Bài 1: Bài toán đặt ra như sau: làm thế nào để biết xem có bao nhiêu người bình luận tích cực, tiêu cực, hoặc trung tính (neutral) trên 1 post. Từ đó có thể tìm hiểu được xem xu hướng của người dùng tương tác với một bài viết đó như thế nào, và cách cộng đồng mạng phản ứng ra sao với nội dung bài viết.

Giả sử đã thu thập được một tập hợp các bài viết trên mạng xã hội (file sample_text_data.txt). Các bạn hãy thực hiện các yêu cầu sau.

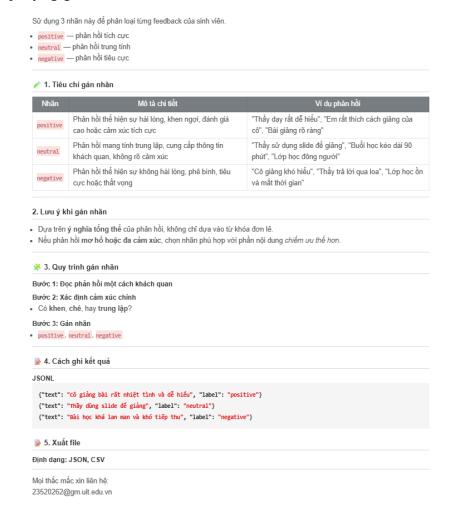
- 1. Các bạn hãy xây dựng hướng dẫn gán nhãn (annotation guidelines) cho bài toán.
- 2. Hãy sử dụng công cụ Doccano, thực hiện annotation cho các bình luân.
- 3. Hãy tiến hành gán nhãn cho các dữ liệu trong tập data.
- (Làm nhóm) Hãy tính độ đồng thuận giữa các thành viên khi gán nhãn. Và cho biết kết quả

Bài làm:

Thu thập dữ liệu:

Sử dụng bộ dữ liệu <u>UIT-VSFC</u>, sử dụng 50 sample đầu tiên để annotate. Em format theo text, lable như template của doccano

Xây dựng guidelines:

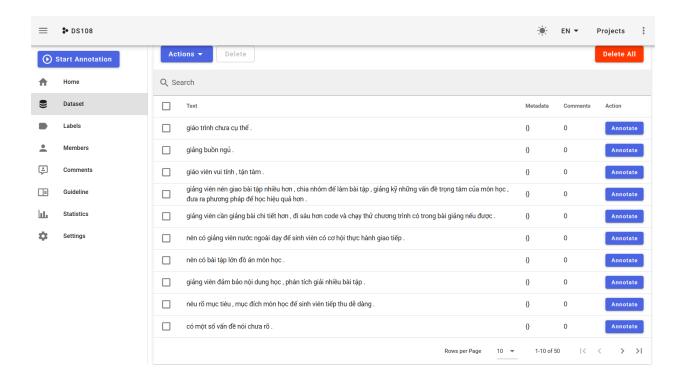


Sử dụng Doccano:

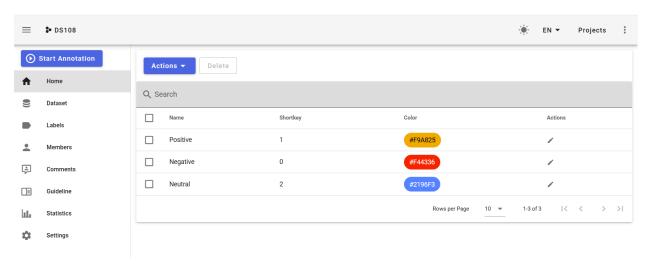
Tạo project dạng text classification

Add Project		
•	Project name DS108	
Ê	Description annotate	
· <u>···</u> ·	Project type Text Classification	
	Allow single label	
	Randomize document order	
	Share annotations across all users	

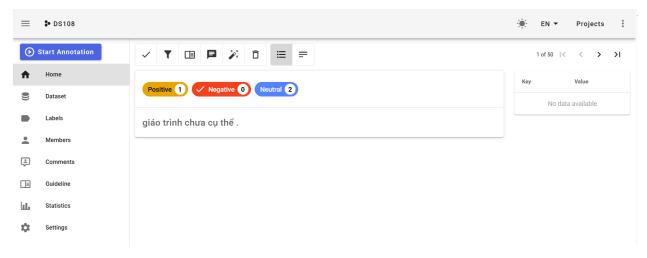
Import dataset



Tạo label



Tiến hành gán nhãn



Tính độ đồng thuận giữa các thành viên khi gán nhãn. Và cho biết kết quả (Do có 1 mình em làm bộ này em xin được tính trên của mình luôn)

Bài 2: Bài toán đặt ra như sau: các bình luận bằng ảnh chế mang tính chất hài hước trên mạng xã hội (gọi là Meme) đang là cách phổ biến của người dùng hiện nay. Tuy nhiên, bên cạnh các meme hài hước và mang tính chất giải trí, có những meme lại mang tính chất miệt thị và xúc phạm người khác (có thể xúc phạm người, hoặc thậm chí là một cộng đồng). Yêu cầu đặt ra là hãy xây dựng một bộ dữ liệu ảnh để xác định xem một Meme có mang tính chất xúc phạm hay không.

Giả sử đã thu thập được một tập dữ liệu ảnh gồm các meme từ các bài viết trên mạng xã hội (file sample_image_data.zip), các bạn hãy thực hiện các yêu cầu sau:

- Các bạn hãy xây dựng hướng dẫn gán nhãn (annotation guidelines) cho bài toán.
- Hãy sử dụng công cụ Doccano, thực hiện annotation cho các bình luận.
- Hãy tiến hành gán nhãn cho các dữ liệu trong tập data.
- 4. (Làm nhóm) Hãy tính độ đồng thuận giữa các thành viên khi gán nhãn. Và cho biết kết quả.

Bài làm:

Thu thập dữ liệu:

Sử dụng bộ dữ liệu <u>Facebook Hateful Meme Dataset</u>, sử dụng 50 sample đầu tiên để annotate.

Xây dựng guidelines:

Mô Tả Bài Toán: Phát Hiện Meme Xúc Phạm

Mục tiêu

Meme (ảnh chế) đang là một hình thức giao tiếp trực tuyến phổ biến, đặc biệt trên các nền tảng mạng xã hội như Facebook, Instagram, Reddit... Tuy nhiên, không phải meme nào cũng là hài hước và vô hại. Một số meme có thể:

- Miệt thị cá nhân
- · Kích động thủ ghét
- · Xúc phạm cộng đồng, tôn giáo, giới tính, dân tộc v.v.

Mục tiêu của bài toán là xây dựng một hệ thống tự động để phân loại một meme (dưới dạng hình ảnh) là:

- hateful xúc phạm, mang tính chất miệt thị, công kích cá nhân hoặc cộng đồng
- non-hateful không xúc phạm, mang tính chất hài hước, châm biếm lành mạnh hoặc trung tính

🎍 Tiêu chí xác định ảnh xúc phạm

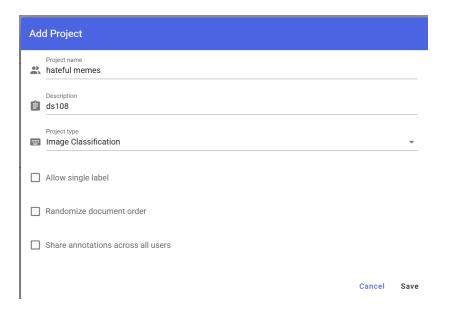
Nhãn	Mô tả
offensive	Có yếu tố miệt thị, xúc phạm cá nhân hoặc cộng đồng, thù ghét, châm biếm nặng nề
non- offensive	Mang tính hài hước thông thường, châm biếm lành mạnh, không xúc phạm

Dịnh dạng dữ liệu

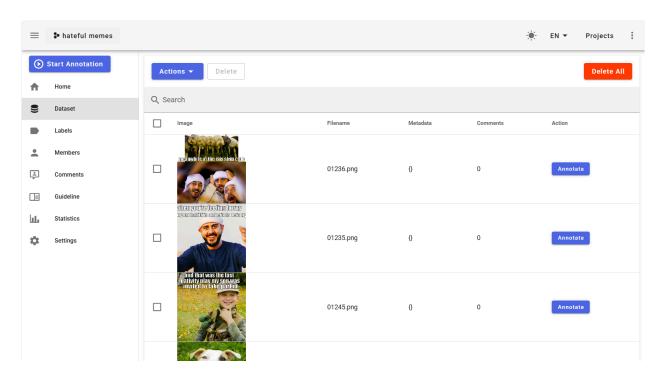
JSONL, CSV

Sử dụng Doccano:

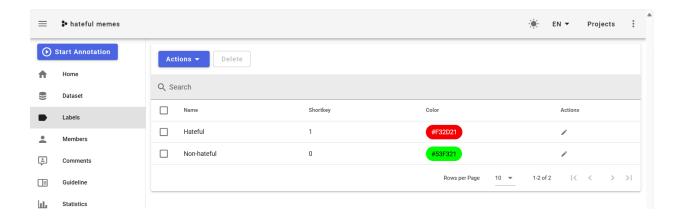
Tạo project dạng text classification



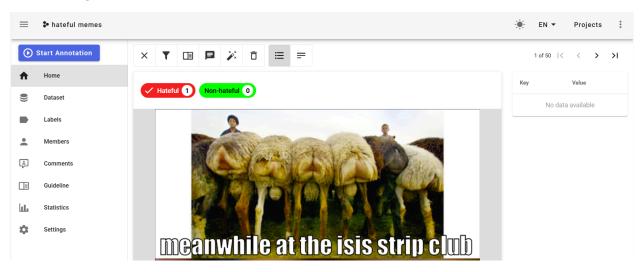
Import dataset



Tạo label



Tiến hành gán nhãn



Tính độ đồng thuận giữa các thành viên khi gán nhãn. Và cho biết kết quả (Do có 1 mình em làm bộ này em xin được tính trên của mình luôn)