

Bộ dữ liệu Hate Speech Dataset

Hướng dẫn gán nhãn

Thông tin bộ dữ liệu: Bộ dữ liệu này là tập hợp các bình luận (comments) từ các bài viết (posts) trên mạng xã hội Facebook.

Mục tiêu bộ dữ liệu: Huấn luyện (training) cho một hệ thống phân lớp đa lớp dùng để xác định xem một bình luận trên mạng xã hội có mang tính xúc phạm hay không. Đồng thời xác định xem bình luận đó xúc phạm tới đối tượng nào.

Nhãn của bộ dữ liệu: Bộ dữ liệu gồm 4 nhãn, mỗi nhãn và ví dụ sẽ được mô tả trong bảng sau.

BẢNG 1: MÔ TẢ Ý NGHĨA TỪNG LOẠI NHÃN.

Loại nhãn	Mô tả	Ví dụ
Not mentioned	Bình luận có nội dung không liên quan tới đối tượng hướng tới mặc dù có xúc phạm hay không.	Ví dụ được xét trên 5 đối tượng là cá nhân, tổ chức/tập thể, chính trị, đồ vật và động vật.
Clean	Bình luận có nội dung liên quan tới đối tượng hướng tới nhưng không có chứa các từ ngữ nhạy cảm và không mang hàm ý xúc phạm tới cá nhân, tổ chức của đối tượng đang đề cập . Lưu ý: Những từ ngữ nhạy cảm là những từ tục tĩu bao gồm cả những từ viết tắt như đm, vcl, vãi lồn,...	Ví dụ 1: “Thầy đéo xóa kênh kêu gọi học trò report cho bay kênh con mẹ nó luôn giờ.” - Xét trên đối tượng là cá nhân: Có nội dung đề cập tới đối tượng “thầy”, có các từ ngữ nhạy cảm như “đéo”, “con mẹ nó luôn” nhưng không có ý nghĩa xúc phạm -> Gán nhãn 2 cho đối tượng là cá nhân. - Xét trên đối tượng là chính trị, tổ chức, đồ vật, động vật thì thấy nội dung không có liên quan tới các đối tượng này -> Gán nhãn 0 cho các đối tượng là chính trị, tổ chức/tập thể, đồ vật, động vật.
Offensive	Bình luận có nội dung liên quan tới đối tượng hướng tới. Có ý nghĩa tiêu cực nhưng không mang hàm ý xúc phạm tới các cá nhân, tổ chức của đối tượng đang đề cập . Có thể chứa các từ ngữ nhạy cảm hoặc không chứa. Có 2 trường hợp xảy ra cho bình luận phản cảm: <ul style="list-style-type: none">- Trường hợp 1: Bình luận có các từ ngữ nhạy cảm, tục tĩu. Nhưng không mang hàm ý xúc phạm.- Trường hợp 2: Bình luận không có các từ ngữ nhạy cảm, tục tĩu. Không mang hàm ý xúc phạm nhưng gây ra cảm giác khó chịu cho người đọc. Lưu ý: Những từ ngữ nhạy cảm là những từ tục tĩu bao gồm cả những từ viết tắt như đm, vcl, vãi lồn,...	Ví dụ 2: “có một sự tồn tại của đàn bò mang tên “ cực kỳ thuyết phục “, có sở thích là nhai đi nhai lại, mục đích của chúng trên hành tinh này vẫn chưa có lời giải đáp, theo như các nhà động vật học cho biết loài bò này có lối sống hoang dã truy lạc, có thể động dục bất cứ lúc nào, con số của loài này đang được gia tăng đáng kể.”

<p>Hate</p>	<p>Bình luận có nội dung liên quan tới đối tượng hướng tới. Mang ý nghĩa xúc phạm tới các cá nhân, tổ chức của đối tượng đang đề cập. Có thể có chứa các từ ngữ nhạt cảm hoặc không chứa.</p> <p>Có 3 trường hợp sau xảy ra cho bình luận xúc phạm:</p> <ul style="list-style-type: none"> - Trường hợp 1: Bình luận mang các từ ngữ tục tĩu, nhậy cảm xúc phạm trực tiếp tới các cá nhân, tổ chức. - Trường hợp 2: Bình luận không mang các từ ngữ tục tĩu, nhậy cảm nhưng lại có hàm ý, ý nghĩa xúc phạm tới các cá nhân, tổ chức. - Trường hợp 3: Bình luận không mang các từ ngữ tục tĩu, nhạt cảm nhưng lại có hàm ý, ý nghĩa xúc phạm được ẩn dụ qua các hình ảnh, các từ ngữ khác. Những bình luận này rất khó để nhận biết vì người gán nhãn phải có những hiểu biết, thông tin về những đối tượng được đề cập trong bình luận. 	<p>- Xét trên đối tượng là cá nhân, tổ chức / tập thể: Tuy các từ ngữ, nội dung thô không đề cập tới đối tượng là cá nhân, tổ chức/ tập thể, cũng không có các từ ngữ nhậy cảm nhưng hàm ý ẩn dụ thì có và mang ý nghĩa xúc phạm tới các cá nhân, tổ chức/ tập thể. -> Gán nhãn 3 cho các đối tượng là cá nhân, tổ chức/ tập thể.</p> <p>- Xét trên đối tượng là động vật: Ta thấy nội dung thô có đề cập tới động vật là loài bò, và nếu xét theo đối tượng câu nói đó muốn đề cập là loài bò thì đây là một câu nói hoàn toàn bình thường, không có ý nghĩa xúc phạm và không có các từ ngữ nhậy cảm. -> Gán nhãn 1 cho đối tượng là động vật.</p> <p>- Xét trên đối tượng là đồ vật, chính trị: Ta thấy nội dung thô lẫn hàm ý đều không có đề cập tới đồ vật hay chính trị. -> Gán nhãn 0 cho đối tượng là đồ vật, chính trị.</p>
--------------------	---	---

Cách gán nhãn:

Ứng với mỗi câu – Là một bình luận trên mạng xã hội Facebook và mỗi chủ đề, mỗi người gán nhãn (Annotator) gán cho mỗi câu trên mỗi chủ đề 1 trong 4 giá trị sau ứng với 4 nhãn trong bộ dữ liệu:

0: Not mentioned

1: Clean

2: Offensive

3: Hate

Ví dụ:

Bình luận	Tôn giáo	Chính trị	Cá nhân	Đồ vật	Quốc gia	Vùng miền	Tổ chức/ tập thể	Dân tộc	Động vật
Đúng là bọn mất híp lò xo thật :))) bên việt nam t cái này ra cách đây 10 năm r và bọn t gọi là cái L :)))	3	3	3	0	3	3	3	3	0
CÔNG ĐỒ CỤC SỨC VÔ NHÂN TÍNH ĐỀ NGHỊ VN. NHÀ NƯỚC VN BAN THƯỞNG.	3	1	3	0	0	0	3	3	0
Thế mà mình nói mấy thằng bắc kì, bọn đó lại bầu vô nói mình phân biệt! :D	0	0	3	0	0	3	3	2	0
Loại này cho dựa cột thôi chứ độ thể nào	2	0	3	0	0	0	3	3	1
Ủa chứ bữa xin lỗi khóc tể om xòm mà giờ chửi nữa rồi bó tay con này hỗn quá	0	0	2	0	0	0	0	0	0
Thật chán với chế độ quản lý dân sinh	0	3	0	0	3	3	3	0	0
Nguyễn Hữu Nhân đốp cả của cháu. Thử mắc dây	0	0	3	0	0	0	0	0	1
Nhậu cái đm chúng mày lũ chó hại dân hại nước	3	0	3	0	0	0	3	3	0
mài có óc để suy nghĩ ko add? cái đồng bằng đó mãi tính đào hồ bnhieu cho đủ hả thằng ngu?	0	0	3	0	0	3	3	0	0
Chỉ 3 vàng ngọc như con heo nọc nhà e	0	0	0	0	0	0	0	0	1
thầy đéo xóa kênh kêu gọi học trò report cho bay kênh thiệt luôn :))	0	0	2	0	0	0	0	0	0
Anh ba lão nháo là Tú tiên giờ	0	0	2	0	0	0	0	0	0
me, sao bộ k mở cửa hàng kinh doanh bán bình luôn cho dân, ko lại phạt vì mua phải bình đếu :)))	0	2	3	1	0	0	3	0	0
Chỉ 3 cái gì cung co nhưng cái lờ la chỉ 3 hết kiếp này cung k bao gio co .	0	0	3	0	0	0	0	0	3
Khắc Thành Nguyễn phục cái máu lờn	3	3	3	0	3	3	3	3	2
Mấy thằng già làm bộ trưởng này nọ mà đầu óc thua cả 1 bác nông dân	0	3	3	0	2	2	3	2	0
có một sự tồn tại của đàn bò mang tên " cực kỳ thuyết phục ", có sở thích là nhai đi nhai lại, mục đích của chúng trên hành tinh này vẫn chưa có lời giải đáp, theo như các nhà động vật học cho biết loài bò này có lối sống hoang dã truy lạc, có thể động dục bất cứ lúc nào, con số của loài này đang được gia tăng đáng kể.	3	3	3	0	3	3	3	3	1

Hình 1: Hình ảnh một số bình luận đã được gán nhãn.

Lưu ý:

- Một số bình luận trống hoặc chỉ có những icon hoặc chỉ có những kí tự đặc biệt thì mặc định gán nhãn 0 cho mọi chủ đề.
- Một bình luận có thể đề cập tới đa đối tượng và những đối tượng đều độc lập với nhau. Vì vậy người gán cần chú ý là cùng một bình luận nhưng có đối tượng sẽ được gán nhãn **Hate**, có đối tượng được gán nhãn **Clean**.
- Trong một câu nói, người gán nhãn cần suy xét cả ý nghĩa thô của nội dung và ý nghĩa ẩn dụ của nội dung đó. Nếu người gán không hiểu hoặc không biết ý nghĩa ẩn dụ thì ưu tiên xét ý nghĩa thô.