SI 618 Fall 2021 HW 7 – Mturk

In this homework, we will utilize Amazon Mechanical Turk (Mturk) to collect labels for a dataset of comments posted on the web. Our goal here is to measure toxicity in discussions on news websites.

Data

Download the file hw7_comments.csv given with the homework. It contains 10000 comments on news websites sampled from a dataset of 1.8 million comments was created for research (Borkan, D., Dixon, L., Sorensen, J., Thain, N., & Vasserman, L. (2019, May)). It has since been used for a large-scale Kaggle competition on conversational bias and toxicity with additional columns. The sample given to you has been balanced based on previous crowd annotation such that approx. 50% of the comments were labeled by at least 50% workers as toxic and the rest were not.

Your team will only use a random sample of 50 comments. Note that you can sample rows randomly either with Pandas or Spark.

Step 1 – Mturk registration

Each team needs to create an account on MTurk as a requester (https://requester.mturk.com/). You will need load funds into that account (only one is needed per team). You need to run your tasks and pay against that account. Once you are done, you should email Rebecca Epstein (rebeps@umich.edu) and request reimbursements with the subject line: **Mturk Reimbursement Request, SI 618, Student Name**. Rebecca will work with you to get the information required for reimbursements. Note that there needs to be one representative who will be responsible for reimbursing other group members.

Step 2 – Setting up the Mturk task

First, in order to make things easy, we will start with an existing project template. Select an existing project template and create a new project.

Creating an MTurk task requires four high level sub-tasks.

1. Properties:

Title: choose a title that is representative of the task at hand
Description: Provide a 1-2 sentence description of the task. You can see the screenshot of the instructions below to get some inspiration.
Keywords: Use “news discussions, toxic language, classification”

Reward per assignment (HIT): $ 0.14, number of assignments (workers) per task: 3, time allotted: 1 hour, hit expires: 7 days, auto-approve: 3 days

Master workers not required
Workers must: (1) have at least 1000 approved HITs, and (2) have at least 98% approval rate.

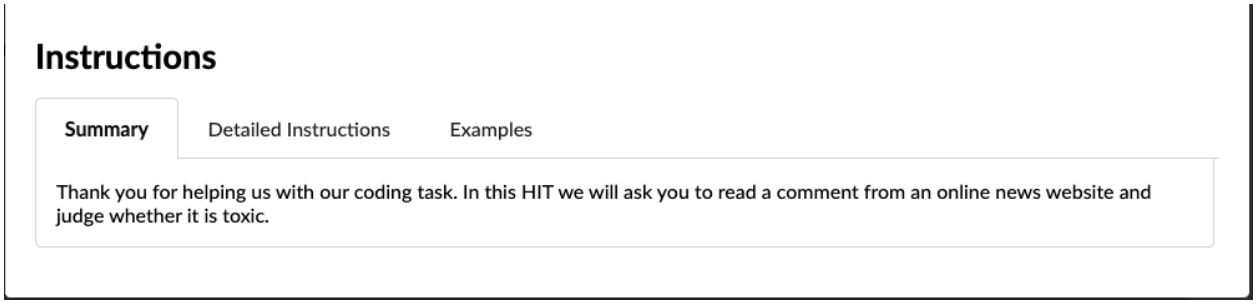
We are going to create 50 HITS (50 comments * one comment per HIT).
Total amount of fee: $0.14 per worker * 3 workers per Hit * 50 HITS = 21 + 4.5(MTurk Fee) = $25.5

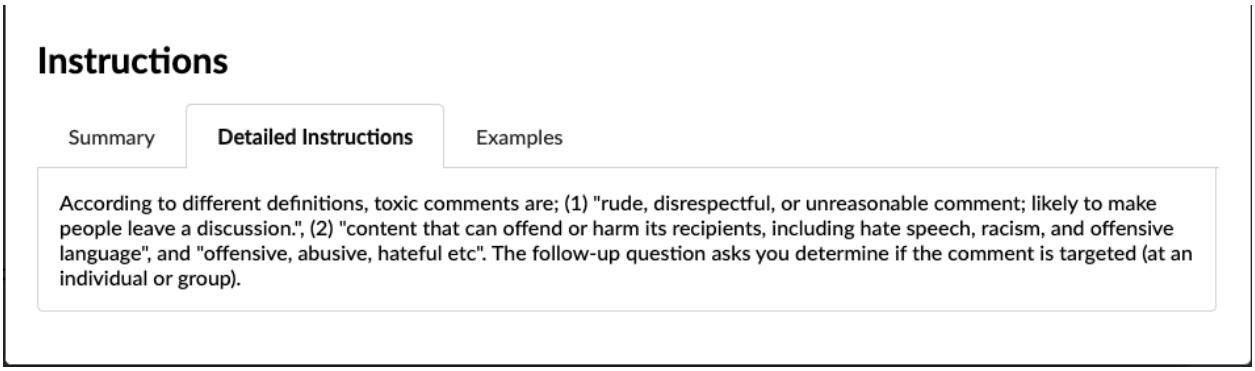
We encourage you to first test your solution by releasing a smaller batch, e.g., 10 HITS. This can allow you to catch any issues before spending a substantial budget. You can use up to $4.5 for these tests. Please do not exceed $30 ($25.5 for the main task + $4.5 for tests) as this is the imbursement limit. However, if any unforeseen issues arise, please reach out to the instructional team.

2. Design Layout & Mturk input file preparation

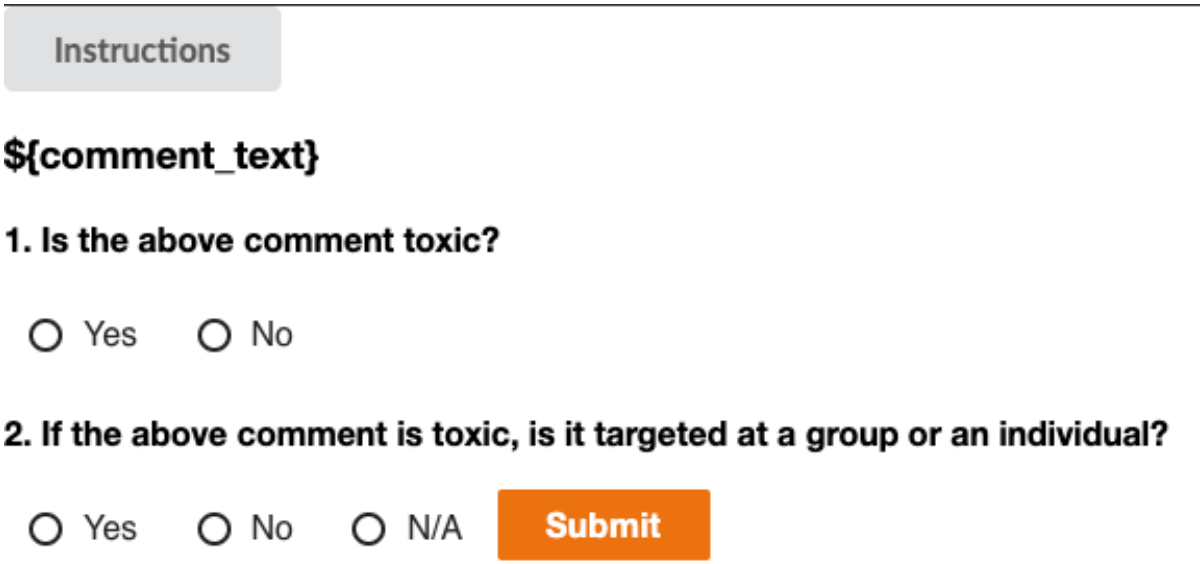
Observe the design layout code written with **Crowd HTML tags** under the "Design Layout" tab. You can see the corresponding layout under the tab "Preview and Finish". Edit the layout code to create the structure seen by the following screenshots.

1. Instructions





2. Questions



You should use radio button groups for the questions. You should use the name and value fields of the input elements such that the in the output file your response fields will have the following column names. You can verify if your column names will be correct by selecting some responses and submitting on the “Preview and Finish” tab.

Question 1 - Yes	toxic1.yes
Question 1 - No	toxic2.no
Question 2 - Yes	target1.yes
Question 2 - No	target2.no
Question 2 - N/A	target3.na

3. Preview and Finish:
This page will show a preview of the task. You can also see the csv file format that your task demands (that can be a clue to you if the template does not match the template csv file we provided).

4. Publish batch:

For this step, you will upload the csv file with your selected random sample of comments. Before you publish you will see the cost (make sure it matches our estimation) and can see some example HITs.
After some time of publishing the batch, you will have workers picking up your HITs. In a relatively short while, your HITs will most likely be completed. But don't leave this to the last minute. The earlier you do this homework, the earlier you can make sure that the HITs are all completed.

Step 4 – Computation
You can check the progress of the batch. Once you have the result, change the result csv file name to uniqname1_uniqname2_si618_hw7_batch_result.csv.

Create uniqname1_uniqname2_si618_ hw7_compute.py. Compute the fraction of comments that were labeled as toxic by at least one, at least 2 and at least 3 workers. Also compute the fraction of toxic comments (agreed by at least 2 workers) that are also labeled as targeted at least once.

toxic_1<tab>XXXX
toxic_2<tab>XXXX
toxic_3<tab>XXXX
targted<tab>XXXX

What to submit:

Mturk result CSV file: uniqname1_uniqname2_si618_hw7_batch_result.csv

Compute result txt file: uniqname1_uniqname2_si618_hw7_compute.txt

Python file: uniqname1_uniqname2_si618_hw7_compute.py

One submission per team is fine! Make sure both uniqnames are on all files when submitted!