

BÁO CÁO BÀI TẬP THỰC HÀNH MÔN HỌC MÁY THỐNG KÊ (DS102.K21)

Giảng viên phụ trách môn : Thầy NGUYỄN TẤN TRẦN MINH KHANG – VÕ DUY NGUYỄN

Họ và tên sinh viên: Võ Hoàng Thông - MSSV: 18521462

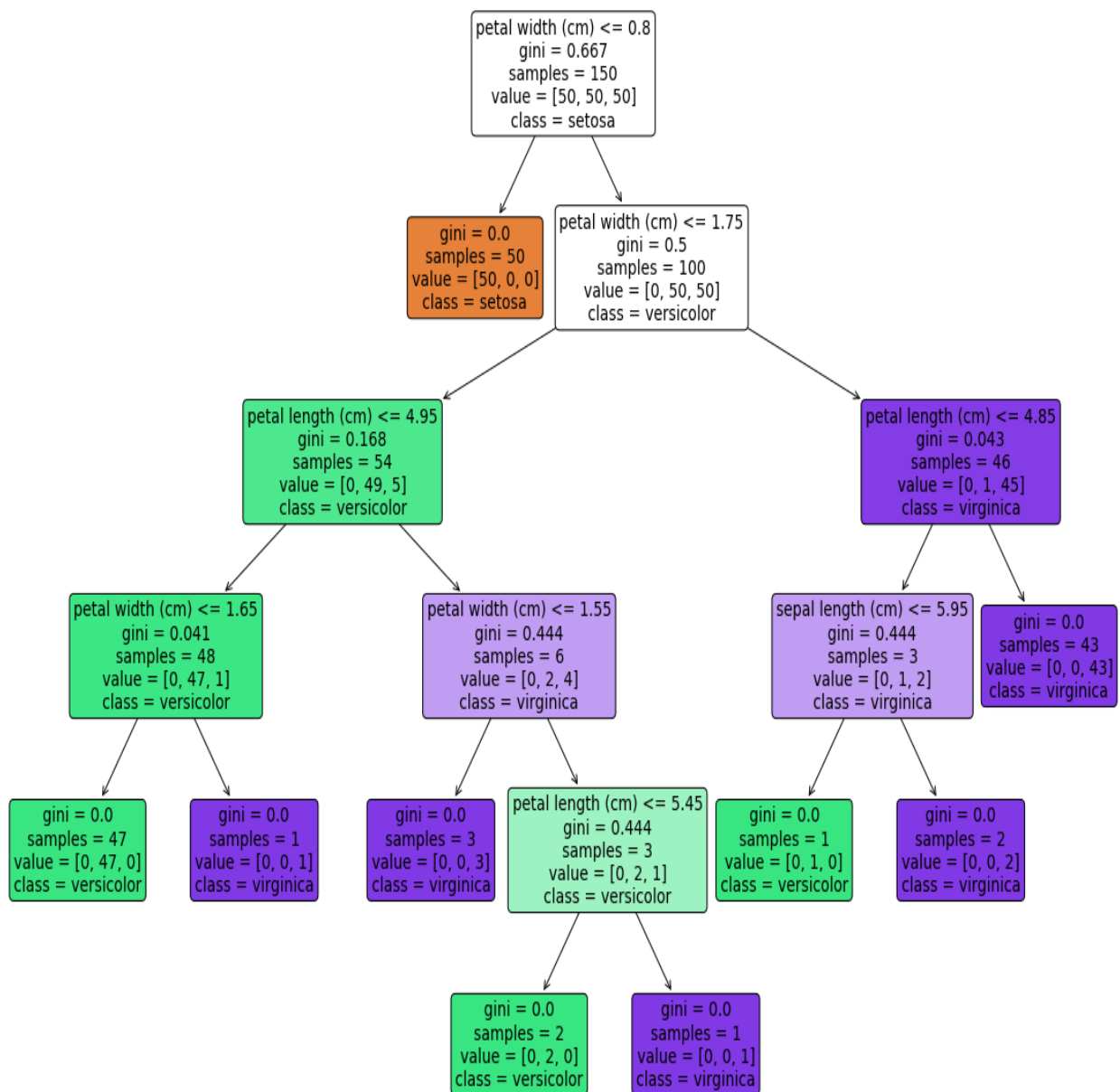
Bài toán: Yêu cầu dựa vào 4 thuộc tính: chiều dài, chiều rộng đài hoa (sepal), và chiều dài, chiều rộng cánh hoa (petal).

Phân lớp ba loại hoa Iris: Iris setosa, Iris virginica và Iris versicolor.

Phương pháp: Sử dụng thuật toán phân lớp Decision Tree.

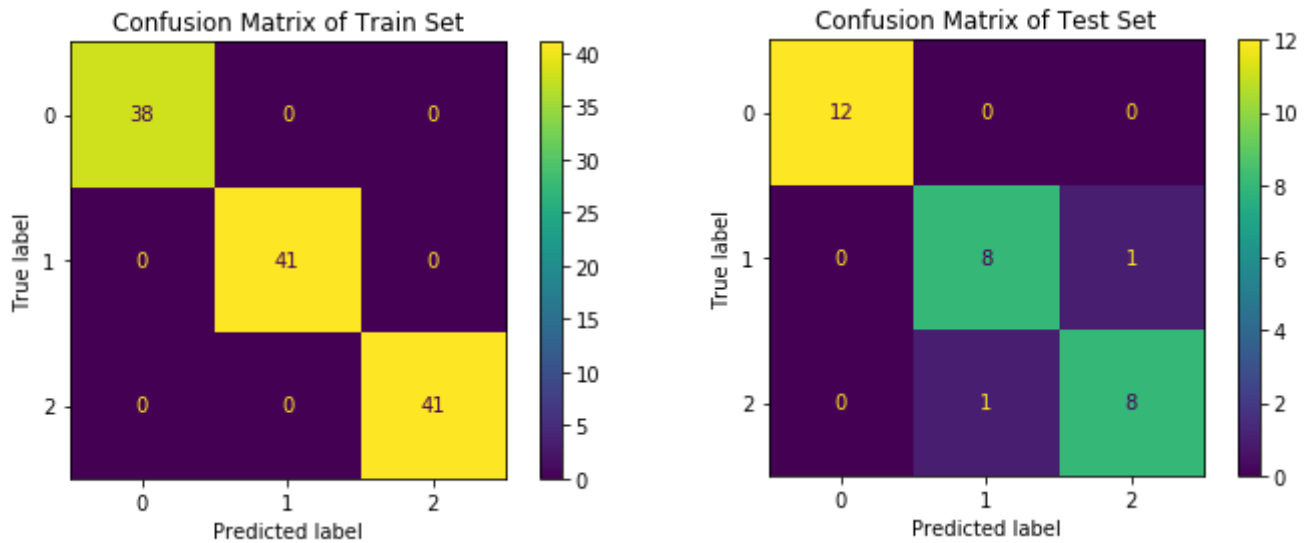
Source File: File đính kèm (Iris-Flower.ipynb)

Dataset: Bao gồm 150 điểm dữ liệu, chia đều cho ba loại hoa.



Hình ảnh cây quyết định cho bài toán Iris Flower

Confusion matrix



Đánh giá model

Accuracy Train Set 1.0

Accuracy Test Set 0.9333333333333333

Nhận xét:

- + Thuật toán phân lớp Decision Tree hoạt động tốt trên tập training set Iris Flower với kết quả chính xác 100%.
- + Kết quả tập test là 0.93 tương đồng với kết quả trên confusion matrix.
- + Bởi vì đây là cơ sở dữ liệu dễ, chỉ với 150 điểm dữ liệu, nên ta có thể thu được kết quả tốt như vậy. Giả sử ta có nhiều hơn những điểm dữ liệu trên, kết quả ta nhận được sẽ là một số khác, có thể là tốt và có thể là tệ.