

BÁO CÁO BÀI TẬP LÝ THUYẾT 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 2 thuộc tính:

+ Độ tuổi (Age).

+ Mức lương ước đoán (Estimated Salary).

Dự đoán khách hàng sẽ mua hàng hay không?

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

Source File: File đính kèm (Social-Network-Ads.ipynb)

Trực quan hóa dữ liệu

Trục x: Age

Trục y: Estimated Salary

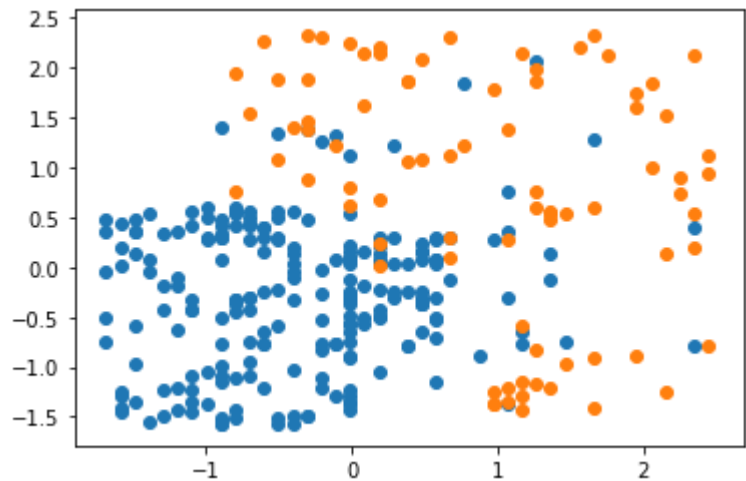
Dữ liệu đã được chuẩn hóa về dạng:

+ Kỳ vọng bằng 0.

+ Phương sai bằng 1.

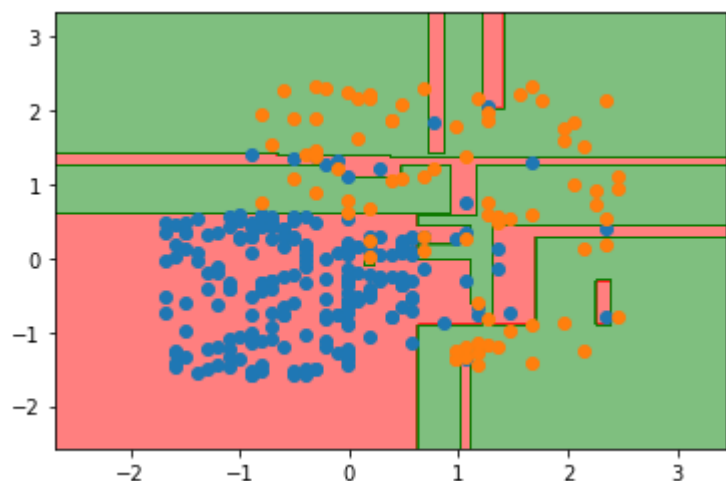
Mảng bên trái dưới màu xanh là các khách hàng không mua hàng.

Mảng bên phải và màu bên trên là màu cam là những khách hàng có mua hàng.



Trực quan hóa dữ liệu huấn luyện

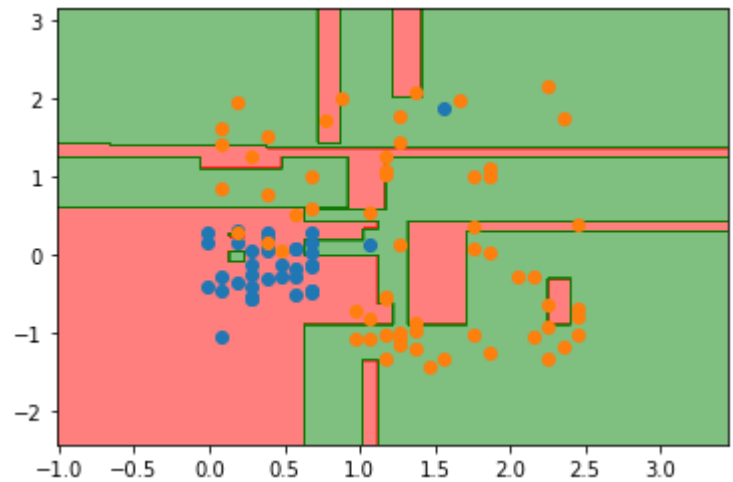
Các điểm dữ liệu phân bố tương đối tốt, tuy nhiên vẫn có một số điểm phân bố chưa tốt và nhiễu.



Trực quan hóa kết quả mô hình

Các điểm dữ liệu phân bố tương đối tốt với tỉ lệ đánh giá R^2 có độ đo là 80%.

Còn training model là gần tuyệt đối, 0.99333.



Confusion matrix của training và testing model

