

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH

## **BÀI TẬP 1**

Môn học: THU THẬP VÀ TIỀN DỮ LÝ DỮ LIỆU.

Học kỳ II (2021-2022)

Sinh viên: Nguyễn Quốc Nam

MSSV: 20520644

Lớp: DS103.M21

Trường: Đại học Công Nghệ Thông Tin

Giảng viên: TS NGUYỄN GIA TUẤN ANH,  
CN LƯU THANH SƠN.

## SỰ KHÁC NHAU GIỮA DỮ LIỆU TRONG LĨNH VỰC Y KHOA (MEDICAL) VÀ LĨNH VỰC MẠNG XÃ HỘI (SOCIAL NETWORK).

*Kiểu dữ liệu:*

- Dữ liệu trong lĩnh vực y khoa bao gồm cả 2 loại là có cấu trúc và phi cấu trúc, định tính.
- Dữ liệu trong lĩnh vực mạng xã hội đa phần là dữ liệu phi cấu trúc và định lượng.

*Công dụng:*

- Dữ liệu trong lĩnh vực y khoa đa phần rất đặc thù. Mỗi bộ dữ liệu thường chỉ được dùng để giải quyết một vấn đề (một loại bệnh, một nhóm bệnh) khác nhau. Dữ liệu này đa số được các công ty, nhà tiêu dùng sử dụng cho mục đích thương mại.
- Dữ liệu trong lĩnh vực mạng xã hội có sự liên quan nhất định. Các bộ dữ liệu có thể kết hợp với nhau để cùng giải quyết một số bài toán. Dữ liệu này được sử dụng trong mục đích nghiên cứu là chủ yếu.

### BÀI TOÁN ỨNG VỚI DỮ LIỆU ĐÃ THU THẬP.

*Dữ liệu:* các bài đăng về sức khỏe trên mạng xã hội của những chuyên gia và bệnh nhân.

- Bài toán 1: Phân loại bài đăng của bệnh nhân và chuyên gia trong các bài viết.
- Bài toán 2: Nhận diện vấn đề sức khỏe trong bài đăng.