



## Sentiment Analysis for E-Commer

### Ngữ cảnh đề án

Trong thương mại điện tử, việc liên tục nâng cao sản phẩm và dịch vụ để đáp ứng nhu cầu khách hàng nhằm nâng cao uy tín của doanh nghiệp là công việc hàng đầu của các doanh nghiệp tham gia sàn thương mại điện tử.

Hệ thống hỗ trợ doanh nghiệp phân loại các phản hồi của khách hàng thành 3 nhóm: tích cực, tiêu cực, trung tính dựa trên dữ liệu dạng văn bản.

Xây dựng hệ thống dựa trên lịch sử những đánh giá của các khách hàng đã có trước đó, dữ liệu được thu thập từ phần bình luận và đánh giá của khách hàng ở trang web thương mại điện tử (VD: nhánh ngành thời trang)...

Hệ thống giúp doanh nghiệp có thể biết được những phản hồi nhanh chóng của khách hàng về sản phẩm hay dịch vụ của họ, điều này giúp cho doanh nghiệp hiểu được tình hình kinh doanh, hiểu được ý kiến của khách hàng từ đó giúp doanh nghiệp cải thiện hơn trong dịch vụ, sản phẩm.

### Yêu cầu chung

Thu thập dữ liệu lịch sử đánh giá của khách hàng từ các trang thương mại điện tử (tiki, shopee, lazada) nhóm ngành thời trang và áp dụng thuật toán phù hợp để xây dựng models phân loại đánh giá của khách hàng theo 3 mức độ cảm xúc “Tích cực, tiêu cực, trung tính”

**Số HV tham gia: 01**

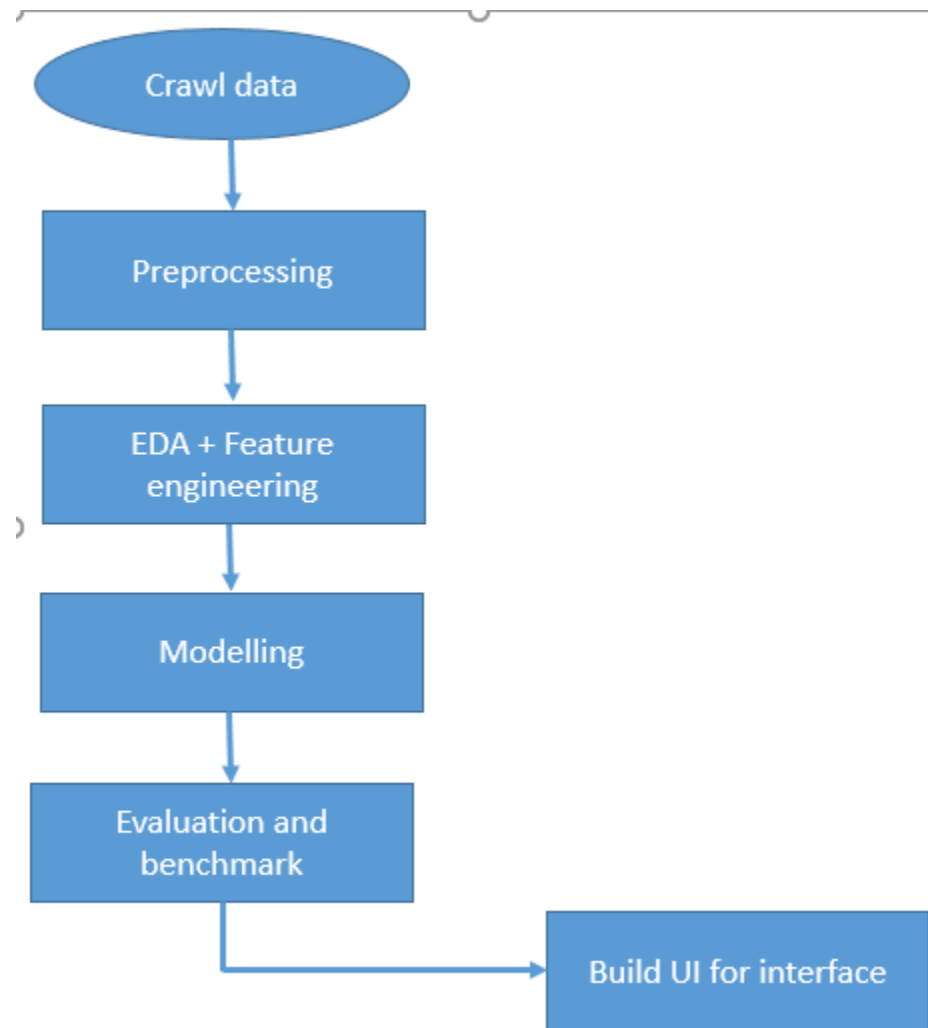
### Yêu cầu chi tiết của đề án

- Yêu cầu 1: Thực hiện crawl dữ liệu từ phần bình luận và đánh giá của các website bán hàng hay cung cấp dịch vụ
- Yêu cầu 2: Tiền xử lý dữ liệu
- Yêu cầu 3: Trích xuất đặc trưng
- Yêu cầu 4: Xây dựng mô hình. Thực hiện nhiều phương pháp rồi thực hiện so sánh độ chính xác
- Yêu cầu 5: Xây dựng UI bằng streamlit để cho người dùng được trực tiếp nhập đánh giá và trả ra kết quả phân loại.



### **Yêu cầu chi tiết về kiến trúc của đồ án**

- Đồ án được thực hiện dành cho đối tượng nào sử dụng?  
Các doanh nghiệp trong lĩnh vực cung cấp sản phẩm, dịch vụ trên sàn thương mại điện tử
- Mô hình kiến trúc



### **Ghi chú:**

- Học viên sử dụng ngôn ngữ lập trình Python, Editor: Jupyter Notebook/ Jupyter Lab/ Google Colab/ Markdown Text và các package cần thiết
- Đề có thể hoạt động được: 5 điểm
- Crawl data: +1.5 điểm
- Có sự nghiên cứu và áp dụng những phương pháp mới, thuật toán mới: +1.5 điểm
- Có demo: +1 điểm
- Có sự so sánh giữa các phương pháp với nhau: +1 điểm

