ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA CÔNG NGHỆ THÔNG TIN

BÁO CÁO ĐỒ ÁN

QUAN HÓA DỮ LIỆU

Môn học: **Trực quan hóa dữ liệu**

Thành viên nhóm:

- **Lê Nguyễn Kiều Oanh** **MSSV: 21127129**
- **Đoàn Ngọc Mai** **MSSV: 21127104**

Giảng viên hướng dẫn:

- **Bùi Tiến Lên**
- **Lê Ngọc Thành**
- **Lê Nguyễn Nhật Trường**

Thành phố Hồ Chí Minh, ngày 16 tháng 4 năm 2024

Mục lục

I. Thông tin nhóm

Họ và tên	Mã số sinh viên
Đoàn Ngọc Mai	21127104
Lê Nguyễn Kiều Oanh	21127129

II. Phân công công việc

Họ và tên	Mã số	Công việc	Mức độ hoàn thành
Đoàn Ngọc Mai	21127104	Hoàn thành mục EDA 3D và đặt câu hỏi	100%
		Hoàn thành phần Insight	100%
		Viết Báo cáo và trang trí lại notebook	100%
Lê Nguyễn Kiều Oanh	21127129	Hoàn thành Data Understanding	100%
		Hoàn thành EDA 1D	100%
		Hoàn thành EDA 2D	100%

III. Những phát hiện ý nghĩa

- Từ histogram, có thể thấy rằng các biến độ dài và độ rộng của cánh hoa **PetalLengthCm** và **PetalWidthCm** có xu hướng phân chia thành các nhóm riêng biệt, có lẽ phản ánh sự khác biệt giữa các loài trong nhóm Iris. Trong khi đó, độ dài và độ rộng của đài hoa **SepallengthCm** và **SepalwidthCm** có phân phối ít rõ ràng hơn nhưng vẫn có sự tập trung dữ liệu ở một số khoảng giá trị nhất định.

- Sự phân phối đều nhau của số lượng mẫu cho mỗi loài mỗi loài có **50** mẫu cho thấy tập dữ liệu này được cân bằng tốt giữa các loài, điều này rất hữu ích cho các phân tích đa dạng sinh học và máy học, bởi vì mỗi loài được đại diện bằng một số lượng mẫu như nhau.

- Giữa độ dài và độ rộng của cánh hoa, cũng như giữa chiều dài của đài hoa và cánh hoa có mối tương quan mạnh mẽ.

- **SepalWidthCm** có thể được sử dụng như một đặc trưng phân biệt **Iris-setosa** với hai loài còn lại trong một mô hình phân loại hoa Iris.

- **PetalLengthCm** là một biến đáng giá trong việc xác định loài của hoa Iris, và nó có thể được sử dụng để phát triển các mô hình phân loại có độ chính xác cao.

- Trong khi chiều dài đài hoa có liên quan mật thiết đến chiều dài cánh hoa, chiều rộng đài hoa không cho thấy mối liên kết rõ ràng đối với chiều dài cánh hoa trong phạm vi dữ liệu này.

- Cả **PetalLengthCm** và **PetalWidthCm** đều có thể được sử dụng làm chỉ số dự đoán mạnh mẽ cho nhau và cả hai đều liên quan đến **SepalLengthCm**. Thông tin này có thể được sử dụng để xây dựng các mô hình dự đoán hoặc phân loại chính xác hơn cho các loài Iris dựa trên kích thước của đài hoa và cánh hoa.

- Kích thước cánh hoa là một đặc điểm phân loại tốt giữa ba loài Iris này. Độ dài và độ rộng của cánh hoa có thể được sử dụng như một chỉ số sinh học để phân biệt chúng trong nghiên cứu sinh học và ứng dụng thực tế.

- Mặc dù **SepalWidthCm** có thể không phải là đặc điểm phân loại rõ ràng giữa Iris-versicolor và Iris-virginica, **SepalLengthCm** lại có tiềm năng làm đặc điểm phân loại tốt giữa ba loài hoa này, đặc biệt khi kết hợp cả hai kích thước đài hoa.

- **PetalWidthCm** có khả năng phân biệt tốt giữa **Iris-setosa** và hai loài còn lại, nhưng phân biệt giữa **Iris-versicolor** và **Iris-virginica** dựa trên **PetalWidthCm** và **SepalWidthCm** có thể không rõ ràng. Phát hiện thú vị là sự phân chia rõ rệt của Iris-setosa dựa trên cả hai đặc điểm, điều này có ý nghĩa trong việc phân loại loài dựa trên đặc điểm hình thái của hoa.

- Sự phân tách rõ rệt của **Iris-setosa** so với hai loài khác, không chỉ bởi kích thước nhỏ của cả đài hoa và cánh hoa, mà còn bởi sự không chồng lấp dữ liệu giữa chúng, cho thấy sự đa dạng sinh học đáng kể giữa các loài trong chi Iris.

- Chiều dài đài hoa có thể là một đặc trưng có ích trong việc phân loại **Iris-virginica** so với hai loài còn lại. Tuy nhiên, sự phân biệt giữa **Iris-setosa** và

Iris-versicolor dựa trên chỉ số này có thể không rõ ràng và cần thêm thông tin từ các đặc trưng khác để phân biệt chính xác hơn.

- **Iris-setosa** có đặc điểm cánh hoa độc đáo so với hai loài khác, với tỷ lệ chiều dài và chiều rộng cánh hoa cao đáng kể. Điều này có thể là do đặc điểm di truyền hoặc thích nghi sinh học mà loài này có, phản ánh sự khác biệt sinh học đáng kể giữa các loài trong chi Iris. Sự biến thiên lớn trong tỷ lệ này cho Iris-setosa cũng gợi ý rằng có sự đa dạng về hình thái cánh hoa trong loài, có thể liên quan đến các biến số sinh thái khác nhau.