

BẢN MÔ TẢ DỰ ÁN

Đội thi: Thiên Khôi

Tên dự án: Sử dụng công cụ AI để phân loại cảm giác đánh giá của khách hàng

Lĩnh vực: AI

Keyword: phân loại cảm giác, AI, Roberta Transformers

1. Vấn đề đặt ra:

Trong thời đại 4.0, các doanh nghiệp đang phải đối mặt với một lượng lớn đánh giá và phản hồi từ khách hàng trên các nền tảng kỹ thuật số. Cảm giác, ý kiến của khách hàng đôi khi sẽ khác so với số sao được đánh giá. Để hiểu được cảm nhận thật của khách hàng (không dựa trên số sao từ 1-5) và phản ứng của họ đối với sản phẩm hoặc dịch vụ, các doanh nghiệp có thể tích hợp công cụ AI vào hệ thống của mình để phân loại các cảm giác khác nhau trong đánh giá của tất cả khách hàng.

Dự án này đặt ra vấn đề làm thế nào để tự động phân loại cảm giác từ đánh giá của khách hàng và áp dụng nó vào thực tế để cải thiện chất lượng sản phẩm hoặc dịch vụ.

2. Giải pháp đề xuất

2.1. Tổng quan về giải pháp

Sử dụng công cụ Python xử lý ngôn ngữ tự nhiên để tạo bộ phân loại Phân tích tình cảm với Huggingface Roberta Transformers của NLTK. Từ đó, phân loại cảm giác của khách hàng trên sàn thương mại điện tử dựa trên những dòng chữ được viết bởi cảm nhận của khách hàng khi trải nghiệm sản phẩm, dịch vụ. Có kết quả phân loại cảm giác, doanh nghiệp sẽ có những thông tin hữu ích để cải thiện trải nghiệm của khách hàng.

2.2. Mô tả cách thức thực hiện

Data được lấy từ cuộc thi “Hackathon.QuyNhon.AI 2022” nói về đánh giá (review) của du khách về du lịch Bình Định trên mạng xã hội ở 5 phương diện: Dịch vụ vui chơi giải trí, Dịch vụ lưu trú (Khách sạn, nhà nghỉ), Hệ thống nhà hàng phục vụ khách du lịch, Dịch vụ ăn uống, Dịch vụ vận chuyển, Dịch vụ mua sắm. Sau đó, chúng ta sẽ thực hiện một số phân tích tình cảm bằng Huggingface Roberta Transformers.

Trước khi đi vào các bước cụ thể ta cần làm rõ một số thuật toán được sử dụng trong bài:

- NLTK (Natural Language Toolkit) là một thư viện phân mềm mã nguồn mở cho xử lý ngôn ngữ tự nhiên bằng Python. Nó cung cấp các công cụ và tài nguyên để xử lý ngôn ngữ tự nhiên như:
 - Tokenization: chia văn bản thành các token như từ, cụm từ, ký tự.
 - POS tagging: gán nhãn từ loại cho mỗi token. Ví dụ: danh từ, động từ, tính từ...
 - Classification: phân loại văn bản.
 - Clustering: nhóm cụm văn bản.

- Topic modeling: tìm chủ đề chính trong văn bản.
- Huggingface Roberta Transformers là một thư viện trong hệ sinh thái Huggingface Transformers cung cấp các mô hình transformer đã được huấn luyện sẵn cho nhiệm vụ xử lý ngôn ngữ tự nhiên (NLP).
- Một số đặc điểm chính:
 - + Là mô hình dựa trên kiến trúc Transformer sử dụng cơ chế self-attention để xử lý dữ liệu văn bản.
 - + Roberta là viết tắt của Robustly optimized BERT pretraining approach, là phiên bản cải tiến của mô hình BERT.
 - + Cung cấp các mô hình đã được huấn luyện sẵn cho nhiều tác vụ NLP như phân loại văn bản, trích xuất đặc trưng, tóm tắt văn bản, đối thoại, dịch máy, ecc.
 - + Cho phép dễ dàng tích hợp vào các ứng dụng thực tế với hiệu năng cao mà không cần huấn luyện lại mô hình.

Step 0: Đọc data và sử dụng NLTK (Natural Language Toolkit)

- Kết nối với Google Drive để có thể truy cập dữ liệu
- Cài đặt thư viện NLTK trong môi trường Colab
- Nhập những thư viện cần thiết vào chương trình
- Đọc dữ liệu từ tệp CSV có đường dẫn '/content/review_data.csv' vào DataFrame
- Làm sạch data
- NLTK (Natural Language Toolkit):
 - lấy ví dụ trong cột "Review" thứ 100 xong in
 - chạy `nlk.word_tokenize` để tách câu văn thành các từ riêng lẻ và lưu chúng vào danh sách tokens (lấy 10 chữ đầu)
 - chạy `nlk.pos_tag(tokens)` để phân loại từng từ trong danh sách tokens và lưu kết quả vào danh sách tagged. Kết quả này bao gồm từ và loại từ (VD: danh từ, động từ, tính từ, ...). (lấy 15 chữ đầu)
 - chạy `nlk.chunk.ne_chunk` để phân tích cú pháp và xác định các thực thể trong danh sách

Step 1: Roberta Pretrained Model

Trong bài, thư viện Roberta Transformers được sử dụng nhằm mục đích để đánh giá sự tích cực (positive), trung lập (neutral), hoặc tiêu cực (negative) của câu trên nội dung của nó.

- Cài đặt thư viện Transformers, nó giúp tải và sử dụng các mô hình ngôn ngữ tự nhiên đã được đào tạo sẵn.
- Import các thư viện cần thiết
- Load mô hình Roberta và tokenizer
- Token hóa và đánh giá đoạn văn bản
- Xử lý kết quả và in ra các xác suất

Step 2: Đánh giá theo sao

- Nhập thư viện matplotlib để tạo biểu đồ.
- Đặt kích thước của nhãn trục x, y của biểu đồ

- Tính toán số lượng đánh giá cho từng số sao và sắp xếp chúng theo thứ tự tăng dần (sử dụng dữ liệu từ cột 'Score' trong DataFrame)
- Vẽ biểu đồ

2.3. Kết quả chạy thử

https://drive.google.com/drive/folders/1OEVkv9fUzUXcZOjXDSiXHqEwvbx7uLUp?usp=drive_link

So sánh 2 phương pháp sử dụng : Kết quả cho ra bằng 2 phương pháp dưới đây chênh lệch nhau khá nhiều nhưng ta có thể nhận thấy một số điều rằng:

- Đánh giá theo sao(1-5): kết quả cho ra khá khả quan nhưng nó không mang lại kết quả chính xác nhất. Dựa theo số sao mà nhà hàng nhận được, nhiều khách hàng đã kỳ vọng cao vào các nhà hàng ấy và tìm đến để trải nghiệm. Tuy nhiên, họ không nhận lại được kết quả như mong đợi. Với lý do là số sao mà nhà hàng nhận được không hoàn toàn đúng vì nhiều khách hàng đã “du di” cho nhà hàng đánh giá tốt.
- Roberta model đánh giá “Review” theo từng từ trong câu cho ra kết quả sát với thực tế hơn. Bởi vì đôi lúc số sao sẽ không phản ánh được đúng với cảm nhận khách hàng mà phải dựa vào ngữ cảnh câu. Từ đây nhà đầu tư có thể nhận thấy những điều chưa hoàn thiện và cải tiến phát triển sản phẩm của mình tốt hơn mang đến cho khách hàng.

2.4. Đánh giá giải pháp

- Đánh giá dựa trên những dòng chữ được viết bởi khách hàng giúp doanh nghiệp xác định những khía cạnh mà họ làm tốt và điểm yếu trong sản phẩm hoặc dịch vụ của mình. Việc này giúp cải thiện chất lượng và tối ưu hóa trải nghiệm của khách hàng.
- Việc hiểu được cảm nhận của khách hàng giúp doanh nghiệp tối ưu hóa chiến lược tiếp thị và quảng cáo để thu hút và duy trì khách hàng.
- Bằng cách lắng nghe và phản hồi vào đánh giá của khách hàng, doanh nghiệp có thể xây dựng mối quan hệ lâu dài và tạo sự trung thành từ phía khách hàng.

3. Dự kiến kết quả áp dụng vào thực tế

- Việc sử dụng trí tuệ nhân tạo để phân tích đánh giá khách hàng là một công cụ mạnh mẽ và quan trọng trong việc quản lý chất lượng sản phẩm, dịch vụ và tạo trải nghiệm tốt cho khách hàng.
- Kết quả của dự án có thể áp dụng vào thực tế để cải thiện chất lượng sản phẩm hoặc dịch vụ dựa trên cảm nhận của khách hàng.