

Trí Tuệ Nhân Tạo K58 – Thực hành 4

Bài 1: Đoán loại phòng học với cây quyết định

Cho dataset về loại phòng học trong trường ĐH Thủy Lợi như mô tả dưới đây:

STT	Có bảng viết	Có máy chiếu	Có điều hòa	Có máy tính	Loại phòng
1	Có	Có	Không	Có	Phòng học
2	Có	Có	Có	Có	Phòng máy
3	Không	Không	Có	Có	Phòng máy
4	Có	Không	Không	Không	Phòng học
5	Không	Có	Có	Có	Phòng máy
6	Có	Không	Có	Không	Phòng học
7	Có	Không	Có	Có	Phòng máy

Yêu cầu:

1. Tính lượng thông tin thu được nếu chọn thuộc tính “Có bảng viết” làm gốc phân hoạch
2. Xây dựng cây quyết định bằng thuật toán Quinlan (thuật toán ID3)
3. Tìm tập luật ứng với cây quyết định sinh từ câu trên

Bài 2: Đặc điểm nhân khẩu học

Cho dataset gồm 8 mẫu đặc điểm của người châu Âu và người châu Á như dưới đây:

STT	1	2	3	4	5	6	7	8
Dáng	To	Nhỏ	Nhỏ	To	Nhỏ	To	Nhỏ	To
Cao	TB	Cao	TB	Cao	TB	TB	Cao	Cao
Giới	Nam	Nam	Nam	Nam	Nữ	Nữ	Nữ	Nữ
Châu	Á	Á	Âu	Âu	Á	Á	Âu	Âu

Yêu cầu:

1. Xây dựng cây quyết định bằng ID3
2. Tìm tập luật xác định một người là châu Âu hay châu Á

Bài 3: Cắt tỉa tập luật sinh từ cây quyết định

Cho dataset gồm 8 mẫu đặc điểm khi khảo sát người dùng kem chống nắng như sau:

STT	Tên	Màu Tóc	Chiều Cao	Cân Nặng	Dùng Kem	Kết Quả
1	Sarah	Vàng	T.Bình	Nhẹ	Không	Cháy
2	Dana	Vàng	Cao	T.Bình	Có	Không

3	Alex	Nâu	Thấp	T.Bình	Có	Không
4	Annie	Vàng	Thấp	T.Bình	Không	Cháy
5	Emilie	Đỏ	T.Bình	Nặng	Không	Cháy
6	Peter	Nâu	Cao	Nặng	Không	Không
7	John	Nâu	T.Bình	Nặng	Không	Không
8	Kartie	Vàng	Thấp	Nhẹ	Có	Không

Yêu cầu:

1. Tính lượng thông tin thu được nếu chọn thuộc tính “Cân nặng” làm gốc phân hoạch
2. Xây dựng cây quyết định tối ưu cho tập ví dụ trên theo thuật toán Quinlan (thuật toán ID3)
3. Phát sinh tập luật ứng với cây quyết định sinh từ câu trên
4. Tập luật trên có thể được tối ưu (bằng cách bỏ điều kiện hoặc kết hợp các luật với nhau) hay không? Hãy chỉ ra nếu có
5. Có ý tưởng (phương pháp) nào giải quyết vấn đề tập luật của cây quyết định không tối ưu hay không?

Bài 4: Làm quen với Weka

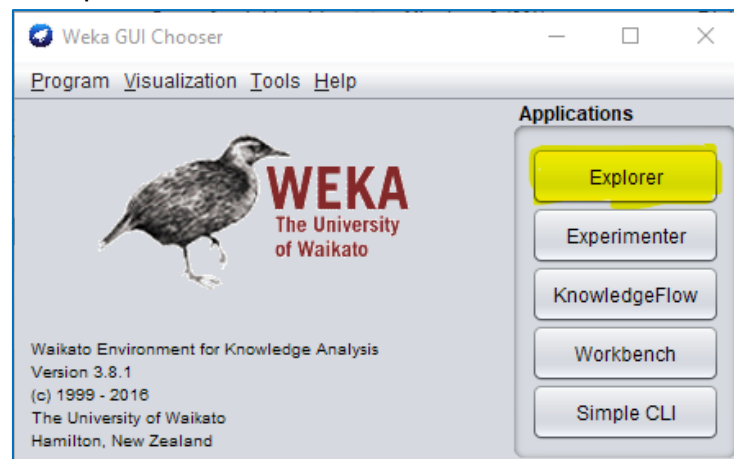
Weka là phần mềm viết bằng java cho phép thực hiện nhiều thuật toán học máy cơ bản, link tải: <https://www.cs.waikato.ac.nz/ml/weka/downloading.html>

Yêu cầu:

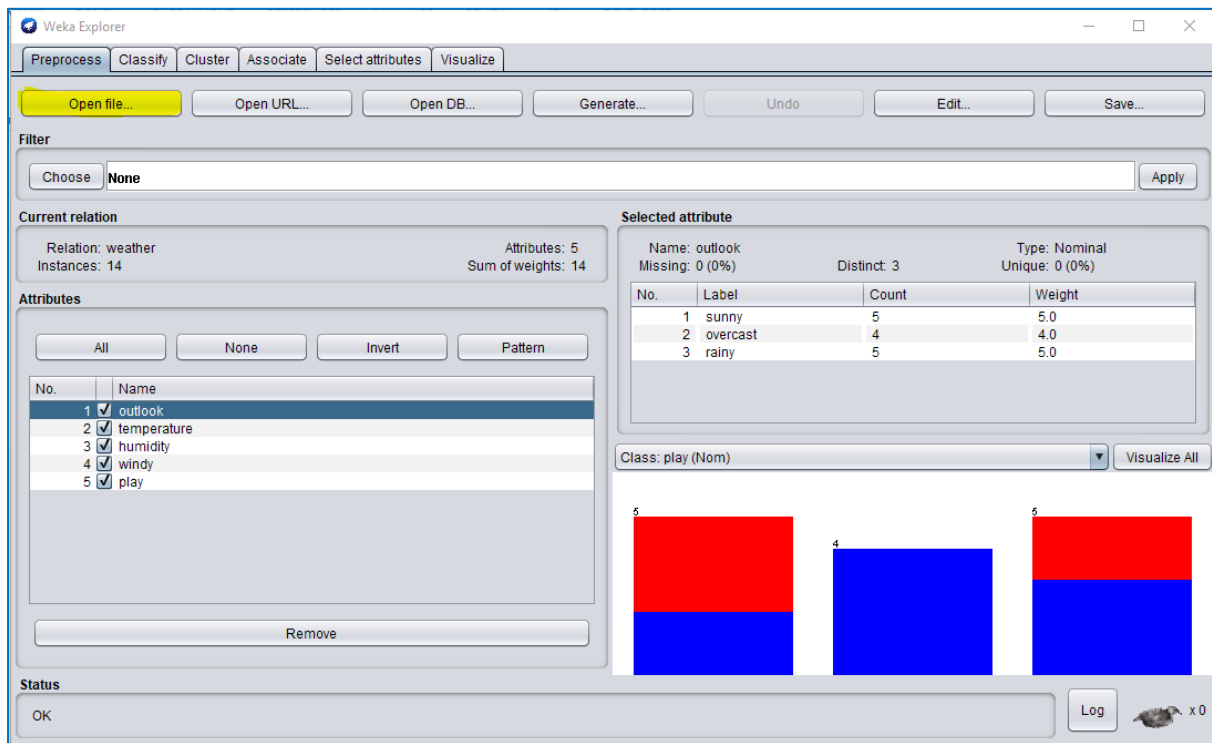
- Tải và cài đặt Weka lên máy tính của bạn
- Tải dataset minh họa: dataset của Weka có thể được lưu trên file (dạng arff), hãy tải về dataset minh họa bài toán quản lý Golf (phần cuối slide lý thuyết 12). Liên kết để tải file này có trong phần bài tập (file weather.arff)

Chạy thử nghiệm Weka với tập dữ liệu weather.arff:

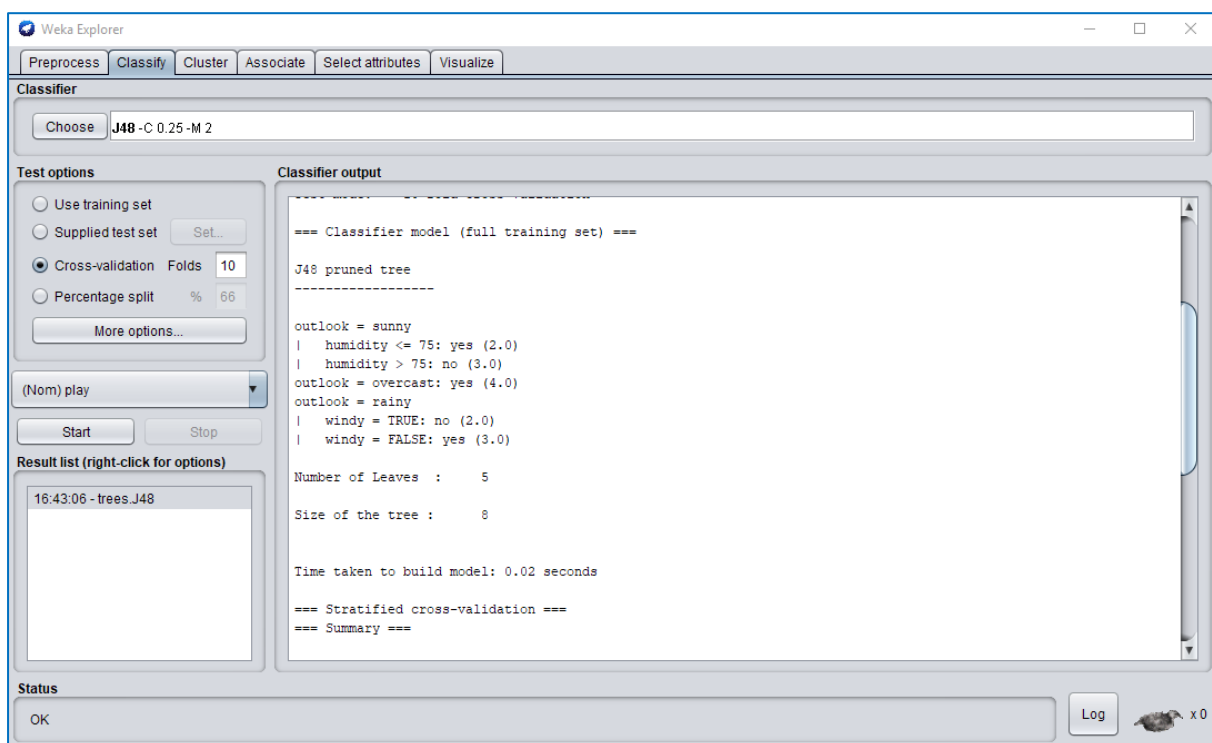
1. Chạy Weka, chọn Explorer



- Chọn dataset: chọn “Open file...” và mở file weather.arff đã được tải về. Hệ thống sẽ đọc và hiển thị một số thông tin về dataset này, đánh dấu tất cả các thuộc tính.



- Huấn luyện: chuyển sang tab Classify và bấm vào Choose, chọn thuật toán J48 (J48 là thuật toán mở rộng từ ID3, trên Weka không cài sẵn ID3, có thể tự thêm bằng cách cài gói mở rộng). Bấm Start để thực hiện thuật toán.



4. Hiển thị kết quả dạng cây (sinh viên tự tìm hiểu)

