

GUIDELINE GÁN NHÃN DỮ LIỆU

- Đồ án môn học DS108 -

I. CHUẨN BỊ DỮ LIỆU

1. Dữ liệu thô

- Yêu cầu đầu vào: File .jsonl với định dạng:

```
{"text": "Đẹp, chắc chắn, quạt chạy êm!", "label": []}
```

II. THIẾT LẬP TRÊN DOCCANO

1. Tạo project:

- Loại project:
 - + Chọn Sequence Labeling
 - + Tiến hành đặt tên project và description
 - + Tick vào ô “Allow overlapping spans”
- Import dữ liệu:
 - + Upload file .jsonl đã được tiền xử lý

2. Tạo nhãn:

2.1. Nhóm Aspect:

Gồm các nhãn sau:

Nhãn	Mô tả	Ví dụ
PRODUCT	Chất lượng, tính năng sản phẩm.	"Màn hình sắc nét", "Pin tụt nhanh"
SERVICE	Dịch vụ CSKH, đổi trả, bảo hành, vận chuyển, đóng gói.	"Nhân viên nhiệt tình", "Đổi trả chậm", "Giao hàng nhanh", "Hộp bị méo"
PRICE	Giá cả, khuyến mãi, so sánh giá.	"Giá rẻ", "Đắt hơn Shopee"

Lưu ý: Đối với các cụm không thuộc phạm trù trên hoặc mang nghĩa chung chung so với các nhãn trên thì BỎ QUA, không cần label nhãn đó.

2.2. Nhóm Sentiment:

Gồm các nhãn:

Nhãn	Mô tả	Ví dụ
POSITIVE	Khen ngợi, hài lòng.	“Chất lượng tuyệt vời, đóng gói cẩn thận”
NEUTRAL	Trung tính, không rõ cảm xúc	“Chưa dùng nên chưa biết chất lượng như thế nào”
NEGATIVE	Phàn nàn, không hài lòng.	“Với giá như này thì chất lượng quá tệ”

III. QUY TẮC GÁN NHÃN

1. Nguyên tắc chung:

Nguyên tắc	Mô tả
Gán nhãn chồng (multi-label)	Một câu có thể chứa nhiều cặp [ASPECT, SENTIMENT]
Gán theo cụm từ	Ưu tiên gán từng cụm từ nhỏ mang cảm xúc, không gán toàn câu nếu chứa nhiều ý
Tách cảm xúc rõ ràng	Phân biệt rõ các khía cạnh như: PRODUCT, PRICE, SERVICE, ...
Gán cảm xúc theo ngữ nghĩa thực (xử lý theo ngữ cảnh)	Không chỉ dựa vào từ khóa (ví dụ: "Không thể chê vào đâu được" là POSITIVE)

- Sau khi gán nhãn xong đảm bảo tick vào dấu “X” ở hàng trên để chuyển sang dấu “✓” tức là đã gán xong một sample đó.

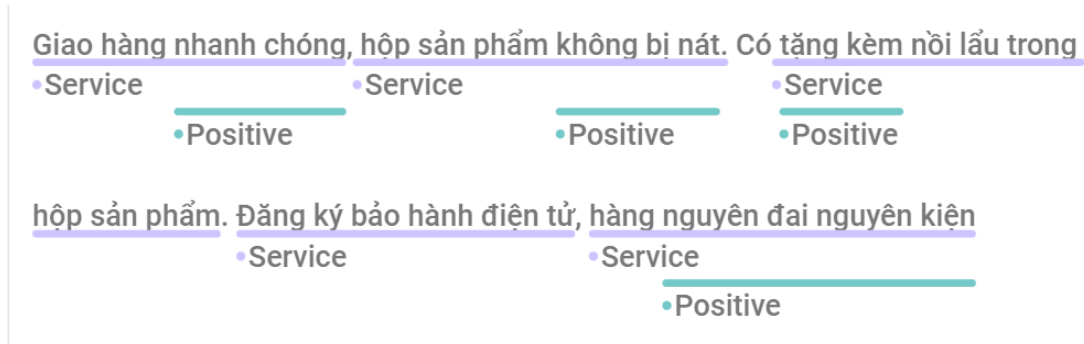


- Ví dụ: "Máy ảnh đẹp nhưng giá cao"
→ "Máy ảnh đẹp" = [PRODUCT, POSITIVE]
→ "giá cao" = [PRICE, NEGATIVE]

2. Ví dụ minh họa:

VD1: “Giao hàng nhanh chóng, hộp sản phẩm không bị nát. Có tặng kèm nồi lẩu trong hộp sản phẩm. Đăng ký bảo hành điện tử, hàng nguyên đai nguyên kiện”

- “Giao hàng nhanh chóng” = [SERVICE, POSITIVE]
- “Hộp sản phẩm không bị nát” = [SERVICE, POSITIVE]
- “Có tặng kèm nồi lẩu trong hộp sản phẩm” = [SERVICE, POSITIVE]
- “Đăng ký bảo hành điện tử, hàng nguyên đai nguyên kiện” = [SERVICE, POSITIVE]



VD2: “Sản phẩm được đóng gói cẩn thận. Nồi chiên dung tích 1,5l, dùng rất tiện. Thức ăn chín đều, vàng giòn, có lưới lọc chỗ thoát hơi nên khi chiên không bị bay mùi. Dùng hơi hao dầu và vệ sinh phải cẩn thận. Mua lúc giá tốt, dùng rất ổn”

- “Sản phẩm được đóng gói cẩn thận” = [SERVICE, POSITIVE]
- “Nồi chiên dung tích 1,5l, dùng rất tiện” = [PRODUCT, POSITIVE]
- “Thức ăn chín đều, vàng giòn, có lưới lọc chỗ thoát hơi nên khi chiên không bị bay mùi.” = [PRODUCT, POSITIVE]
- “Dùng hơi hao dầu và vệ sinh phải cẩn thận.” = [PRODUCT, NEGATIVE]
- “Mua lúc giá tốt, dùng rất ổn” = [PRICE, POSITIVE, NEUTRAL]

Sản phẩm được đóng gói cẩn thận. Nồi chiên dung tích 1,5l, dùng rất tiện. Thức

•Service

•Positive

•Positive

•Product

•Product

ăn chín đều, vàng giòn, có lưới lọc mỡ thoát hơi nên khi chiên không bị bay mùi.

•Positive

•Positive

Dùng hơi hao dầu và vệ sinh phải cẩn thận. Mua lúc giá tốt, dùng rất ổn.

•Product

•Price

•Negative

•Positive

•Neutral

3. Các trường hợp đặc biệt:

- Câu phủ định:

+ "Không thể chê vào đâu được" → POSITIVE (dù có từ "chê").

- Câu so sánh:

+ "Đắt hơn Lazada nhưng chất lượng tốt hơn" → "Đắt hơn Lazada" = [PRICE, NEGATIVE], "chất lượng tốt hơn" = [PRODUCT, POSITIVE].

- Câu nghi vấn:

+ "Mua ở đây có an toàn không?" → [SERVICE, NEUTRAL]