

# Danh sách đề tài môn học máy thống kê

## 1) Xây dựng dữ liệu và mô hình phân loại chủ đề văn bản báo chí

- Đối với đề này, SV có thể đăng ký nhiều nhóm. Mỗi nhóm là một cách chia về các chủ đề sao cho hợp lý.

## 2) Xây dựng dữ liệu và mô hình phân loại bình luận theo cảm xúc

- Đối với đề này, SV có thể đăng ký nhiều nhóm. Mỗi nhóm là một miền dữ liệu/một loại sản phẩm nào đó.

## 3) Xây dựng dữ liệu và mô hình phân loại bình luận theo cung bậc cảm xúc

- Đối với đề này, SV có thể đăng ký nhiều nhóm. Mỗi nhóm đăng ký xây dựng dữ liệu trên một mạng xã hội nào đó. Ví dụ: Twitter, Youtube, Tiktok, Facebook, các trang báo, ...

## 4) Xây dựng dữ liệu và mô hình phân loại những bài hát theo chủ đề

- Đối với đề này, SV có thể đăng ký nhiều nhóm. Mỗi nhóm là một cách chia về các chủ đề bài hát sao cho hợp lý.

Những đề tài có thể phát triển thành khóa luận tốt nghiệp

## 5) Xây dựng dữ liệu và mô hình phân loại câu hỏi tương đồng trong hệ thống hỏi đáp

## 6) Xây dựng dữ liệu và mô hình tạo sinh văn bản từ dữ liệu

## 7) Xây dựng dữ liệu và mô hình phân loại bình luận xúc phạm thân thể (body shaming)

## 8) Xây dựng dữ liệu và mô hình suy luận ngôn ngữ

## 9) Xây dựng dữ liệu và mô hình phân loại bệnh dựa trên hình ảnh (ảnh Xquang, CT, MRI,...)

## 10) ....

Yêu cầu:

- Bộ dữ liệu ít nhất khoảng 2000 điểm trở lên. Qui trình xây dựng dữ liệu rõ ràng và có đánh giá chỉ số độ đồng thuận (**inter-annotator agreement**). Có thể tham khảo từ những công trình nghiên cứu đi trước và **thiết kế các qui trình sao cho hợp lý**.
- Các thông số đánh giá rõ ràng (Ví dụ: Accuracy, Precision, Recall, F1-macro).
- Hiểu, cài đặt, và thử nghiệm ít nhất khoảng 03 thuật toán trở lên.

Và các đề tài khác do các SV đề xuất!