

# Machine Learning

Bộ môn Khoa học dữ liệu  
Khoa Công nghệ thông tin  
Trường Đại học Công nghiệp thành phố Hồ Chí Minh-IUH

Cho bộ dữ liệu ở đường link

<https://drive.google.com/file/d/1-xM3vbO5D9sO71Ord2ip6UVzevdyhdKs/>

hãy thực hiện các yêu cầu sau:

1. Sử dụng thống kê mô tả, mô tả về bộ dữ liệu trên
2. Trực quan hóa dữ liệu với các biểu đồ box, histogram, scatter matrix, grid line
3. Cho biết những dữ liệu còn thiếu và đề xuất cách xử lý dữ liệu thiếu đó
4. Chia bộ dữ liệu thành 2 phần Train/Test với các tỉ lệ nhất định
5. Đánh giá bộ dữ liệu với giải thuật KNN bằng phương pháp 10-Fold (k-fold)
6. Huấn luyện dữ liệu cho bài toán phân lớp sử dụng với các giải thuật Decision Tree, Bayes, SVM và XGBoost.
7. Tính độ đo F1 score
8. Tính độ đo Accuracy
9. Tính độ đo Confusion Matrix
10. So sánh kết quả các độ đo 6,7,8 với các giải thuật học máy ở trên
11. Lưu model với giải thuật đạt kết quả tốt nhất
12. Xây dựng ứng dụng với đầu vào là 1 dữ liệu hay từ 1 file, in kết quả ra màn hình.