**GUIDELINE HƯỚNG DẪN GÁN NHÃN DỮ LIỆU**

**I. Gán nhãn dữ liệu là gì.**

Nói một cách dễ hiểu, gán nhãn dữ liệu chính là thu thập dữ liệu lại, đưa chúng vào một tổ hợp được lặp sẵn giúp việc nhận biết và sử dụng những dữ liệu đó trở nên dễ dàng hơn. ần 95% dữ liệu được tạo ra là không có cấu trúc. Nói một cách dễ hiểu, dữ liệu phi cấu trúc có thể ở khắp nơi và không được xác định đúng cách. Nếu bạn đang xây dựng một mô hình AI, bạn cần cung cấp thông tin cho một thuật toán để nó xử lý và cung cấp kết quả đầu ra và suy luận.

**Tóm lại, ghi nhãn dữ liệu và chú thích dữ liệu là tất cả về việc gắn nhãn hoặc gắn thẻ thông tin / siêu dữ liệu có liên quan trong tập dữ liệu để cho phép máy móc hiểu chúng là gì. Tập dữ liệu có thể ở bất kỳ dạng nào, ví dụ như hình ảnh, tệp âm thanh, cảnh quay video hoặc thậm chí là văn bản.** Khi chúng tôi gắn nhãn các phần tử trong dữ liệu, các mô hình ML hiểu chính xác những gì chúng sẽ xử lý và lưu giữ thông tin đó để tự động xử lý thông tin mới hơn được xây dựng dựa trên kiến ​​thức hiện có để đưa ra quyết định kịp thời.

Có rất nhiều loại dữ liệu có thể gán nhãn được, ví dụ như hình ảnh, âm thanh, văn bản, đồ họa hay hỗn hợp các định dạng. Tùy thuộc các chức năng và tham số được chỉ định, mô hình phân loại sẽ tiến hành gán nhãn.

Trong Guideline này, chúng tôi sẽ chỉ hướng dẫn các bạn gán nhãn một dữ liệu thuộc định dạng văn bản, cụ thể là comment (bình luận) trên mạng xã hội mà chúng tôi đã thu thập được để phục vụ cho **bài toán nhận diện cảm xúc qua comment** trên tiki của chúng tôi.

**II. Thông tin chi tiết các nhãn.**

Loại nhãn cơ bản dùng nhiều trong tiki mà chúng ta cần biết: **Sentiments ( 4 nhãn positive, negative, neutral và undefined)**.

Nhằm phục vụ cho mục đích tăng độ đồng thuận cho các a**nnonators**, chúng tôi yêu cầu các bạn phải hiểu định nghĩa và công dụng của các nhãn, các nhãn chúng tôi định nghĩa như sau:

**2.1. Nhãn Sentiments**

**2.1.1. Positive:**

Thường thấy ở những comment mang ý nghĩa tích cực, yêu đời, vui mừng phấn khích như:

* Vui mừng khi nhận được sản phẩm theo ý mình
  + VD: Hàng đẹp đóng gói kỹ càng, chất lượng sản phẩm quá tốt so với số tiền.
* Khen sản phẩm lẫn người bán
  + VD: Shop gửi hàng đúng như lúc đặt, rất ok cho 5 sao.
* Giới thiệu tích cực tới những khách hàng khác.
  + VD: Qúa xịn xò, pin to bằng iphone 6 tiện lợi.
* Khen trang giao dịch trung gian (tiki)
  + VD: Tuiii uiiii giao cực kì nhanh đặt hôm qua hôm nay đã giao, hàng rất là ổn áp nha mình xài hãng này roiiiii. Opng shipper dễ thương xĩu, nói chuyện cực kì lịch sự. Dịch vụ của tiki chưa bao giờ làm mình thất vọng. Muốn chấm hẳn 10 sao ❤️
* Hài lòng với chất lượng dịch vụ
  + VD: Nhân viên hỗ trợ khách hàng tốt.
* Vài yếu tố chưa rõ hay lỗi vặt nhưng tổng thể vẫn đánh giá cao
  + VD: chưa kiểm tra nên chưa biết chất lượng. giao hàng, đóng gói tốt, nhanh
  + VD: Chưa dùng thử nhưng rất đẹp

**2.1.2. Negative**

Là câu mang ý nghĩa tiêu cực, thể hiện sự không hài lòng, khó chịu với sản phẩm nhận được:

* Thất vọng với sản phẩm
  + VD: Pin bị lỗi, thân pin bị phù nhẹ, cắm sạc thì báo pin ảo hay sao ý, mới cắm vào sạc đã lên 2 đèn và chớp đèn thứ 3, ipad đang 64% cắm pin dự phòng vào sạc thử đến khi 91% thì pin dự phòng tắt đèn và hết pin. Hơi thất vọng vì giá cả lẫn chất lượng sản phẩm.
* Bất mãn với số review tích sự so với chất lượng sản phẩm
  + VD: Không thể tin nổi sản phẩm như thế này mà được review 5 sao
* Tức giận với chất lượng
  + VD: cực kỳ tệ, xả chậm như rùa, làm gì có 3.0, lừa đảo
* Review chất lượng sản phẩm sau khi sử dụng
  + VD: đã hư sau 1 tuần sử dụng
  + VD: Thời gian sử dụng lâu
* Có khen nhẹ tuy nhiên sau đó phản ánh nhiều thể hiện không hài lòng
  + VD: Màu sắc đẹp. Nhưng không sạc nhanh được. Này phiên bản sạc thường
* Không hài lòng với việc giao hàng
  + VD: đối tác vận chuyển Best giao hàng quá lâu

**2.1.3. Neutral**

Là các câu mang ý nghĩa trung tính: không tiêu cực cũng không tích cực, hay không liên quan đến sản phẩm cũng như đánh giá sản phẩm ở mức trung bình (không cao không thấp)

* VD: Tiền nào của đó
* VD: có bán các loại khẩu trang
* VD: Ko có xem nên ko biết
* VD: Cũng đc, hơi nặng