

BÁO CÁO KẾT QUẢ THỰC HIỆN BÀI TẬP CUỐI KỲ - XỬ LÝ NGÔN NGỮ TỰ NHIÊN

Tên Đề tài: PHÂN TÍCH CẢM XÚC CỦA KHÁCH HÀNG VỀ CƠ SỞ Y TẾ

Họ và tên thành viên: Lê Hồng Anh

Lớp: 21AD

Chữ ký:

I. Thuật toán, mô hình, thư viện, mã nguồn sử dụng trong đề tài:

- numpy: Thư viện tính toán số học hiệu năng cao, hỗ trợ mảng nhiều chiều và thao tác ma trận.
- pandas: Cung cấp cấu trúc dữ liệu bảng (DataFrame) để xử lý và phân tích dữ liệu hiệu quả.
- torch (PyTorch): Thư viện học sâu mã nguồn mở cho xây dựng và huấn luyện mô hình neural network.
- transformers: Thư viện của Hugging Face cho mô hình NLP hiện đại như BERT, GPT, RoBERTa,...
- datasets: Hỗ trợ tải, xử lý và chia nhỏ tập dữ liệu NLP nhanh chóng (cũng của Hugging Face).
- sklearn (Scikit-learn): Cung cấp các công cụ huấn luyện, đánh giá và chia tập dữ liệu cho học máy truyền thống.
- bs4: Thư viện Python dùng để phân tích và trích xuất dữ liệu từ HTML hoặc XML theo cấu trúc cây DOM một cách đơn giản và nhanh chóng.
- selenium: Thư viện tự động hóa trình duyệt, cho phép điều khiển các thao tác như click, nhập liệu và tương tác với JavaScript, thường dùng để thu thập dữ liệu từ các trang web động.
- accelerate: Tự động hóa việc tối ưu hóa và phân phối mô hình trên nhiều phần cứng (CPU, GPU,...). Ví dụ: sử dụng mã nguồn mở nào không? Sử dụng thư viện gì không?

II. Mô tả Dữ liệu (ngắn gọn):

- Nguồn dữ liệu: Dữ liệu gồm các bình luận người dùng về một số bệnh viện ở Việt Nam được thu thập từ Google Maps.
- Số mẫu (dòng): 3063
- Số cột: 4
- Các cột gồm:
 - o comment: Bình luận gốc từ người dùng trên Google Maps
 - o hospital: Tên bệnh viện được đánh giá
 - o star: Số sao đánh giá (chuỗi như "5 sao")
 - o rating: Điểm trung bình (VD: 3,6)

III. Liệt kê các chức năng đã hoàn thành

1. Huấn luyện thành công một mô hình phân tích và dự đoán cảm xúc dựa trên bình luận của khách hàng
2. Có giao diện người dùng để nhận bình luận và đưa ra kết quả phân tích và dự đoán cảm xúc của bình luận (Tệ, Trung tính, Tốt)