

Machine Learning

Lecturer: Doctor Bui Thanh Hung

Data Science Laboratory

Faculty of Information Technology

Industrial University of Ho Chi Minh city

Email: hung.buithanhcs@gmail.com (buithanhhung@iuh.edu.vn)

Website: <https://sites.google.com/site/hungthanhbui1980/>

Bài 1:

Dựa trên bộ dữ liệu Iris được cung cấp ở đường link <https://archive.ics.uci.edu/dataset/53/iris> hãy thực hiện các yêu cầu sau:

1. Trực quan hóa dữ liệu với các biểu đồ box, histogram và các phân tích thống kê với 4 đặc trưng đầu
2. Chuyển đổi các lớp thành số một cách tự động
3. Sử dụng 4 đặc trưng của bộ dữ liệu, chia bộ dữ liệu thành 2 phần Train/Test với tỉ lệ 7:3, cho biết trong bộ dữ liệu đã cho có dữ liệu nào null không, phân tích bộ dữ liệu theo từng lớp qua các biểu đồ.
4. Đánh giá bộ dữ liệu với giải thuật KNN bằng phương pháp 5-Fold (k-fold)
5. Huấn luyện dữ liệu cho bài toán phân lớp sử dụng 4 đặc trưng trên với các giải thuật Bayes, SVM, Linear Regression và Decision Tree.
6. Tính độ đo F1 score
7. Tính độ đo Accuracy
8. Tính độ đo Confusion Matrix
9. So sánh kết quả các độ đo 6,7,8 với các giải thuật học máy ở trên
10. Lưu model với giải thuật đạt kết quả tốt nhất
11. Xây dựng ứng dụng với đầu vào là dữ liệu, có thể tự gõ hay từ 1 file, in kết quả ra màn hình

Bài 2:

Hãy tạo một tài khoản ChatGPT và tìm câu trả lời cho 2 câu hỏi sau từ ChatGPT:

1. Hãy cho biết kiến trúc tổng quát của ChatGPT và từng thành phần cụ thể của các phần trong kiến trúc này.
2. Những tiến bộ gần đây của Deep Learning và 40 bài báo về chủ đề ứng dụng deep learning trong phát hiện đối tượng trong ảnh (20 bài) và dịch máy (20 bài) có thời gian publish trong năm 2021-2022 và phân tích các phương pháp của từng bài báo