BÀI TẬP NHÓM (Group Project)

Cho trước bộ dữ liệu (dataset(s) và đường dẫn (link) có liên quan.

1) **Tìm hiểu bộ dữ liệu này (data exploration) và đưa ra một báo cáo về chất lượng dữ liệu (data quality report).**

Chú ý:
- trình bày việc chuyển đổi bộ dữ liệu này về dạng tập tin dữ liệu thông dụng (csv, xls) nếu cần.
- trình bày cách thức xử lý về các giá trị dữ liệu bị thiếu (missing values) và giá trị dữ liệu ngoại biệt (outliers) nếu có.
- vận dụng các kỹ thuật trực quan hóa (visualization) để thể hiện bộ dữ liệu đã cho một cách phù hợp nếu cần.

2) **Trình bày chi tiết mục tiêu của bài toán khoa học dữ liệu trên bộ dữ liệu này.**

Chú ý:
- trình bày các thuộc tính thông thường (features/attributes/predictors), và thuộc tính đích/kết quả(class label/target/response/outcome) một cách phù hợp.
- trình bày việc chuẩn hóa dữ liệu (normalization) nếu có.

3) **Dùng ít nhất 3 nhiệm vụ/giải thuật (tasks/algorithms) để thực hiện và so sánh kết quả (performance) của mô hình** (vận dụng các công cụ đánh giá có liên quan như - confusion matrix, accuracy, precision, error, recall, ROC, AUC, lift curves, v.v nhiều nhất đến có thể).

Chú ý:
- trình bày các thuộc tính dữ liệu (data attributes/features) bỏ qua không dùng đến trong nhiệm vụ/giải thuật đã chọn, nếu có.
- trình bày các thuộc tính dữ liệu (data attributes) bổ sung vào nếu có (không dùng trong nhiệm vụ/giải thuật đã chọn nhưng có thể hỗ trợ việc thảo luận và diễn dịch kết quả).
- đối với từng nhiệm vụ/giải thuật, trình bày các thông số của các toán tử (operator parameters) một cách phù hợp (nếu có khác với các giá trị mặc định).
- điều chỉnh tỉ lệ training và test dataset nếu cần và so sánh kết quả (performance) của mô hình (vận dụng các công cụ đánh giá có liên quan).

4) **Thảo luận và diễn dịch kết quả, chỉ ra các hàm ý và/hay đề xuất áp dụng.**

Chú ý:
- vận dụng các kỹ thuật trực quan hóa (visualization) để thể hiện kết quả một cách phù hợp nếu có thể.

----------HẾT----------