

A. Nội dung tiểu luận:

Gồm 3 phần:

1. Thu thập dữ liệu:

- SV phải TỰ thu thập dữ liệu (tự crawl).
- Số lượng: >1000 mẫu, dẫn nguồn dữ liệu và mô tả cách thức thu thập.
- Xuất ra các thống kê mô tả trực quan về dữ liệu.

2. Trích xuất đặc trưng:

- Trình bày việc lựa chọn đặc trưng, làm sạch và chuẩn hóa dữ liệu, giảm chiều, ...
- Trực quan hoá kết quả của các quá trình trên.

3. Mô hình hóa dữ liệu:

- Chọn các mô hình phù hợp (ít nhất 2 mô hình/thuật toán).
- Chia dữ liệu thành Train/Validation/Test set theo tỉ lệ phù hợp.
- Trình bày các đồ thị thể hiện hiệu quả của các mô hình trong quá trình huấn luyện/Hiệu chỉnh/Kiểm thử.
- So sánh hiệu quả của các mô hình bằng bảng dữ liệu hoặc đồ thị.

B. Phân nhóm và chấm thi:

- Các SV đăng ký nhóm và đề tài theo lớp học phần: 3 SV/nhóm (trường hợp đặc biệt thì 2 SV/nhóm).
- Mỗi SV trình bày RIÊNG phần việc mình làm thông qua nội dung viết trong tiểu luận và phần trình bày trên slide chung của cả nhóm để GV chấm điểm. Mỗi SV ghi rõ (những) phần việc mình làm trong bảng phân công công việc ở đầu tiểu luận và trình bày slide/demo code phần mình làm để GV vấn đáp và chấm điểm.
- Thời gian trình bày slide và demo code cho mỗi nhóm: 15 phút.