

BÀI TẬP KIỂM TRA BUỔI 2 – CÂY QUYẾT ĐỊNH

Sáng 4 – 4/9/2019

1. Xây dựng cây quyết định dựa vào **chỉ số độ lợi thông tin** và dự đoán nhãn
 - a. Đọc dữ liệu từ tập dữ liệu đánh giá chất lượng rượu vang **trắng** trên trang UCI <https://archive.ics.uci.edu/ml/datasets/wine+quality>
 - b. Tập dữ liệu có bao nhiêu phần tử, Có bao nhiêu nhãn? Ghi chú kết quả trong file code sau câu lệnh.
 - c. Sử dụng **8 phần để** xây dựng mô hình và sử dụng **2 phần để** thực hiện đánh giá mô hình. Anh/chị ghi chú lại số lượng phần tử trong tập test và nhãn của các phần tử thuộc tập test trong file code
 - d. Xây dựng mô hình cây quyết định dựa trên tập dữ liệu học tạo ra ở bước c.
 - e. Đánh giá độ chính xác tổng thể và độ chính xác của từng lớp cho toàn bộ dữ liệu trong tập test
 - f. Đánh giá độ chính xác tổng thể và độ chính xác của từng lớp cho 6 phần tử đầu tiên trong tập test
2. Cho tập dữ liệu gồm 5 phần tử như bảng bên dưới,

STT	Chiều cao	Độ dài mái tóc	Giọng nói	Nhãn
1.	180	15	0	Nam
2.	167	42	1	Nữ
3.	136	35	1	Nữ
4.	174	15	0	Nam
5.	141	28	1	Nữ

- Sử dụng 5 phần tử trong tập dữ liệu trên để xây dựng mô hình dựa vào **chỉ số độ lợi thông tin** với thuộc tính: chiều cao, độ dài tóc và giọng nói để dự đoán nhãn là nam giới hay nữ giới.
- Dự báo phần tử mới tới có thông tin chiều cao=135, độ dài mái tóc = 39 và giọng nói có giá trị là 1 thì người này là nam hay nữ?