

## BÀI THỰC HÀNH 04 (Tính điểm quá trình)

### Yêu cầu:

Từ slide BUỔI 06 và 02 tập dữ liệu Chess, Mushroom (classroom), sinh viên thực hiện các thực nghiệm:

Sử dụng thuật toán khai thác tập phổ biến từ các thư viện được giới thiệu trên lớp:

1. So sánh thời gian thực hiện, bộ nhớ sử dụng và số lượng itemset phổ biến của thuật toán từ các thư viện (2 thư viện tự chọn) – thực nghiệm trên 5 giá trị ngưỡng *minsup* khác nhau. Theo dữ liệu đầu vào:
  - Dùng dữ liệu gốc;
  - Dữ liệu được sắp xếp các item thứ tự tăng dần theo độ phổ biến;
  - Lọc DataFrame chỉ chứa các item thỏa ngưỡng *minsup* trước khi thực hiện khai thác tập phổ biến;
2. Phân tích ưu và nhược điểm của từng hàm khai thác tập phổ biến trên giữa các thư viện: dữ liệu đầu vào, thời gian thực hiện, tính mở (các đối số) và kết quả trả về có dễ dàng truy vấn báo cáo.

### Quy cách nộp bài:

- Tập tin chương trình .py;
- Tập tin .docx:
  - Mô tả tập dữ liệu;
  - Mô tả phân tích ưu và nhược điểm;
  - Có hình ảnh biểu đồ trực quan hóa so sánh thời gian thực hiện và bộ nhớ sử dụng;
- Tạo Folder là <Họ tên SV> chứa các tập tin – nén và nộp trên classroom.