

**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  
PHÂN TÍCH DỮ LIỆU**

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

**Sinh viên thực hiện:**

**Nguyễn Nhật Khương                      B1609777**

**Nguyễn Đăng Khoa                        B1611129**

**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  
PHÂN TÍCH DỮ LIỆU**

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

**Sinh viên thực hiện:**

**Nguyễn Nhật Khương**

**B1609777**

**Nguyễn Đăng Khoa**

**B1611129**

**Khóa : 42  
Cần Thơ, 7/6/2022**

[illegible]

## LỜI CẢM ƠN

Lời đầu tiên, chúng em xin trân trọng cảm ơn **Khoa Công nghệ thông tin và Truyền thông** đã tạo điều kiện về cơ sở vật chất và tài liệu phong phú thuận lợi cho việc tìm kiếm, nghiên cứu thông tin cho đề tài nghiên cứu của chúng em.

Chúng em xin gửi lời cảm ơn trân thành đến thầy **Lưu Tiền Đạo – giảng viên khoa Công nghệ thông tin và Truyền thông** đã nhiệt tình hướng dẫn chúng em trong suốt quá trình thực hiện đề tài.

Trong phạm vi giới hạn về thời gian cũng như kiến thức, chắc chắn bài làm của em còn nhiều thiếu sót . Em rất mong được sự góp ý và chỉ dẫn của các Thầy Cô để bài làm được hoàn chỉnh hơn.

*Em xin chân thành cảm ơn!*

*Cần Thơ, ngày 07 tháng 06 năm 2020*

Người viết

Nguyễn Nhật Khương

Nguyễn Đăng Khoa

# MỤC LỤC

<b>I. PHẦN GIỚI THIỆU .....</b>	<b>6</b>
1. Đặt vấn đề .....	6
2. Mục tiêu của đề tài .....	6
3. Phạm vi nghiên cứu.....	6
4. Đối tượng đề tài.....	7
5. Kết quả đạt được.....	7
<b>II. PHẦN NỘI DUNG.....</b>	<b>8</b>
1. Nguyên lý hoạt động: .....	8
2. Hướng dẫn sử dụng: .....	8
2.1 Truy cập Website: .....	8
2.2 Chọn tập dữ liệu:.....	9
<b>IV. KẾT LUẬN .....</b>	<b>13</b>
1. Kết quả đạt được.....	13
2. Hướng phát triển .....	13
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>14</b>

# I. PHẦN GIỚI THIỆU

## 1. Đặt vấn đề

Chúng ta đang sống trong thời đại công nghệ 4.0, ứng dụng của công nghệ tin học đang ngày càng phát triển trên toàn thế giới và phát triển theo đó chính là dữ liệu, chúng được sinh ra ngày càng nhiều, mỗi ngày có vô số dữ liệu được sản sinh ra trong không gian mạng. Dữ liệu sẽ mang lại rất nhiều lợi ích khi chúng ta biết dùng chúng một cách hợp lý.

Trên thế giới sở hữu lượng dữ liệu vô tận sẵn có để làm việc. Các công ty lớn như Google và Microsoft sử dụng các kiến thức chuyên môn về phân tích dữ liệu để đưa ra kết quả chính xác của thông tin và sử dụng các dữ liệu đã được phân tích nhằm đưa ra các quyết định hay định hướng công việc phù hợp với tình hình thực tế. Nhưng vấn đề phân tích dữ liệu đối với người dùng bình thường hoặc các tổ chức nhỏ còn quá khó khăn, vì rất ít ứng dụng có thể áp dụng đầy đủ các tiêu chí đại trà mà người dùng phổ thông có thể tiếp cận.

Để sử dụng được các ứng dụng phân tích dữ liệu hiện nay, người dùng phải mất phí để có thể sử dụng được, bên cạnh đó người dùng phổ thông thường rất khó tiếp cận vì không có kiến thức chuyên môn. Nắm được yêu cầu thiết yếu đó nhóm em xin được chọn đề tài **“Xây dựng Website phân tích dữ liệu”** để làm đề tài cho bài báo cáo.

## 2. Mục tiêu của đề tài

- Xây dựng Website phân tích dữ liệu.
- Bao gồm 3 giải thuật: SVM, Decision Trees, Navie Bayes.
- Cho phép người dùng nạp lên các tập tin: csv, excel, json.
- Trang web được sử dụng miễn phí.
- Giao diện dễ sử dụng đối với người dùng phổ thông.

## 3. Phạm vi nghiên cứu

### – Về mặt lý thuyết:

- Tìm hiểu cách sử dụng ngôn ngữ python.
- Ứng dụng Flask framework để xử lý dữ liệu web (backend) [1]

- Tìm hiểu cách xây dựng trang web từ các ngôn ngữ lập trình: html, css, bootstrap framework.[3]
- **Về mặt thực hiện:**
  - Xây dựng trang web dữ liệu dựa trên tất cả các tiêu chí nêu trên.
- **Về mặt hoạt động:**
  - Website hoạt động hoàn chỉnh các tính năng phân tích dữ liệu trên 3 giải thuật (SVM, Decision Trees, Navie Bayes)

#### 4. Đối tượng đề tài

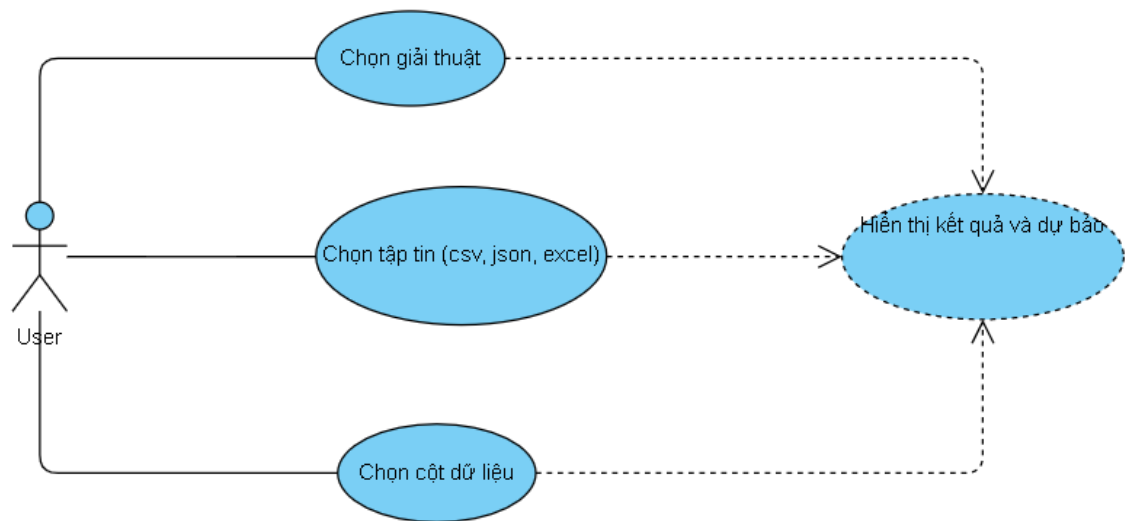
- Hướng đến người dùng phổ thông, người dùng không có kiến thức chuyên môn về lập trình vẫn có thể sử dụng được
- Bên cạnh đó website không áp dụng tính phí đối với các tính năng sẵn có.

#### 5. Kết quả đạt được

- **Ý nghĩa đối với người dùng:**
  - Phân tích dữ liệu dễ dàng với giao diện dễ sử dụng, không cần tốn kém chi phí sử dụng.
- **Ý nghĩa đối với cá nhân sinh viên:**
  - Hoàn thành đề tài một cách tốt nhất qua đó nâng cao trình độ hiểu biết, làm tiền đề để làm luận văn tốt nghiệp

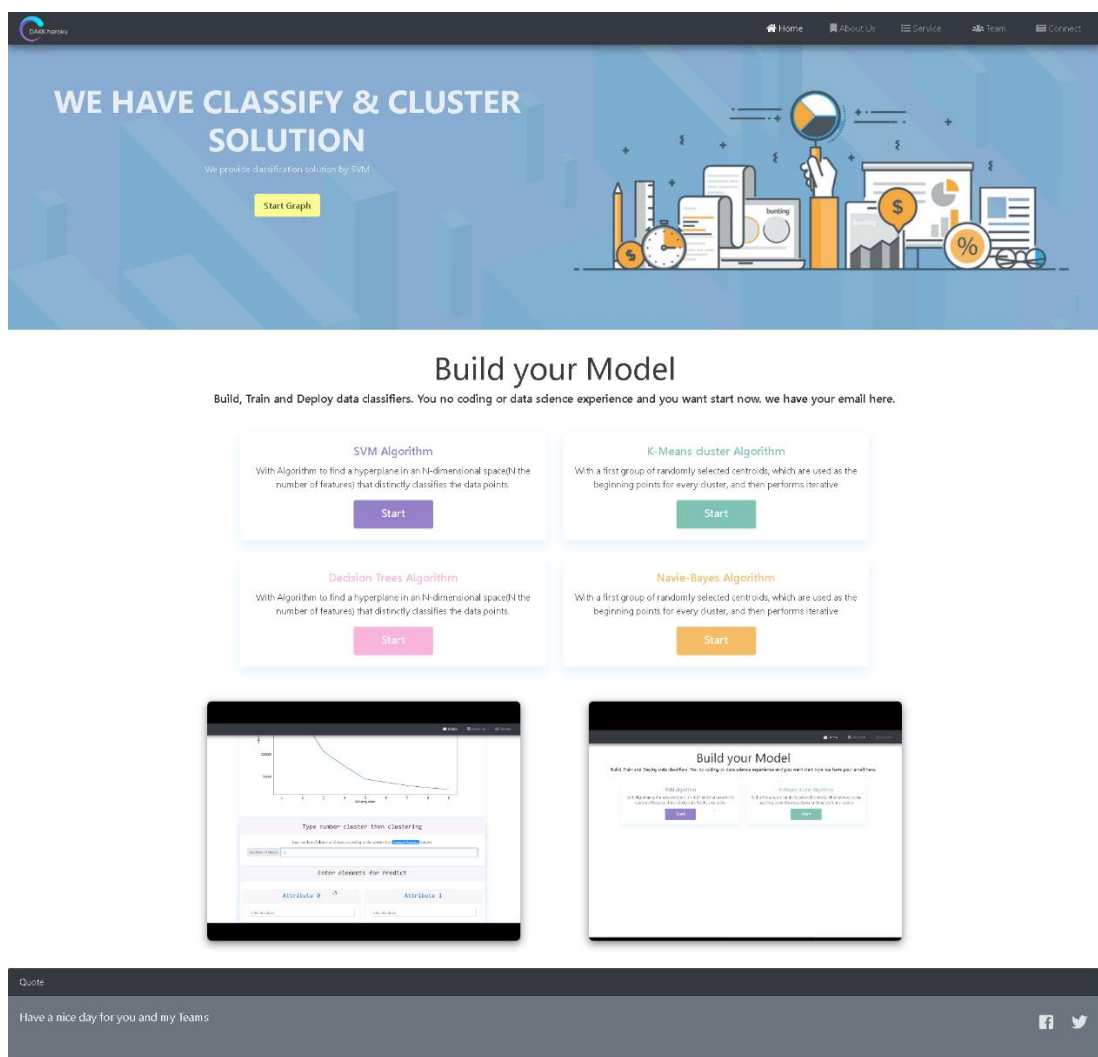
## II. PHẦN NỘI DUNG

### 1. Nguyên lý hoạt động:

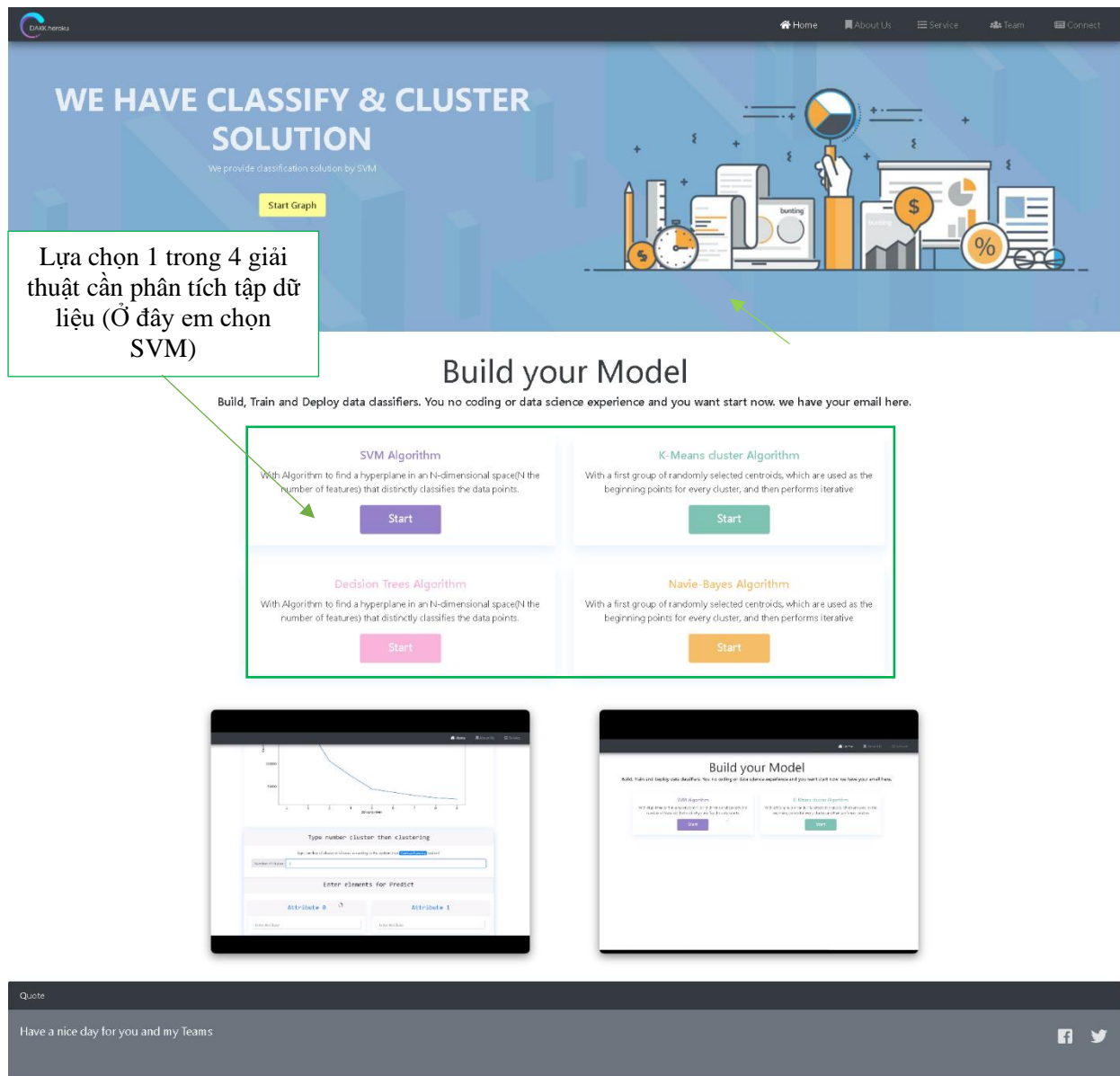


### 2. Hướng dẫn sử dụng:

#### 2.1 Truy cập Website:

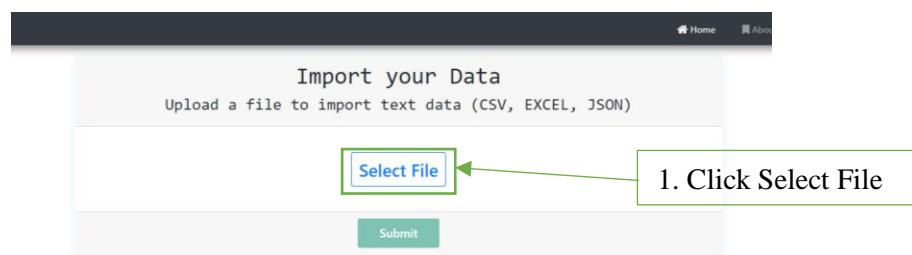


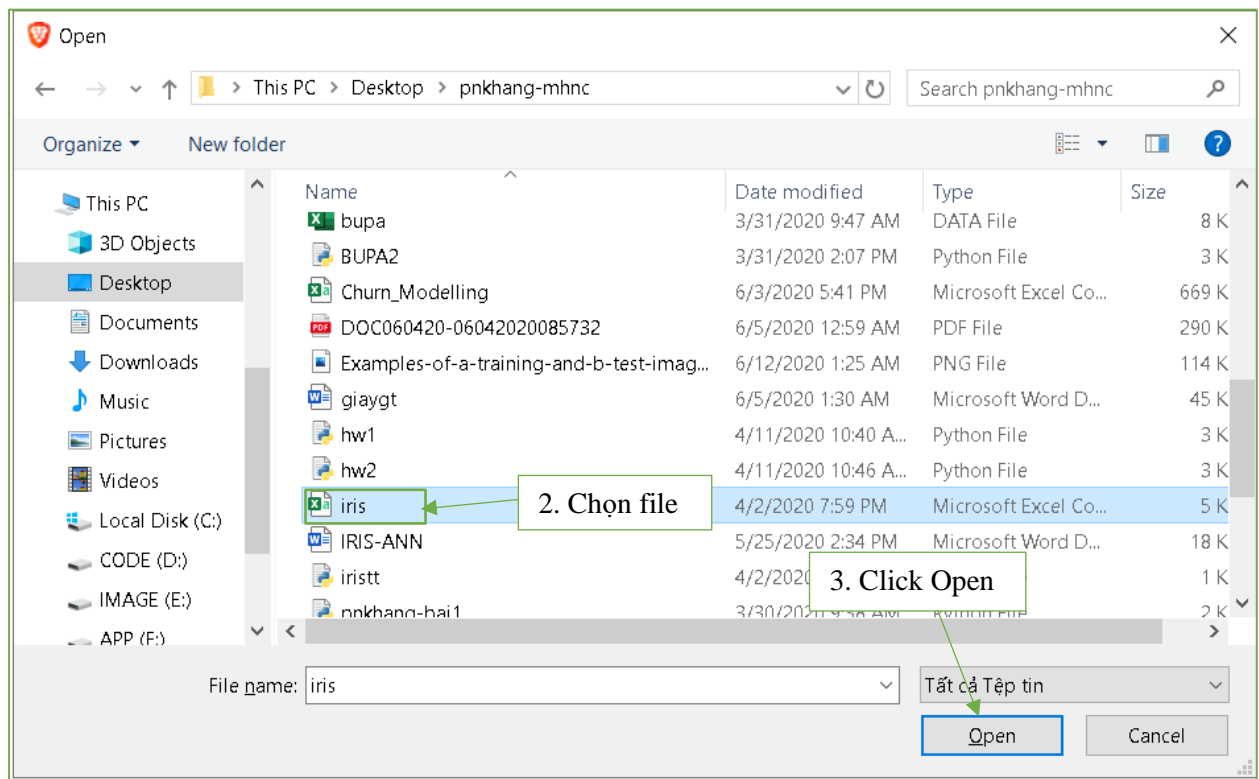




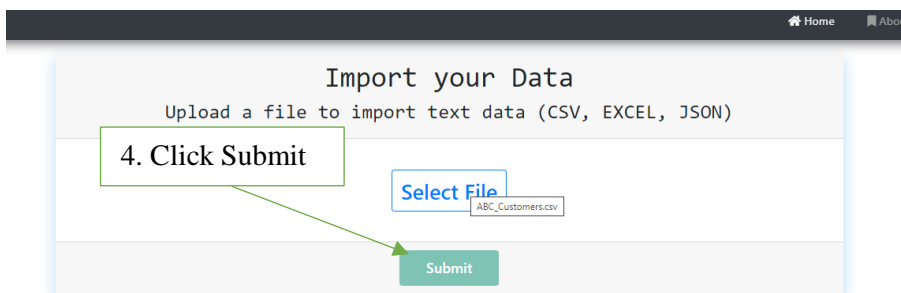
## 2.2 Chọn tập dữ liệu:

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.



DAIX.herku

Home About Us Service Team Connect

### Your Dataset

Show 10 entries Search

col_0	col_1	col_2	col_3	col_4
4.3	3.0	1.1	0.1	Iris-setosa
4.4	2.9	1.4	0.2	Iris-setosa
4.4	3.0	1.3	0.2	Iris-setosa
4.4	3.2	1.3	0.2	Iris-setosa
4.5	2.3	1.3	0.3	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa
4.6	3.6	1.0	0.2	Iris-setosa
4.6	3.2	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa

Showing 1 to 10 of 150 entries Previous 1 2 3 4 5 ... 15 Next

### Select a column is label then SVM

☐ Column: 0
 ☐ Column: 1
 ☐ Column: 2
 ☐ Column: 3
 ☒ Column: 4

### Select columns then SVM

☒ Column: 0
 ☒ Column: 1
 ☒ Column: 2
 ☒ Column: 3
 ☒ Column: 4
 ☒ Check all

### Select kernel for SVM Model

☐ Linear Kernel
 ☒ Polyminal Kernel
 ☐ RBF Kernel
 ☐ Sigmod Kernel

Classify

Chọn cột nhãn

Chọn nhân kernel

Chon cột dữ liệu phân nhóm

Chọn classify để phân nhóm

DAIX.herku

Home About Us Service Team Connect

### Your Dataset

Show 10 entries Search

col_0	col_1	col_2	col_3	col_4
4.3	3.0	1.1	0.1	Iris-setosa
4.4	2.9	1.4	0.2	Iris-setosa
4.4	3.0	1.3	0.2	Iris-setosa
4.4	3.2	1.3	0.2	Iris-setosa
4.5	2.3	1.3	0.3	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa
4.6	3.6	1.0	0.2	Iris-setosa
4.6	3.2	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa

Showing 1 to 10 of 150 entries Previous 1 2 3 4 5 ... 15 Next

### Accuracy

100.0

### New Element

Column: 0: 4.0  
 Column: 1: 3.1  
 Column: 2: 1.8  
 Column: 3: 0.8

Predict

Chọn dữ liệu cho phần tử mới đến

Dự báo nhãn cho phần tử mới đến

**Hiển thị kết quả dự báo.**

Kết quả dự báo nhãn cho phần tử mới đến	Accuracy
	95.55555555555556
New_element	
[4. 3.1 1.8 0.8]	
Class	
['Iris-setosa']	

## **IV. KẾT LUẬN**

### **1. Kết quả đạt được**

- Hoàn thành tốt các tính năng đề ra
- Đảm bảo đủ yêu cầu về kiến thức
- Xây dựng web trực quan, dễ sử dụng
- Đúng với các mục đích đặt ra

### **2. Hướng phát triển**

- Phát triển thêm tính năng upload file để test dữ liệu.
- Do hạn chế về mặt phần cứng và kinh phí nên không thể xử lý được các tập dữ liệu dung lượng lớn.
- Cải thiện về giao diện
- Phát triển thêm nhiều giải thuật.

## **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/>