá mạnh mẽ.

Với sự phát triển không ngừng về kĩ thuật máy tính và mạng điện tử, công nghệ thông tin cũng được những công nghệ có đẳng cấp cao và lần lượt chinh phục hết đỉnh cao này đến đỉnh cao khác. Mạng Internet là một trong những sản phẩm có giá trị hết sức lớn lao và ngày càng trở nên một công cụ không thể thiếu, là nền tảng chính cho sự truyền tải, trao đổi thông tin trên toàn cầu.

Cùng với đó là sự bùng phát dữ liệu đòi hỏi chúng ta phải tốn rất nhiều công sức để xử lý chúng. Để tạo sự tiện lợi và rút ngắn thời gian xử lý, chúng em tạo ra một Website gom nhóm dữ liệu giúp phân chia các dữ liệu một cách hợp lý, nhanh chóng. Từ đó có thể dự đoán các giá trị của dữ liệu, ứng dụng vào thực tế để dự báo Thời tiết, dự báo rầy nâu,...

2. Lịch sử giải quyết:

Trước đây đã có rất nhiều dự án tương tự do các anh/chị khóa trước, thầy cô đã làm, phần mềm hoạt động rất tốt nhưng độ tiện dụng không cao. Do người dùng phải tự cài đặt các ứng dụng khác để chạy phần mềm này (gây khó khăn cho người sử dụng, đặc biệt là những người không chuyên về CNTT).

Nhận ra được hạn chế đó, chúng em quyết định thực hiện chủ đề này dưới dạng một Website để tiện cho người sử dụng và tiết kiệm được nhiều thời gian.

3. Mục tiêu của đề tài

- Xây dựng Website gom nhóm dữ liệu
- Ngôn ngữ lập trình: Flask, python.
- Cho phép người dùng nạp lên các tập tin: csv, excel, json
- Hiển thị dữ liệu người dùng vừa nạp lên.
- Cho phép người dùng chọn các cột dữ liệu để tiến hành phân loại
- Cài đặt ít nhất một giải thuật phân loại: K-Means
- Hiển thị kết quả gom nhóm.

4. Đối tượng và phạm vi nghiên cứu

- Nghiên cứu ứng dụng Website bằng ngôn ngữ Python
- Phạm vi nghiên cứu: Youtube, các tài liệu của thầy Lưu Tiến Đạo.

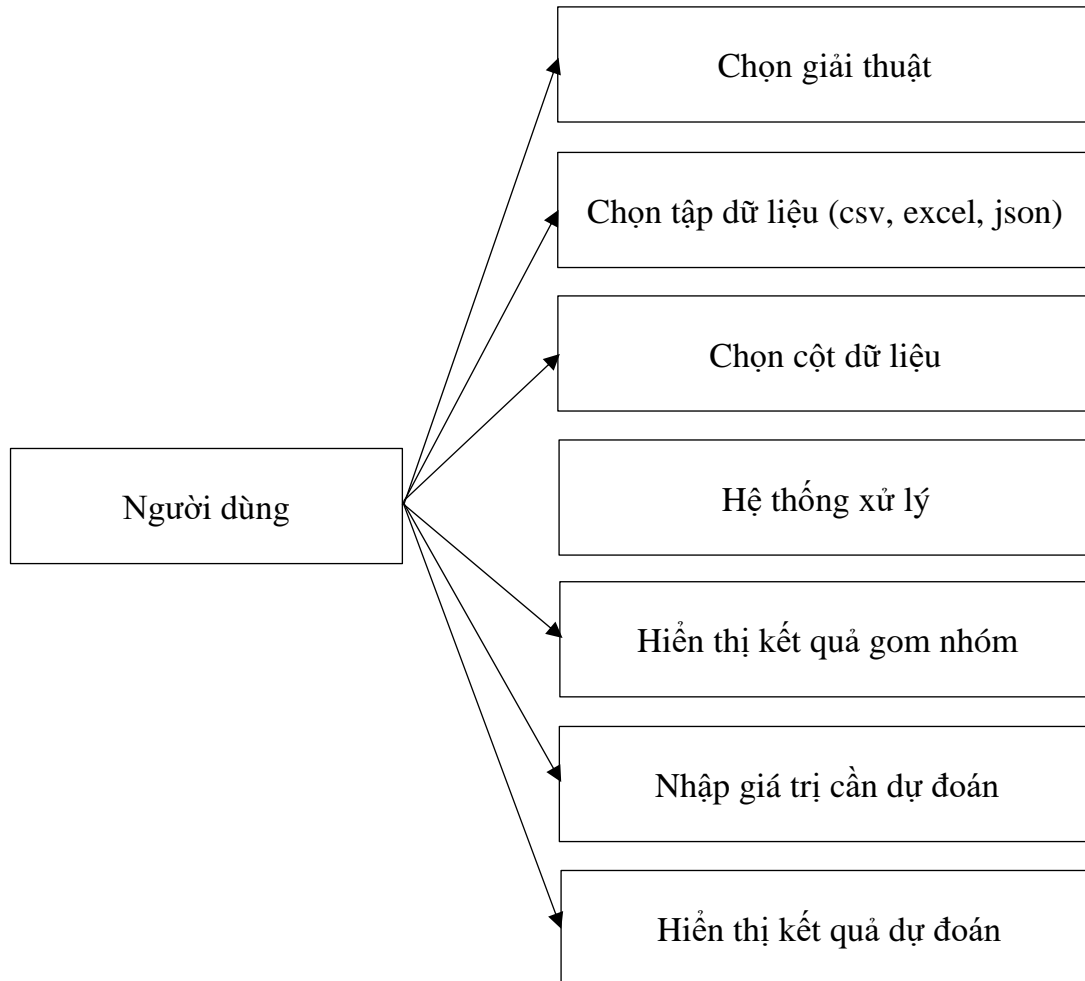
5. Ý nghĩa của đề tài

Đề tài này sẽ mang lại ý nghĩa to lớn về thực tiễn lẫn lý luận:

- **Ý nghĩa thực tiễn:** Với sự thành công của đề tài sẽ góp phần tối ưu hóa cho quá trình gom nhóm dữ liệu lớn giúp tiết kiệm thời gian, chi phí, hiệu quả và tính chuyên môn cao.
- **Ý nghĩa lý luận:** Chương trình và bản thuyết minh của đề tài rất thiết thực, dễ sử dụng, thích hợp cho tất cả mọi người đặc biệt là những người yêu thích bộ môn công nghệ Web và ứng dụng, nhất là lĩnh vực xây dựng Web quản lý trên môi trường INTERNET.

III. PHẦN NỘI DUNG

1. Nguyên lý hoạt động:

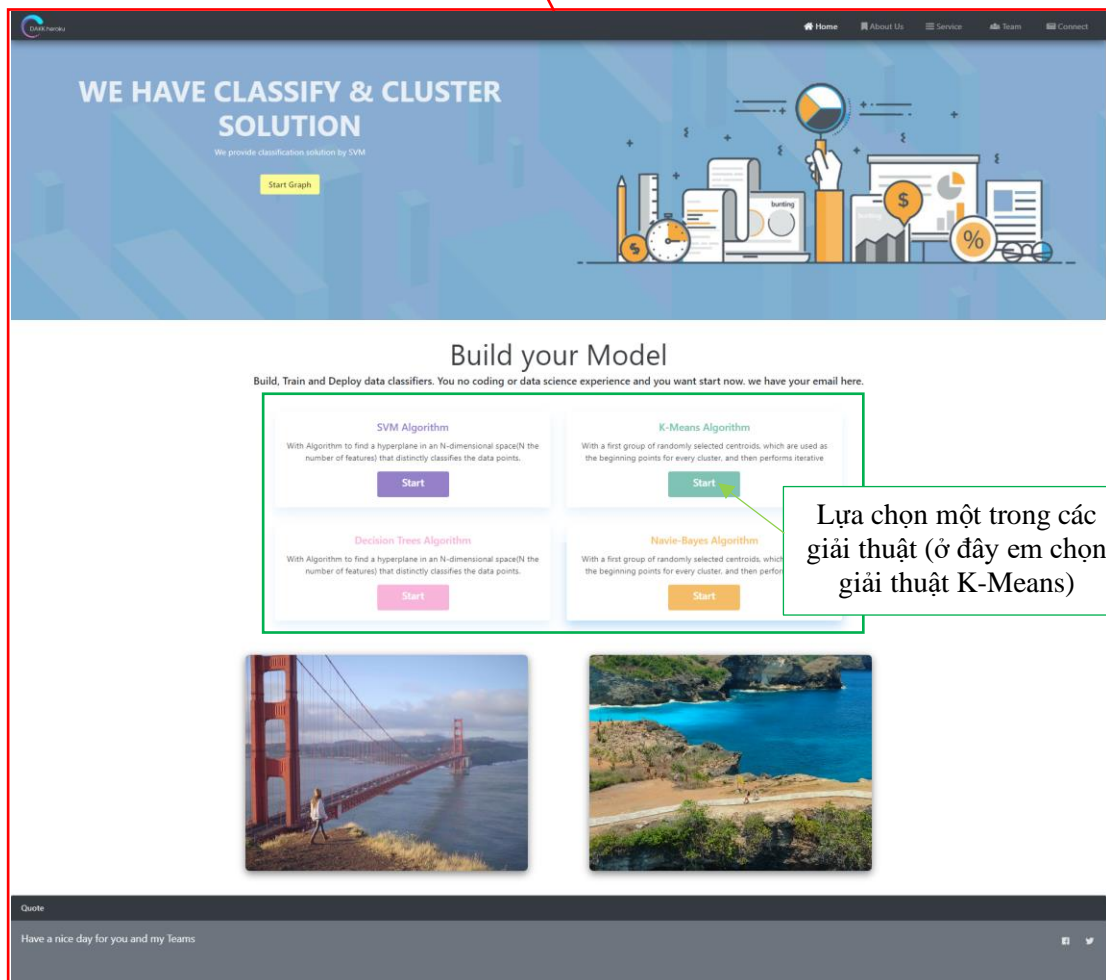


3. Hướng dẫn sử dụng:

3.1 Truy cập Website:

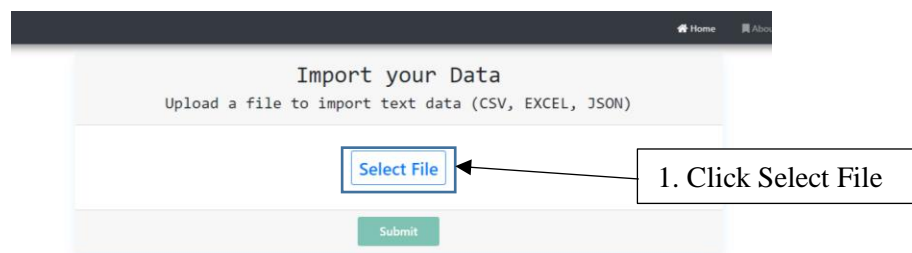
Sau khi truy cập Website, giao diện hệ thống sẽ hiện ra, bên dưới là các giải thuật cho mình lựa chọn.

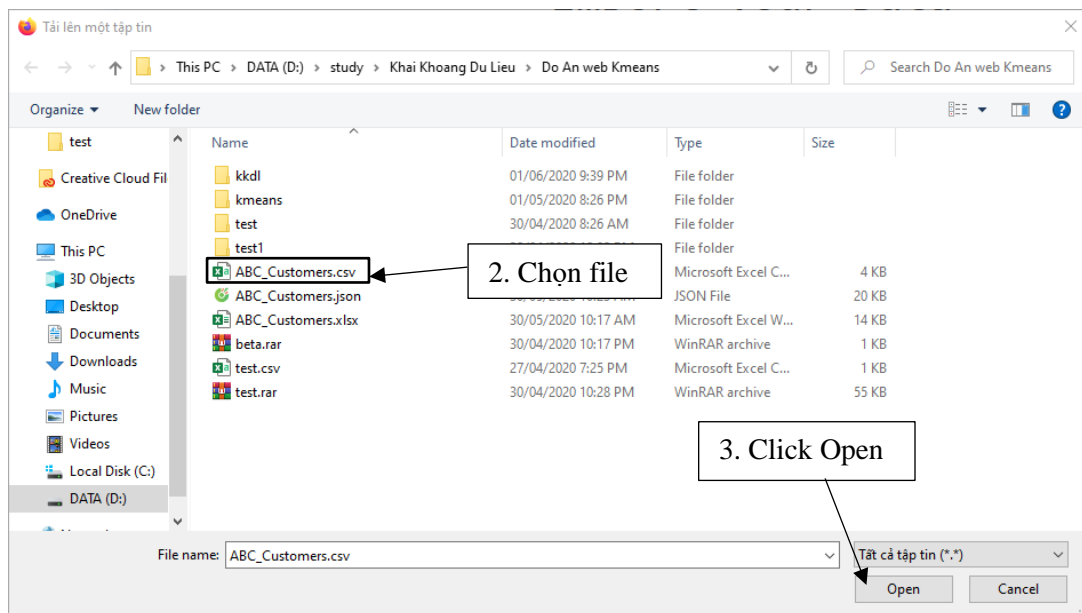
Giao diện hệ thống



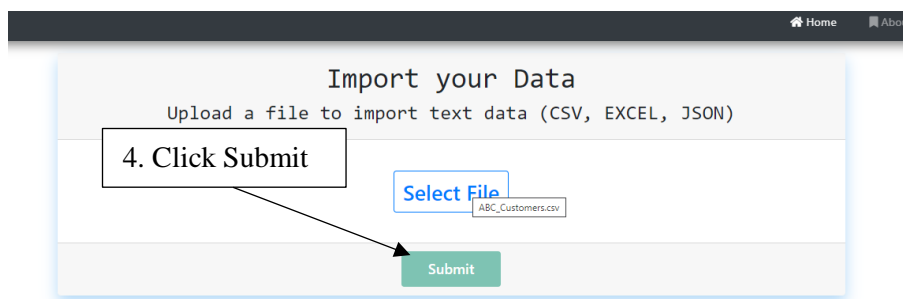
3.2 Chọn tập dữ liệu:

Sau khi chọn giải thuật, cửa sổ Import hiện ra, ta tiến hành Upload file bằng các click Select File, chọn file CSV, EXCEL, JSON và nhấn Open.



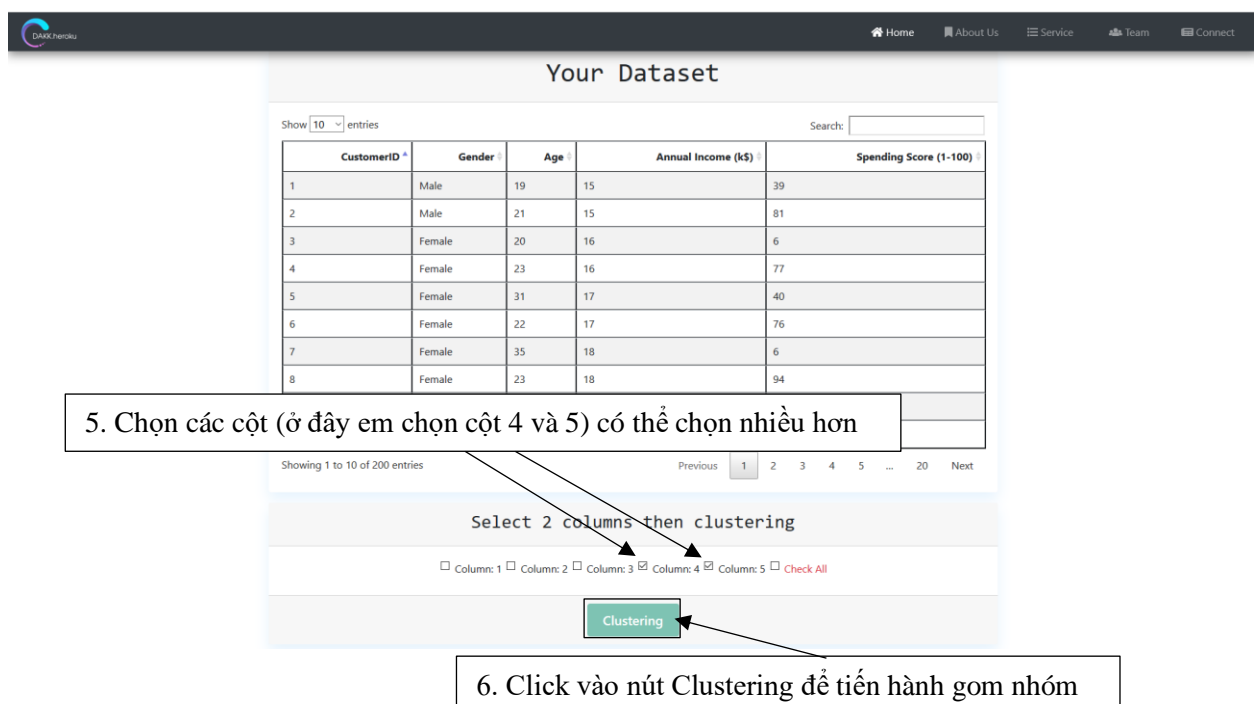


Click tiếp nút Submit để xác nhận.



2.3 Chọn cột:

Dữ liệu hiện ra sau khi Submit, tiếp theo là tiến hành lựa chọn các cột để gom nhóm.



2.4 Hiện thị sơ đồ Elbow

Dựa vào sơ đồ Elbow để chọn số nhóm, trong trường hợp này em chọn 3 nhóm (nằm trong khoảng gấp khúc của đồ thị). Nhập các giá trị cần dự đoán.

Recommend number of cluster

Đồ thị Elbow

Khoảng nhóm có thể chọn

7. Nhập số nhóm

Type number cluster then clustering

Type number of cluster or Choose according to the system then [Continue Clustering](#) button!.

Number of cluster 3

8. Nhập các giá trị cần dự đoán

Enter elements for Predict

Attribute 0

15

Attribute 1

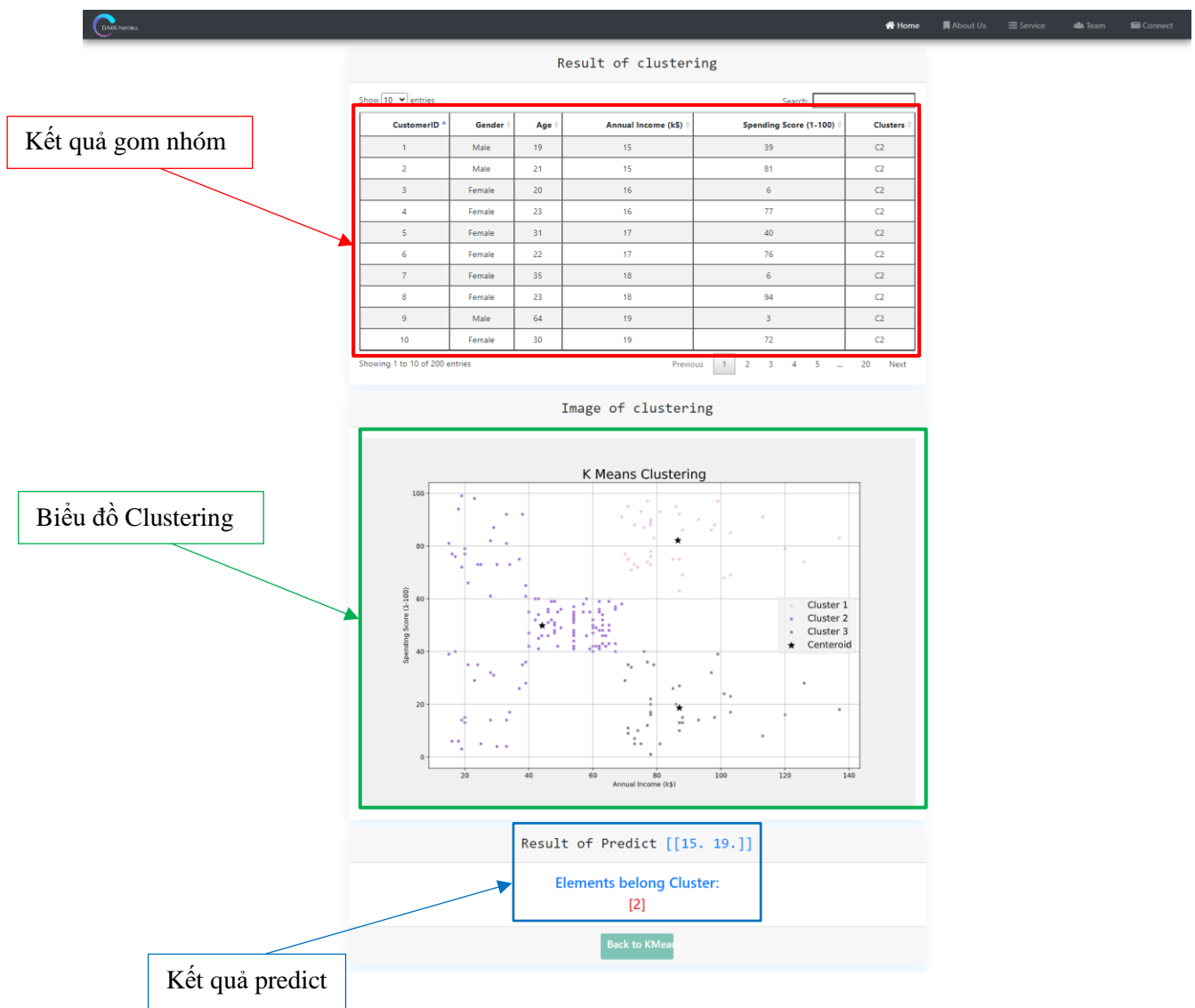
19

Continue Clust

9. Click Continue Clustering để tiến hành Predict các giá trị vừa nhập

2.5 Hiện thị kết quả:

Hiện thị kết quả gom nhóm, biểu đồ Clustering và kết quả Predict các giá trị vừa nhập.



IV. PHẦN TỰ ĐÁNH GIÁ VÀ HƯỚNG PHÁT TRIỂN

1. NHỮNG ĐIỂM ĐÃ LÀM ĐƯỢC

- ✓ Giao diện bắt mắt, thân thiện với người sử dụng.
- ✓ Có thể Upload nhiều file khác nhau.
- ✓ Hệ thống xử lý dữ liệu nhanh.
- ✓ Người dùng có thể chọn tùy ý các cột để xử lý.
- ✓ Dự đoán được các giá trị người dùng nhập vào.
- ✓ Hiển thị kết quả bằng các biểu đồ chi tiết, dễ hiểu.
- ✓ Các chức năng này rất dễ sử dụng trên giao diện web.

2. HƯỚNG PHÁT TRIỂN

Mong muốn Website có thể phát triển hơn hiện nay, có thêm nhiều chức năng như là Upload tập tin để Predict, giúp người dùng có thể dự đoán nhiều dữ liệu trong một lần.

TÀI LIỆU THAM KHẢO

1. Flask Python - <https://flask.palletsprojects.com/en/1.1.x/patterns/flashing/>
2. Lưu Tiến Đạo [2020] – Tài liệu hướng dẫn thực hành gom nhóm dữ liệu
3. Bootstrap 4 - <https://getbootstrap.com/docs/4.0/>