

## MỤC LỤC

<b>1.GIỚI THIỆU VỀ HOA IRIS:</b>	2
<b>2. ĐỊNH NGHĨA BÀI TOÁN:</b>	3
<b>3.THU THẬP VÀ TIỀN XỬ LÝ SỐ LIỆU:</b>	3
3.1.LÀM SẠCH DỮ LIỆU( DATA CLEANING):	4
3.2.CHỌN LỌC DỮ LIỆU (DATA SELECTION):	4
3.3.BIẾN ĐỔI DỮ LIỆU (DATA TRANSFORMATION):	4
3.4.RÚT GỌN DỮ LIỆU (DATA REDUCTION):	5
<b>4.KHAI PHÁ DỮ LIỆU (DATA MINING):</b>	5
4.1.THUẬT TOÁN “K – NGƯỜI HÀNG XÓM GẦN NHẤT” LÀ GÌ:	5
4.2.MÔ TẢ THUẬT TOÁN K-NN:	5
4.3.ÁP DỤNG THUẬT TOÁN K-NN VÀO BÀI TOÁN PHÂN LỚP HOA IRIS:	7
<b>5.ĐÁNH GIÁ ĐỘ CHÍNH XÁC CỦA MÔ HÌNH VÀ CHƯƠNG TRÌNH CHẠY:</b>	10
5.1.CÁCH THỨC ĐÁNH GIÁ ĐỘ CHÍNH XÁC CỦA MÔ HÌNH:	10
5.2.CHƯƠNG TRÌNH CHẠY:	12
<b>TÀI LIỆU THAM KHẢO</b>	16

# 1.GIỚI THIỆU VỀ HOA IRIS:

Hoa Iris có tên tiếng việt là hoa Diên Vĩ. Tiếng Hy Lạp, "Iris" có nghĩa là Cầu Vồng. Trong thần thoại Hy Lạp, Iris là tín sứ của thần Zeus và nàng thường xuất hiện dưới hình một chiếc cầu vồng. Nàng là người đưa tin trên đỉnh Olympus, Iris mang thông điệp của các vị thần linh từ "con mắt Thiên Đường" xuống cho nhân loại trên trái đất qua vòng cung cầu vồng rực rỡ. Từ Iris cũng có nghĩa là "con mắt Thiên Đường" (The eye of Heaven). Iris, như ta đã biết, là tên của một nữ thần Hy Lạp, một loài hoa, và nó còn có nghĩa là trông đen trong con mắt chúng ta. Điều này ngụ ý rằng mỗi chúng ta đều mang trong mình một mảnh của Thiên Đường. Những người đàn ông Hy Lạp thường trồng hoa Diên Vĩ tím trên mộ những người phụ nữ mà họ yêu thương để tỏ lòng tôn kính nữ thần Iris, người mang sứ mệnh dẫn dắt những linh hồn phụ nữ này đến chốn Thiên Đàng (the Elysian fields).

Diên Vĩ là loài cây lưu niên có thân thảo vươn cao, lá hình lưỡi kiếm và những đóa hoa to nhiều màu sắc với ba cánh và ba đài hoa rủ xuống. Có hơn 300 loài hoa Diên Vĩ xinh đẹp khác nhau với các màu xanh da trời nhạt, tím, vàng, trắng, hồng, cam, nâu đỏ...đa dạng như màu sắc cầu vồng. Hoa Diên Vĩ được xem như sứ giả mang đến những niềm hy vọng. Ba cánh hoa Diên Vĩ đại diện cho lòng Trung Thành, sự Khôn Ngoan và lòng Dũng Cảm. Hoa Diên Vĩ vàng là biểu tượng của ngọn lửa và niềm đam mê

Hoa Diên Vĩ đã từng được thấy ở sa mạc, đầm lầy hay cả miền Bắc cực Siberia lạnh giá, nhiều nhất vẫn là ở các vùng khí hậu ôn hòa

Hoa Diên Vĩ đã là biểu tượng của hoàng gia và sự che chở thần thánh suốt hàng thế kỷ trên khắp thế giới. Loài hoa đầy sức thu hút này được rất nhiều người ngưỡng mộ. Các vị vua chúa nước Pháp đã dùng nó làm biểu tượng hoàng gia, gọi nó là Fleur-de-lis. "Fleur-de-lis" có nguồn gốc từ tên "Fleur-de-Louis", sau thời vua Louis VII, năm 1147. Theo thời gian, tên đó chuyển thành "Fleur-de-luce", có nghĩa là hoa của ánh sáng (flower of light), cuối cùng đến ngày nay, nó được gọi là "Fleur-de-Lys", hay Flower of the Lily (Lily : Hoa Huệ Tây, Loa Kèn, Bách Hợp). Fleur-de-Lis đã là biểu tượng của nước Pháp từ thế kỷ 13. Hoàng gia Pháp trang trí hoa Diên Vĩ trên áo choàng, các đồ vật trong cung điện và trên những bức tường như biểu hiện của sự toàn bích, ánh sáng và cuộc sống. Có nhiều truyền thuyết khác nhau giải thích tại sao hoa Diên Vĩ được chế độ quân chủ Pháp chọn làm biểu tượng . Tương truyền rằng, Clovis, vua nước Pháp triều đại Mêrôve khi đối mặt với đội quân thiện chiến của Alamanni (Đức) đến xâm chiếm vương quốc mình, ông đã nói với hoàng hậu Clotilda rằng ông sẽ theo đạo và chịu rửa tội nếu như Chúa phù hộ ông đánh thắng trận đấu này (trước đó hoàng hậu đã nhiều lần khuyên chồng mình vào đạo nhưng ông vẫn

không nghe). Cuối cùng, ông thắng thật và nhận fleur-de-lis làm biểu tượng. Tiếp đó, vào thế kỷ 12, vua Louis trở thành hoàng đế nước Pháp đầu tiên khắc họa hoa Diên Vĩ trên chiếc khiên của mình. Nữ anh hùng nước Pháp, Joan of Arc (Jeanne d'Arc) đã mang theo lá cờ trắng có biểu tượng Chúa hộ mệnh của hoàng gia (hoa Diên Vĩ) khi bà đánh bại quân Anh tại Orléans (1429).

Ở Nhật, hoa Diên Vĩ tượng trưng cho chí khí anh hùng và dòng dõi quý phái. Hoa Diên Vĩ là một phần quan trọng trong lễ hội mùa xuân dành cho các bé trai.

## 2. ĐỊNH NGHĨA BÀI TOÁN:

Với 1 bộ dữ liệu đồ sộ về hoa Iris đã có về các cá thể hoa thuộc các loài Iris cụ thể, ta không cần phải phân tích về gen hay các phân tích về sinh học phức tạp để nhận biết ra 1 cá thể hoa đang xét thuộc loài Iris nào, bằng các dữ liệu đã có sẵn và dựa vào các thuộc tính của cá thể, ta có thể đưa ra nhận biết nhanh chóng cá thể đó thuộc loài hoa Iris nào bằng các tính toán số học trên máy tính (Machine Learning)

## 3. THU THẬP VÀ TIỀN XỬ LÝ SỐ LIỆU:

Bộ dữ liệu về hoa Iris được thu thập bởi Edgar Anderson – 1 nhà thực vật học người mỹ - trên bán đảo [Gaspé](#) (là một bán đảo dọc theo bờ biển phía nam của sông Saint Lawrence ở Quebec, Canada, mở rộng vào vịnh Saint Lawrence)

Sau đó bộ dữ liệu trên được [Ronald Aylmer Fisher](#) – 1 nhà thống kê người Anh đồng thời là 1 nhà di truyền học rút gọn lại, bỏ đi các thuộc tính mang tính chất mô tả, không đo đạt được giá trị theo số học (như thời điểm gieo trồng, thời điểm nở hoa, màu sắc hoa, tốc độ tăng trưởng, điều kiện ánh sáng, đất nước và độ ẩm,...)

Bộ dữ liệu sau khi được rút gọn bao gồm 5 thuộc tính :

- 4 thuộc tính kiểu số: chiều dài đài hoa, chiều rộng đài hoa, chiều dài cánh hoa, chiều rộng cánh hoa (đơn vị là cm)
- 1 thuộc tính còn lại là tên của loài hoa Iris (có 3 loài tất cả : Iris Setosa, Iris Versicolour, Iris Virginica)

Theo tổng kết thống kê:

	Giá trị nhỏ nhất	Giá trị lớn nhất	Giá trị trung bình
Chiều dài đài hoa	4.3	7.9	5.84
Chiều rộng đài hoa	2.0	4.4	3.05
Chiều dài cánh hoa	1.0	6.9	3.76
Chiều rộng cánh hoa	0.1	2.5	1.20

Tỷ lệ phân chia cho mỗi loài trong 3 loài Iris trên là : 33,3%

### 3.1.LÀM SẠCH DỮ LIỆU( DATA CLEANING):

- Thiếu giá trị: khi xảy ra sự thiếu thông tin ở 1 thuộc tính nào đó trong 1 bản ghi của bộ dữ liệu thu thập, khi mà tính đảm bảo số bản ghi chia đều cho 3 loài trên đã có ( 33,3%) và số lượng bản ghi thiếu là ít ta có thể áp dụng phương pháp loại bỏ.(thực tế trong bộ dữ liệu trên, không tồn tại giá trị thuộc tính bị thiếu hay không được điền).mặt khác khi cần phải điền vào các giá trị thiếu, ta có thể áp dụng phương pháp điền tự động, thay thế các giá trị thiếu bằng giá trị trung bình đã được thống kê ở trên
- Nhiễu dữ liệu: khi xuất hiện 1 giá trị bất ngờ nào đó, đột nhiên vượt quá các giá trị biên đã được thống kê, ta có thể sửa lại giá trị đó thành các giá trị ở vùng biên theo bảng thống kê ở trên

### 3.2.CHỌN LỌC DỮ LIỆU (DATA SELECTION):

- Tích hợp và dư thừa dữ liệu: do dữ liệu trên thu thập từ 1 nguồn duy nhất nên việc tích hợp là không cần thiết, các thuộc tính của dữ liệu là độc lập nhau, không có mối quan hệ tương quan nào, các thuộc tính đã được rút gọn chọn lọc nên không cần việc phân tích dư thừa dữ liệu

### 3.3.BIẾN ĐỔI DỮ LIỆU (DATA TRANSFORMATION):

- Biến đổi dữ liệu: ta sử dụng phương pháp chuẩn hóa dữ liệu về khoảng giá trị [0,1], phần này sẽ được nói rõ hơn ở phần sau tùy vào phương pháp khai phá dữ liệu được lựa chọn

### 3.4.RÚT GỌN DỮ LIỆU (DATA REDUCTION):

- Như đã nói từ đầu, bộ dữ liệu đã được loại bỏ đi các thuộc tính mang tính chất mô tả, chỉ giữ lại 1 thuộc tính mô tả duy nhất là tên loài Iris cụ thể. các thuộc tính mang tính số học còn lại vẫn đảm bảo được tính phân loại (đã được [Ronald Aylmer Fisher](#) – 1 nhà di truyền học người Anh – chọn lọc)

### 4.KHAI PHÁ DỮ LIỆU (DATA MINING):

Dựa vào phát biểu và mục đích của bài toán, dễ dàng nhận thấy đây là 1 bài toán phân lớp: với 1 bộ dữ liệu đã cho và đã được gán nhãn các lớp cho trước (thể hiện thông qua thuộc tính loài Iris cụ thể), ta cần gán các mẫu mới (chưa biết thuộc nhãn lớp nào) vào các lớp với độ chính xác cao nhất có thể

Bộ dữ liệu được chọn có số thuộc tính không lớn, để đơn giản và dễ sử dụng để phân lớp ta lựa chọn dùng phân lớp theo thuật toán “*k – người hàng xóm gần nhất*” (*k-Nearest Neighbor Algorithm*)

#### 4.1.THUẬT TOÁN “K – NGƯỜI HÀNG XÓM GẦN NHẤT” LÀ GÌ:

Thuật toán *k – người hàng xóm gần nhất* (viết tắt là *K-NN*) là thuật toán có mục đích phân loại lớp cho một mẫu mới (Query point) dựa trên các thuộc tính và lớp của các mẫu sẵn có (Training Data), các mẫu này được nằm trong một hệ gọi là không gian mẫu.

Một đối tượng được phân lớp dựa vào *K* láng giềng của nó. *K* là số nguyên dương được xác định trước khi thực hiện thuật toán. Người ta thường dùng khoảng cách Euclidean để tính khoảng cách giữa các đối tượng.

#### 4.2.MÔ TẢ THUẬT TOÁN K-NN:

Các mẫu được mô tả bằng *n – chiều* thuộc tính số. Mỗi mẫu đại diện cho một điểm trong một chiều không gian *n – chiều*. Theo cách này tất cả các mẫu được lưu trữ trong một mô hình không gian *n – chiều*

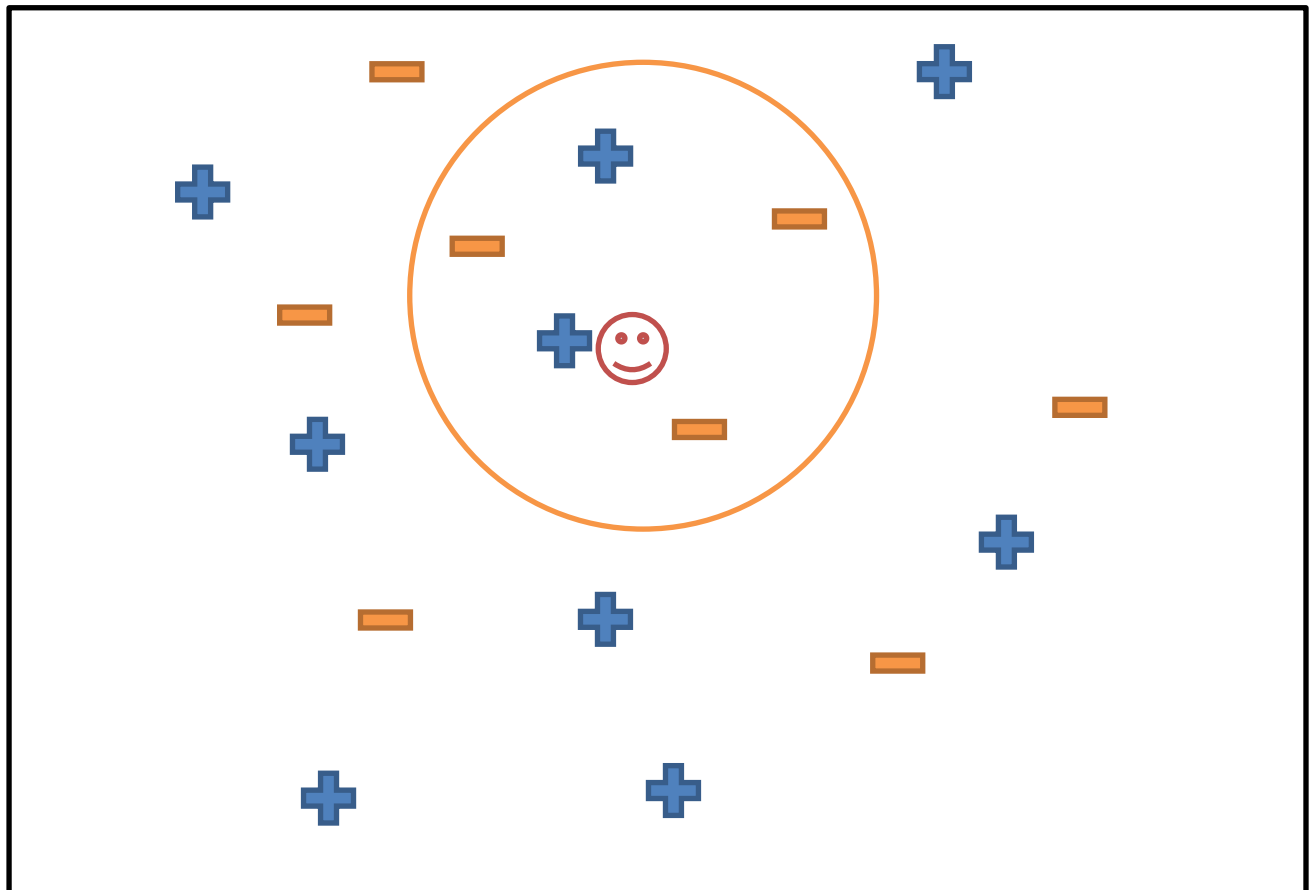
Các bước thực hiện của Thuật toán K-NN được mô tả như sau:

1. Xác định giá trị tham số *K* (số láng giềng gần nhất)

2. Tính khoảng cách giữa đối tượng cần phân lớp (Query Point) với tất cả các đối tượng trong training data (thường sử dụng khoảng cách Euclidean)
3. Sắp xếp khoảng cách theo thứ tự tăng dần và xác định K láng giềng gần nhất với Query Point
4. Lấy tất cả các lớp của K láng giềng gần nhất đã xác định
5. Dựa vào phần lớn lớp của láng giềng gần nhất để xác định lớp cho Query Point

Để hiểu K-NN được dùng để phân lớp thế nào ta xem minh họa dưới đây:

Trong hình dưới đây, training Data được mô tả bởi dấu **(+)** và dấu **(-)**, đối tượng cần được xác định lớp cho nó (Query point) là **hình mặt cười đỏ**. Nhiệm vụ của chúng ta là ước lượng (hay dự đoán) lớp của Query point dựa vào việc lựa chọn số láng giềng gần nhất với nó. Nói cách khác chúng ta muốn biết liệu Query Point sẽ được phân vào lớp **(+)** hay lớp **(-)**



**Ta thấy rằng:**

1-Nearest neighbor : Kết quả là + (Query Point được xếp vào lớp dấu +)

2-Nearest neighbors : không xác định lớp cho Query Point vì số láng giềng gần nhất với nó là 2 trong đó 1 là lớp + và 1 là lớp - (không có lớp nào có số đối tượng nhiều hơn lớp kia).vấn đề này sẽ được nói rõ cụ thể ở phần sau

5-Nearest neighbors : Kết quả là - (Query Point được xếp vào lớp dấu - vì trong 5 láng giềng gần nhất với nó thì có 3 đối tượng thuộc lớp - nhiều hơn lớp + chỉ có 2 đối tượng).

### **4.3.ÁP DỤNG THUẬT TOÁN K-NN VÀO BÀI TOÁN PHÂN LỚP HOA IRIS:**

“Sự gần gũi” được xác định bằng khoảng cách Euclide, với khoảng cách Euclide giữa hai điểm

$X=(x_1, x_2, x_3, \dots, x_n)$  và  $Y=(y_1, y_2, y_3, \dots, y_n)$  là :

$$D(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Sự phân loại lớp cho mẫu chưa biết lúc này dựa trên sự chiếm đa số của một lớp nào đó trong “ k – người hàng xóm gần nhất “ đã tìm được

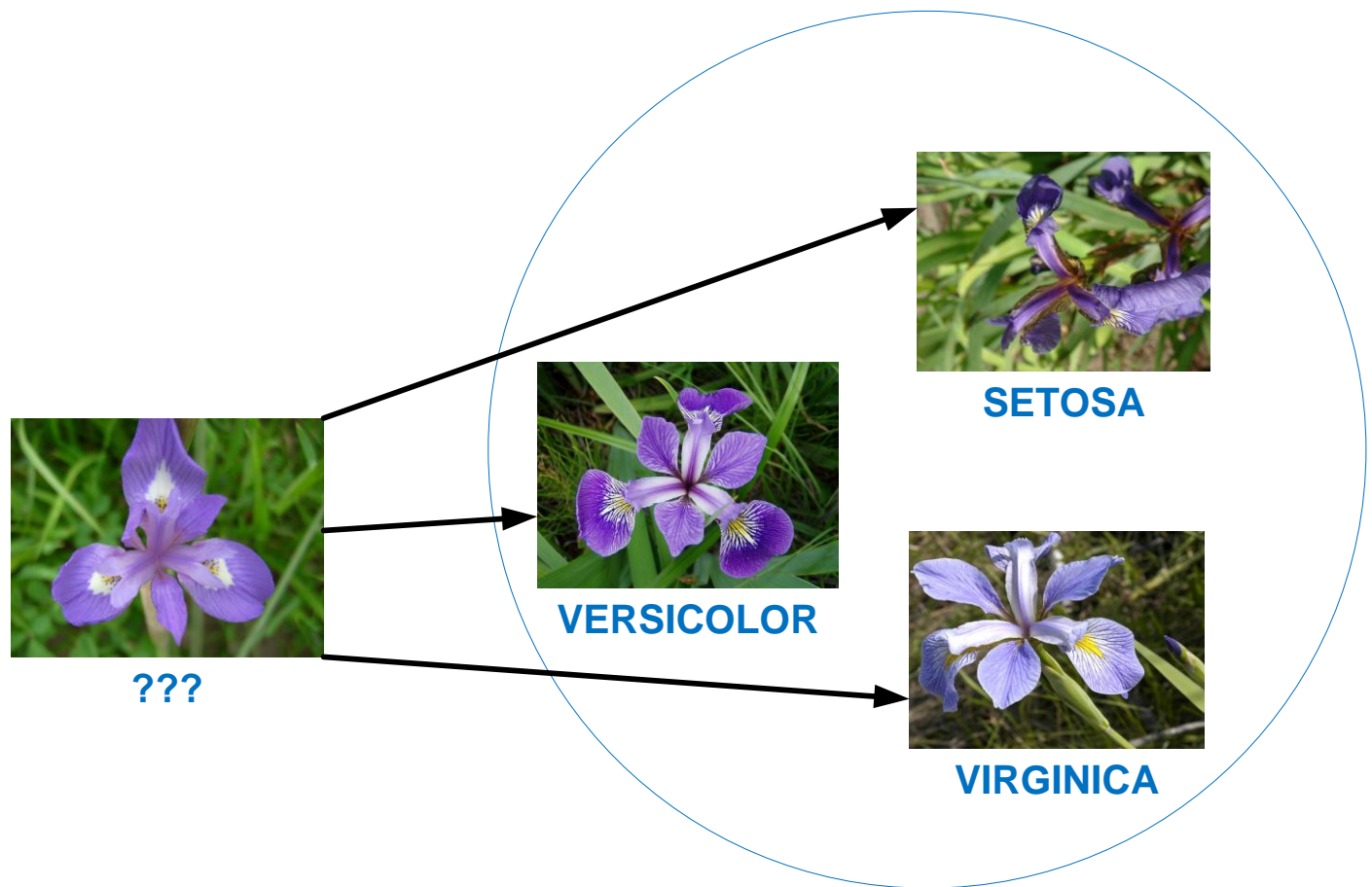
ở đây ta cần biến đổi dữ liệu để cho hiệu quả của thuật toán được nâng cao hơn.cần lưu ý rằng:

- Khoảng cách Euclide bị ảnh hưởng bởi khoảng giới hạn giá trị thuộc tính
- Khoảng giá trị lớn thì gây ra nhiều ảnh hưởng hơn so với khoảng giá trị nhỏ

Do vậy khoảng cách giữa các mẫu có thể bị chi phối bởi một số thuộc tính có giá trị tương đối lớn so với thuộc tính khác.

Vì vậy ta có thể khắc phục bằng cách bình thường hóa một số thuộc tính bằng cách chia giá trị của mỗi thuộc tính thuộc mỗi mẫu cho giá trị lớn nhất của thuộc tính đó trong tất cả các mẫu, làm các giá trị sau khi bình thường hóa nằm trong khoảng [0,1]

Áp dụng vào bài toán phân lớp hoa Iris:



Trích dẫn từ bộ dữ liệu hoa Iris, ta lập ra 1 bảng dữ liệu nhỏ và tiến hành thực hiện theo thuật toán K-NN để minh họa:

Chiều dài đài hoa	Chiều rộng đài hoa	Chiều dài cánh hoa	Chiều rộng cánh hoa	Tên loài
5.1	3.5	1.4	0.2	<b>Iris-setosa</b>
4.7	3.2	1.3	0.2	<b>Iris-setosa</b>
7.0	3.2	4.7	1.4	<b>Iris-versicolor</b>
6.3	3.3	6.0	2.5	<b>Iris-virginica</b>
5.9	3.0	4.2	1.5	<b>Iris-versicolor</b>
5.1	3.8	1.6	0.2	<b>???</b>

Ta tiến hành thường hóa dữ liệu, bảng dữ liệu sẽ bị biến đổi thành bảng sau:



Chiều dài đài hoa	Chiều rộng đài hoa	Chiều dài cánh hoa	Chiều rộng cánh hoa	Tên loài
$5.1/7.0 = 0.73$	$3.5/3.8 = 0.92$	$1.4/6.0 = 0.23$	$0.2/2.5 = 0.08$	<b>Iris-setosa</b>
$4.7/7.0 = 0.67$	$3.2/3.8 = 0.84$	$1.3/6.0 = 0.22$	$0.2/2.5 = 0.08$	<b>Iris-setosa</b>
$7.0/7.0 = 1$	$3.2/3.8 = 0.84$	$4.7/6.0 = 0.78$	$1.4/2.5 = 0.56$	<b>Iris-versicolor</b>
$6.3/7.0 = 0.9$	$3.3/3.8 = 0.87$	$6.0/6.0 = 1$	$2.5/2.5 = 1$	<b>Iris-virginica</b>
$5.9/7.0 = 0.84$	$3.0/3.8 = 0.79$	$4.2/6.0 = 0.7$	$1.5/2.5 = 0.6$	<b>Iris-versicolor</b>
$5.1/7.0 = 0.73$	$3.8/3.8 = 1$	$1.6/6.0 = 0.27$	$0.2/2.5 = 0.08$	<b>???</b>

Ta tính toán khoảng cách Euclide được bằng khoảng cách tương ứng :

Tên loài	Khoảng cách
<b>Iris-setosa</b>	$\sqrt{(0.73 - 0.73)^2 + (0.92 - 1)^2 + (0.23 - 0.27)^2 + (0.08 - 0.08)^2} = 0.008$
<b>Iris-setosa</b>	$\sqrt{(0.67 - 0.73)^2 + (0.84 - 1)^2 + (0.22 - 0.27)^2 + (0.08 - 0.08)^2} = 0.178$
<b>Iris-versicolor</b>	$\sqrt{(1 - 0.73)^2 + (0.84 - 1)^2 + (0.78 - 0.27)^2 + (0.56 - 0.08)^2} = 0.767$
<b>Iris-virginica</b>	$\sqrt{(0.9 - 0.73)^2 + (0.87 - 1)^2 + (1 - 0.27)^2 + (1 - 0.08)^2} = 1.194$
<b>Iris-versicolor</b>	$\sqrt{(0.84 - 0.73)^2 + (0.79 - 1)^2 + (0.7 - 0.27)^2 + (0.6 - 0.08)^2} = 0.715$

➤ Chọn K = 3

Ta chọn K hàng xóm gần nhất là 3 thì 3 khoảng cách nhỏ nhất lần lượt là : 0.008, 0.178, 0.715 tương ứng với tên loài: Iris-setosa, Iris-setosa và Iris-versicolor

Trong 3 hàng xóm gần nhất này có 2 hàng xóm là Iris-setosa và 1 hàng xóm là Iris-versicolor. vậy ta kết luận loài cần phân lớp thuộc loài Iris-setosa

➤ Chọn K = 4

Trong trường hợp ta chọn K hàng xóm gần nhất là 4 thì trong 4 hàng xóm đấy có 2 hàng xóm là Iris-setosa và 2 hàng xóm là Iris-versicolor

Do số hàng xóm ở đây là bằng nhau (đều bằng 2).ta tiến hành thêm 1 bước nữa là đi tính trung bình khoảng cách của mỗi hàng xóm rồi so sánh các khoảng cách trung bình đó với nhau rồi chọn ra khoảng cách nhỏ nhất và lấy loài hoa đại diện cho khoảng cách được chọn gắn nhãn cho cá thể đang xét.

Như ví dụ trên ta đi tính khoảng cách trung bình nên ta có bảng:

Tên loài	Khoảng cách trung bình
Iris-setosa	$(0.008+0.178) / 2 = 0.093$
Iris-versicolor	$(0.767+0.715) / 2 = 0.741$

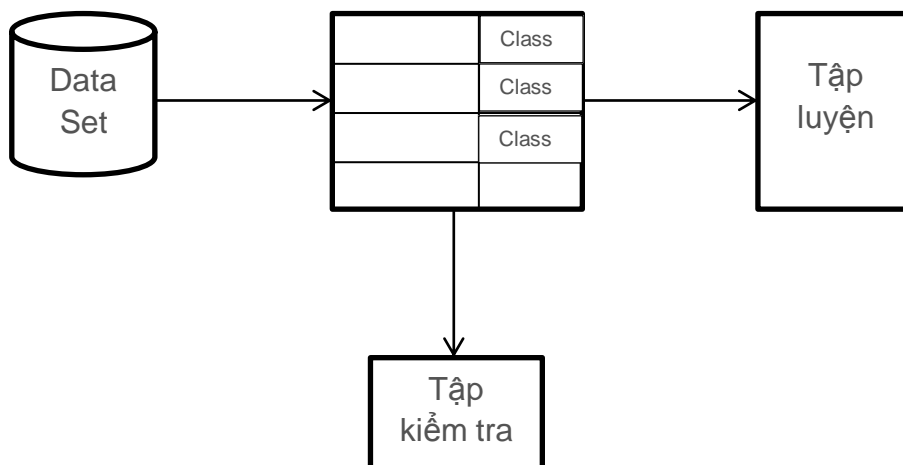
Ta chọn khoảng cách trung bình nhỏ nhất là 0.093 tương ứng với loài hoa: Iris-setosa.ta kết luận cá thể hoa đang xét thuộc loài Iris-setosa

## 5.ĐÁNH GIÁ ĐỘ CHÍNH XÁC CỦA MÔ HÌNH VÀ CHƯƠNG TRÌNH CHẠY:

### 5.1.CÁCH THỨC ĐÁNH GIÁ ĐỘ CHÍNH XÁC CỦA MÔ HÌNH:

Ước lượng độ chính xác của bộ phân lớp là quan trọng bởi nó cho phép dự đoán được độ chính xác của các kết quả phân lớp những dữ liệu trong tương lai. Phương pháp được thực hiện qua hai bước sau:

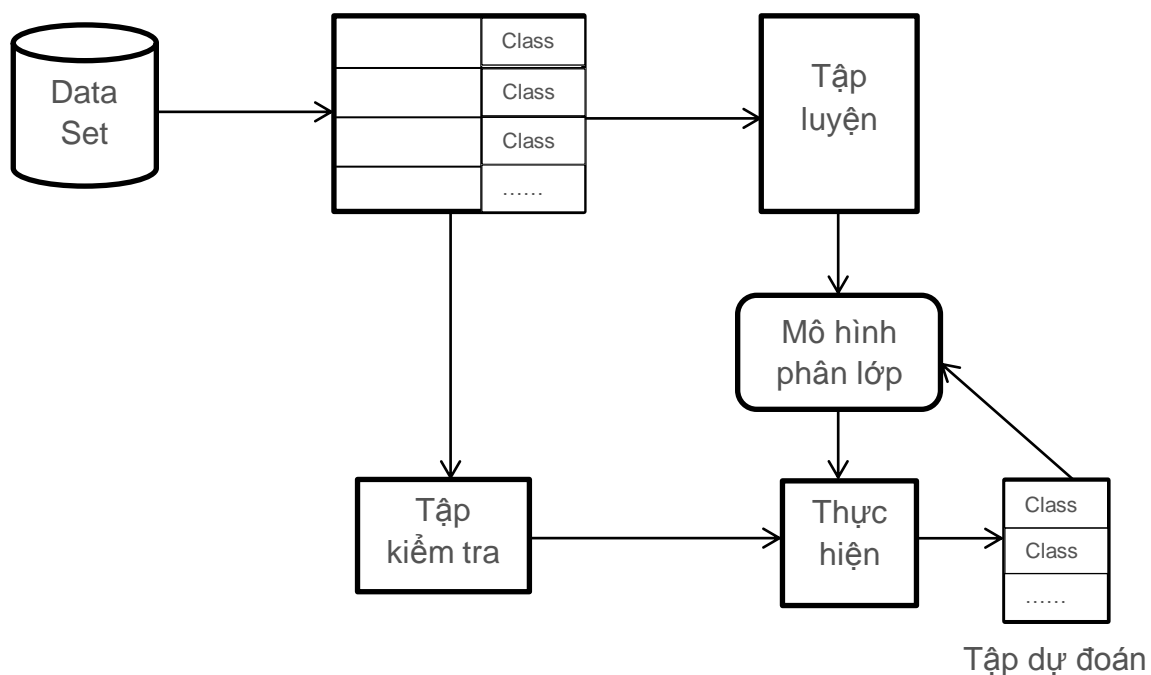
**Bước 1:** phân dữ liệu một cách ngẫu nhiên vào hai tập là tập luyện và tập kiểm tra.



Một cách tổng quát, tập luyện lớn hơn thì việc phân lớp tốt hơn, tập kiểm tra lớn hơn sẽ đúng hơn về việc đánh giá lỗi. Thông thường, người ta thường chọn tập luyện bằng 2/3 và tập kiểm tra bằng 1/3.

Trong trường hợp số phần tử của các lớp trong tập dữ liệu gốc là không đều, ví dụ như về y học thì lớp khỏe mạnh chiếm 90% còn lớp bệnh tật chiếm 10%, như vậy nếu vẫn chọn ra ngẫu nhiên tập kiểm tra thì sẽ không đánh giá đúng tỷ lệ lỗi. Để khắc phục, ta chọn ngẫu nhiên ở mỗi lớp ra số phần tử cho tập kiểm tra bằng 1/3 số phần tử mỗi lớp.

## Bước 2: Đánh giá mô hình thông qua tập kiểm tra



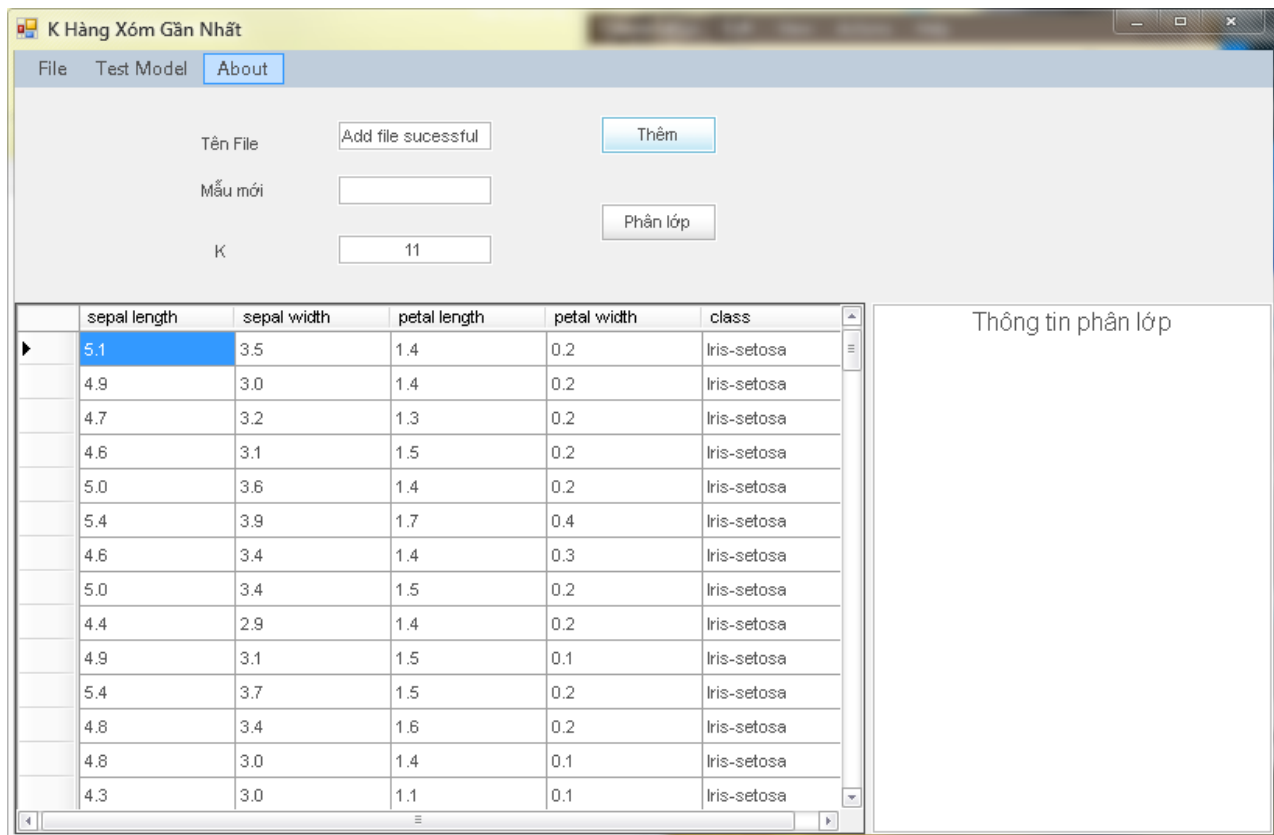
Tỷ lệ lỗi:

$$Error\ rate = \frac{Số\ mẫu\ phân\ sai}{Tổng\ số\ mẫu\ kiểm\ tra} \times 100\%$$

Ta đánh giá sự chính xác của mô hình thông qua tỷ lệ lỗi. Tỷ lệ lỗi càng nhỏ, mô hình có độ chính xác càng cao, có tính phù hợp tốt.

## 5.2.CHƯƠNG TRÌNH CHẠY:

- Giao diện chính:



- Nhập mẫu cần phân lớp (các thuộc tính được nhập vào ngăn cách nhau bởi dấu chấm phẩy):

K Hàng Xóm Gần Nhất

File Test Model About

Tên File Add file sucessful Thêm

Mẫu mới 5.1;3.8;1.6;0.2 Phân lớp

K 11

	sepal length	sepal width	petal length	petal width	class
▶	5.1	3.5	1.4	0.2	Iris-setosa
	4.9	3.0	1.4	0.2	Iris-setosa
	4.7	3.2	1.3	0.2	Iris-setosa
	4.6	3.1	1.5	0.2	Iris-setosa
	5.0	3.6	1.4	0.2	Iris-setosa
	5.4	3.9	1.7	0.4	Iris-setosa
	4.6	3.4	1.4	0.3	Iris-setosa
	5.0	3.4	1.5	0.2	Iris-setosa
	4.4	2.9	1.4	0.2	Iris-setosa
	4.9	3.1	1.5	0.1	Iris-setosa
	5.4	3.7	1.5	0.2	Iris-setosa
	4.8	3.4	1.6	0.2	Iris-setosa
	4.8	3.0	1.4	0.1	Iris-setosa
	4.3	3.0	1.1	0.1	Iris-setosa

Thông tin phân lớp

- Hiện thị thông tin phân lớp:

K Hàng Xóm Gần Nhất

File Test Model About

Tên File Add file sucessful Thêm

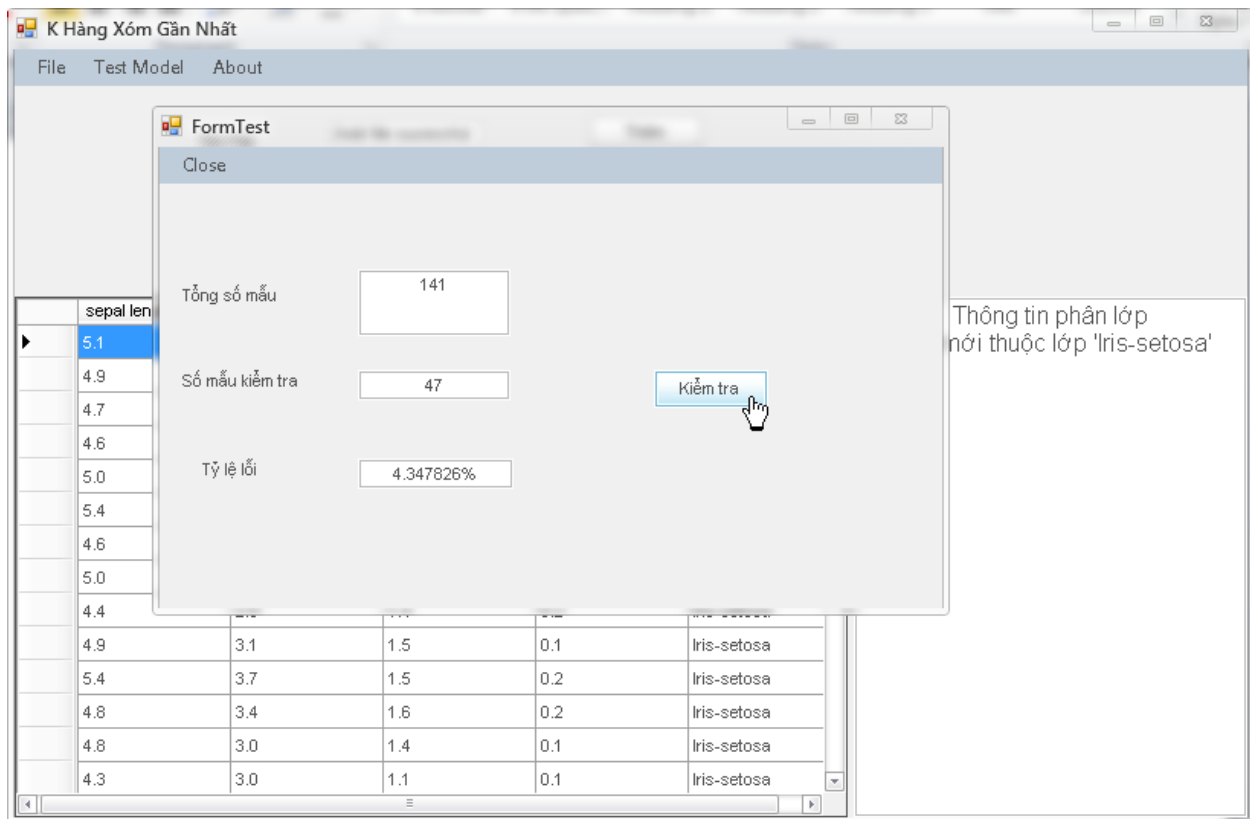
Mẫu mới 5.1;3.8;1.6;0.2 Phân lớp

K 11

	sepal length	sepal width	petal length	petal width	class
▶	5.1	3.5	1.4	0.2	Iris-setosa
	4.9	3.0	1.4	0.2	Iris-setosa
	4.7	3.2	1.3	0.2	Iris-setosa
	4.6	3.1	1.5	0.2	Iris-setosa
	5.0	3.6	1.4	0.2	Iris-setosa
	5.4	3.9	1.7	0.4	Iris-setosa
	4.6	3.4	1.4	0.3	Iris-setosa
	5.0	3.4	1.5	0.2	Iris-setosa
	4.4	2.9	1.4	0.2	Iris-setosa
	4.9	3.1	1.5	0.1	Iris-setosa
	5.4	3.7	1.5	0.2	Iris-setosa
	4.8	3.4	1.6	0.2	Iris-setosa
	4.8	3.0	1.4	0.1	Iris-setosa
	4.3	3.0	1.1	0.1	Iris-setosa

Thông tin phân lớp  
Mẫu mới thuộc lớp 'Iris-setosa'

- Đánh giá tỷ lệ lỗi (nhấn chuột vào **Test Model**):



## TÀI LIỆU THAM KHẢO

- ✓ Slide bài giảng môn DATA MINING – NGUYỄN HÙNG SƠN, ĐẠI HỌC CÔNG NGHỆ
- ✓ Ebook môn khai phá dữ liệu – KHOA CÔNG NGHỆ THÔNG TIN, ĐẠI HỌC HÀNG HẢI
- ✓ Slide bài giảng môn KHAI THÁC DỮ LIỆU & ỨNG DỤNG (DATA MINING) – ThS. NGUYỄN HOÀNG ANH TÚ
- ✓ Các bài viết thông tin về hoa Iris trong trang <http://en.wikipedia.org>
- ✓ Thông tin và bộ nguồn dữ liệu về hoa Iris trong trang <http://archive.ics.uci.edu/ml/datasets/Iris>