

Bài thực hành số 1

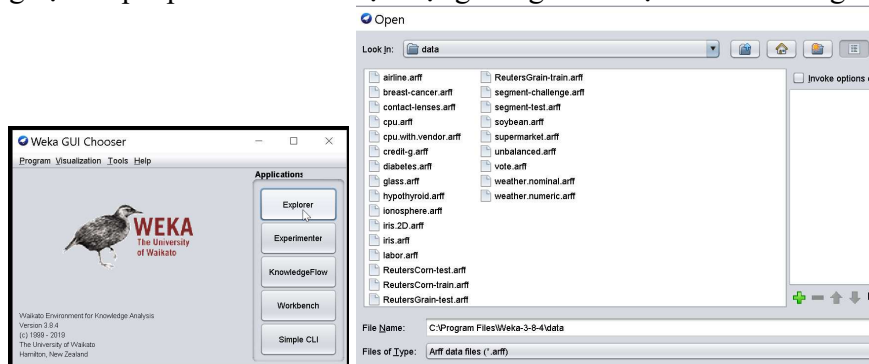
Thành lập nhóm, tối đa 3 sinh viên/nhóm. Nhóm tự chọn 1 tên ngắn gọn sau đó vào diễn đàn, tạo 1 chủ đề với tựa <Tên Nhóm đã đặt>. Vào trang chủ đề ghi ra danh sách (gồm mã số và họ tên) các thành viên trong nhóm và ghi chú thành viên làm nhiệm vụ Trưởng Nhóm.

Lưu ý: Chủ đề trong diễn đàn nhóm tạo ra dùng để nộp các bài tập, các bài báo cáo, các kết quả thảo luận,... trong suốt học kỳ. Các nhóm có thể tham khảo nội dung của các nhóm khác nhưng tuyệt đối không sao chép cho các bài kết quả của nhóm mình (đề cao tính trung thực).

Trong các buổi làm việc: yêu cầu làm việc theo nhóm, các nhóm có thể trao đổi, hỗ trợ nhau khi cần thiết. Nhóm có thể mang máy tính cá nhân cùng làm việc.

Câu 1:

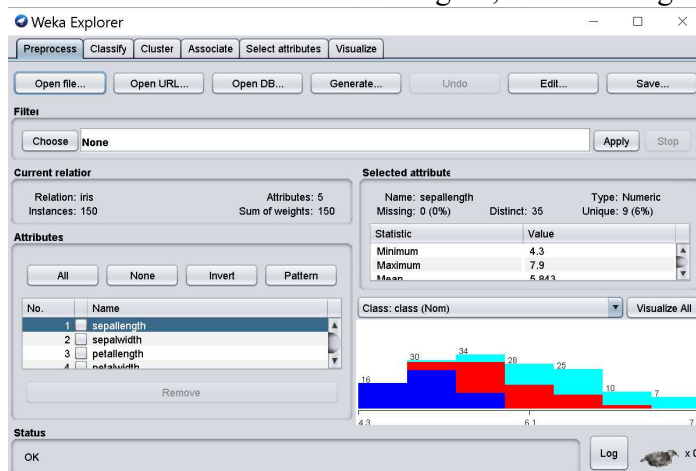
Sử dụng mạng Internet để tìm kiếm, tải về, cài đặt phần mềm WEKA của đại học Waikato New Zealand. Phần mềm này hoạt động dựa trên ngôn ngữ Java. Sau khi cài đặt xong, thử thực hiện Weka và dùng Explorer để thử nạp vào vài tập tin dữ liệu (do Weka cung cấp sẵn) để tìm hiểu các công cụ cho phép tìm hiểu các đại lượng trong các thuộc tính của bảng dữ liệu.



Lưu ý: Tập tin dữ liệu của Weka có phần mở rộng là arff. Có thể xem trực tiếp bằng các công cụ soạn thảo văn bản như NotePad, Note++, ...

Ngoài ra Weka cũng cho phép người dùng nhập (import) dữ liệu từ các nguồn khác nhau (bảng tính, tập tin CSDL,...).

Nhóm cần tìm hiểu tất cả các thông tin, các tính năng trong thẻ tab Preprocess



Câu 2:

Thử tìm xem có những phần mềm khác tương tự như Weka. Sau đó chọn 1 để thử nghiệm (tải về, cài đặt, chạy thử,...)

Câu 3: Làm quen với việc tổng hợp dữ liệu và làm sạch dữ liệu.

Có 4 tập tin dạng bảng tính (Excel) lưu thông tin sinh viên gồm mã số, họ lót, tên, ngày sinh (ngày tháng và năm). Nhóm có nhiệm vụ tổng hợp 4 danh sách này thành 1 danh sách duy nhất (tổng hợp) sau đó làm sạch thông tin (phát hiện các cụm từ sai, các dữ liệu không hợp lệ,...) Kết quả làm việc sẽ có 2 tập tin: tập tin đầu tiên chứa danh sách có thông tin đầy đủ và chính xác, tập tin thứ 2: chứa các sinh viên có thông tin chưa đầy đủ hoặc không chính xác.

Nhóm có quyền sử dụng bất kỳ công cụ gì để thực hiện nhiệm vụ và bất kỳ ứng dụng gì để lưu lại kết quả làm việc (lưu ở dạng text, bảng tính, bảng dữ liệu Access,...). Thực hiện xng câu 3: Nhóm vào diễn đàn và gửi (attach) 2 tập tin này trong phần chủ đề m2 nhóm tạo ra.

Câu 4: Phân thu hoạch

Phân công thực hiện, thực hiện ở nhà, thời hạn 1 tuần và gửi vào chủ đề trong diễn đàn.

- Hướng dẫn cài đặt và làm quen WEKA.
- Hướng dẫn cài đặt và làm quen ngôn ngữ lập trình Python. Minh họa việc dùng Python để viết các chương trình đơn giản sau:
 1. Nhập 4 số nguyên, in ra số lớn nhất và nhỏ nhất trong 4 số.
 2. Nhập (hoặc gán trước) 1 mảng các số thực. Đếm xem có bao nhiêu số dương.
 3. Nhập (hoặc gán trước) các số nguyên vào 1 mảng 2 chiều có số hàng số cột tùy ý chọn trước. Sau đó in ra các phần tử nằm trên hàng thứ k (với k do người sử dụng nhập vào).

Lưu ý: Nhóm có thể dùng bất kỳ ứng dụng để thực hiện báo cáo (Power Point, Word, Clip/Video,...). Phân thu hoạch này cần nộp lên diễn đàn trước buổi học lý thuyết kế tiếp: GV có thể mời 1 nhóm lên trình bày.