

# Task proposal

Our task idea is to take Reddit comments from a variety of subreddits, and annotate them with a family-friendliness label. We are focusing on English comments, and, for simplicity, we will judge the family-friendliness solely using the document(text), ignoring other contexts such as links, attachments, and other media files.

1. The documents for our task are Reddit comments with character lengths ranging from 300 to 400 characters. We want to be good at classifying diverse documents, so we prioritize the number of posts in our training set; however, we don't want it to be so short that it is not a good training document for our model.
2. Our labels are integer values on a Likert scale ranging from 1 to 5 where 1 is the model strongly disagrees that the comment is family-friendly, and 5 is the model strongly agrees vice versa. We don't use real valued labels, because it is very difficult to write a concrete and fair annotation guidelines. Our concern is that if we used real-valued label, our guideline would have to be excessively specific to get good Inter-annotator agreement score.
3. One of the reasons that this task is interesting is that it has strong applications in the emerging field of age-based content filtering. Secondly, we believe that an automated tool for assessing family friendliness would be a practical for parents. On top of that, even though Reddit labels subreddits for explicit content, the comments themselves are not rated. Last but not least, it is an interesting corpus from social media, composed of text written by general population as opposed to professional writers.
4. Our data source is Reddit, specifically comments gathered from the top 200 subreddits. We sample n number of comments from each subreddit through pushshiftAPI.
5. According to the Reddit terms of use ([https://www.reddit.com/wiki/api-terms/#wiki\\_3.\\_fees.3B\\_restrictions\\_on\\_use](https://www.reddit.com/wiki/api-terms/#wiki_3._fees.3B_restrictions_on_use)), section 2.d: "Reddit user photos, text and videos ("User Content") are owned by the users and not by Reddit. Subject to the terms and conditions of these Terms, Reddit grants You a non-exclusive, non-transferable, non-sublicensable, and revocable license to copy and display the User Content using the Reddit API through your application, website, or service to end users. You may not modify the User Content except to format it for such display. You will comply with any requirements or restrictions imposed on usage of User Content by their respective owners, which may

include "all rights reserved" notices, Creative Commons licenses or other terms and conditions that may be agreed upon between you and the owners.". So as long as we don't modify user generated content, and comply with user requests and licenses, we can display it.