

SI 618 Fall 2022 HW 4

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 judge the position a tweet takes towards the topics of immigration, guns and lgbtq.

Please do this homework on only one assigned topic for your team (immigration or guns or lgbtq).

Data

Depending on your chosen topic, download one of the datasets

- hw4_immigration.tsv
- hw4_lgbtq.tsv
- hw4_guns.tsv

Each dataset will have data points spanning across three unique months.

This data was collected from the Twitter decahose using a [set of keywords](#) from [this paper](#), which was focused on evaluating keyword expansion techniques for social movement research. Three months selected for analysis (each high- and low-activity month for each movement) are also directly from that paper. The only additional filtering done was excluding retweets and restricting data collection to English-language tweets (based on the language field in the tweet object).

Your team should only use a truly random sample of 75 tweets selecting 25 random tweets per month. Use some random() function for the sampling and please DO NOT just take first k or put some sequence of indices.

Save your random sampling python code in a [username1_username2_si618_hw4_data_sampling_<topic>.py](#)

Note: When you publish the MTurk HITs, you MUST keep the tweet_id column in your sampled dataset that you upload to MTurk. You will get a warning for not using this column's data in the crowdfunder but this can be ignored. If you fail to add the tweet_id column, you would not be able to do the second computation task in this homework!

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 “tweets, classification, immigration (or) lgbtq (or) guns, social media” |
| Reward per assignment (HIT) | : \$ 0.15 |
| 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:

- have at least 1000 approved HITs
- have at least 98% approval rate.

We are going to create 75 HITS (75 tweets * one comment per HIT).

Total amount of fee: \$0.15 per worker * 3 workers per Hit * 75 HITS = 33.75 + 6.75(MTurk Fee 20%) = \$40.5

We encourage you to first test your solution by releasing a smaller batch, e.g., 8 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 \$45 (\$40.5 for the main task + \$4.5 for tests) as this is the reimbursement 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”. You can add the below definitions for instructions as you edit the layout code to create a structure similar to what is seen in the Sample Instructions. Please use only one of the research definitions (Immigration/Guns/LGBTQ) based on your topic.

Research Definitions:

Immigration

Based on the following social media post, answer whether the author is (1) pro-immigration , (2) anti-immigration, or (3) neutral towards immigration and/or their position is unclear from the text.

1. pro-immigration (e.g. they are in favor of increased immigration or advocate for immigrants’ well-being)
2. anti-immigration (e.g. prefer reducing immigration or criticize pro-immigration talking points)
3. neutral/unclear (author is neutral towards immigration and/or their position is unclear from the text)

Guns

Based on the following social media post, answer whether the author (1) supports gun control, (2) supports gun rights, or (3) is neutral towards guns and/or their position is unclear from the text.

1. supports gun control (e.g. stricter policies about who may purchase guns, what types of guns should be legal, where carrying guns may be permitted, or argues that guns are unnecessary and cause undue harm)
2. supports gun rights (e.g. less restriction on gun ownership and usage, or argues that guns are necessary for protection)
3. neutral/unclear (author is neutral towards guns and/or their position is unclear from the text)

LGBTQ

Based on the following social media post, answer whether the author (1) supports LGBTQ movements, (2) opposes some LGBTQ movements, or (3) is neutral towards LGBTQ rights and/or their position is unclear from the text.

1. supports LGBTQ movements (e.g. believes that LGBTQ people should be fully accepted by society, have the same opportunities as non-LGBTQ people, or advocates for LGBTQ people's well-being)
2. opposes some LGBTQ movements (e.g. believes that LGBTQ people should not be fully accepted by society, clash with traditional/religious values, or that LGBTQ activism has gone too far)
3. neutral/unclear (author is neutral towards LGBTQ movements and/or their position is unclear from the text)

Sample Instructions (Immigration)

Instructions

SummaryDetailed InstructionsExamples

Thank you for helping us with our classification task. In this task, we ask you to read a tweet answer whether the author is (1) pro-immigration , (2) anti-immigration, or (3) neutral towards immigration and/or their position is unclear from the text.

Figure 1

Instructions

Summary**Detailed Instructions**Examples

According to our research definitions, a tweet can be classified as

- pro-immigration (e.g. they are in favor of increased immigration or advocate for immigrants' well-being)
- anti-immigration (e.g. prefer reducing immigration or criticize pro-immigration talking points)
- neutral/unclear (author is neutral towards immigration and/or their position is unclear from the text)

In the follow up questions, we ask you to judge whether the tweet was directed at a particular individual or group.

Figure 2

- **Questions**

You will be asking two questions in this task. The first one will depend on your topic chosen. The second question is common for all three topics.

Tweet : ``${tweet_text}``

Is the text of this tweet pro-immigration, anti-immigration or neutral towards immigration (or) unclear?

☐ Pro-immigration

☐ Anti-immigration

☐ Neutral/Unclear

Is the text of this tweet targeted at a particular individual or group?

☐ Yes

☐ No

☐ N/A

Submit

Figure 3

You should use radio button groups for the questions. You should edit the fields of the input elements such that 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.

For Question 1: (pick one set of three columns depending on your chosen topic)

| Radio Button Option | Column name |
|-------------------------------|------------------------------|
| Question 1 – Pro-immigration | immigration.pro-immigration |
| Question 1 – Anti-immigration | immigration.anti-immigration |
| Question 1 – Neutral/Unclear | immigration.neutral/unclear |

| | |
|---|------------------------------------|
| Question 1 – Supports LGBTQ Movements | lgbtq.supports-lgbtq-movements |
| Question 1 – Opposes some LGBTQ Movements | lgbtq.opposes-some-lgbtq-movements |
| Question 1 – Neutral/Unclear | lgbtq.neutral/unclear |

| | |
|-----------------------------------|---------------------------|
| Question 1 – Supports gun control | guns.supports-gun-control |
| Question 1 – Supports gun rights | guns.supports-gun-rights |
| Question 1 – Neutral/Unclear | guns.neutral/unclear |

For Question 2 (common for all three topics)

| | |
|------------------|--------------|
| Question 2 – Yes | targeted.yes |
| Question 2 – No | targeted.no |
| Question 2 – NA | targeted.n/a |

Please save the full html code you write in the editor under the Design Layout tab as `username1_username2_si618_hw4_crowd_form_design_<topic>.html`

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 Tasks

For this step, you will upload the csv file with your selected random sample of tweets. 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. You can check the progress of the batch after publishing. Once you have the result, change the result csv file name to `username1_username2_si618_hw4_batch_result_<topic>.csv`.

Step 3 – Computation

Create `username1_username2_si618_hw4_compute_<topic>.py` for writing the computation tasks.

For Immigration:

Computation Tasks:

You can consider a tweet to be a pro-immigration-majority/anti-immigration-majority/neutral-majority tweet if at least two workers agreed that it is pro-immigration/anti-immigration/neutral.

1. Compute the fraction of tweets that were labeled as pro-immigration, anti-immigration and neutral about immigration by at least one, at least two and three workers.
2. Compute the ratio of pro-immigration-majority, anti-immigration-majority, neutral-majority, no-majority tweets for each month (Hint: use the tweet's timestamp).
Please sort the results by ascending order of months before printing out to the results file.
3. Compute the fraction of pro-immigration-majority tweets that are also labeled as targeted at least once. Do the same for anti-immigration-majority and neutral/unclear-majority tweets.

Result Format:

Write the results to a `username1_username2_si618_hw4_computations_immigration.txt` file in the following format.

```
pro-immigration1<tab>XX<tab>anti-immigration1<tab>XX<tab>neutral-immigration1<tab>XX<tab>
pro-immigration2<tab>XX<tab>anti-immigration2<tab>XX<tab>neutral-immigration2<tab>XX<tab>
pro-immigration3<tab>XX<tab>anti-immigration3<tab>XX<tab>neutral-immigration3<tab>XX<tab>
```

```
<month1><year1>
```

```
pro-immigration-majority<tab>XX<tab>anti-immigration-majority <tab>YY<tab>neutral-majority<tab>ZZ<tab>no_majority<tab>AA
```

```
<month2><year2>
```

```
pro-immigration-majority<tab>XX<tab>anti-immigration-majority <tab>YY<tab>neutral-majority<tab>ZZ<tab>no_majority<tab>AA
```

```
<month3><year3>
```

pro-immigration-majority<tab>XX<tab>anti-immigration-majority <tab>YY<tab>neutral-majority<tab>ZZ<tab>no_majority<tab>AA

targeted_ pro-immigration-majority<tab>XX<tab>targeted_ anti-immigration-majority<tab>YY<tab>targeted_ neutral-majority<tab>ZZ

For LGBTQ:

Computation Tasks:

You can consider a tweet to be a supports-lgbtq-movements-majority/opposes-some-lgbtq-movements-majority/neutral-majority tweet if at least two workers agreed that it supports-lgbtq-movements/opposes-some-lgbtq-movements/neutral.

1. Compute the fraction of tweets that were labeled as supports-lgbtq-movements, opposes-some-lgbtq-movements and neutral about lgbtq by at least one, at least two and three workers.
2. Compute the ratio of supports-lgbtq-movements-majority, opposes-some-lgbtq-movements-majority, neutral-majority, no-majority tweets for each month (Hint: use the tweet's timestamp). Please sort the results by ascending order of months before printing out to the results file.
3. Compute the fraction of supports-lgbtq-movements-majority tweets that are also labeled as targeted at least once. Do the same for opposes-some-lgbtq-movements-majority and neutral/unclear-majority tweets.

Result Format:

Write the results to a username1_username2_si618_hw4_computations_lgbtq.txt file in the following format.

supports-lgbtq-movements1<tab>XX<tab>opposes-some-lgbtq-movements1<tab>XX<tab>neutral-lgbtq1<tab>XX<tab>
supports-lgbtq-movements2<tab>XX<tab>opposes-some-lgbtq-movements2<tab>XX<tab>neutral-lgbtq2<tab>XX<tab>
supports-lgbtq-movements3<tab>XX<tab>opposes-some-lgbtq-movements3<tab>XX<tab>neutral-lgbtq3<tab>XX<tab>

<month1><year1>

supports-lgbtq-movements-majority<tab>XX<tab>opposes-some-lgbtq-movements-majority <tab>YY<tab>neutral
majority<tab>ZZ<tab>no_majority<tab>AA

<month2><year2>

supports-lgbtq-movements-majority<tab>XX<tab>opposes-some-lgbtq-movements-majority <tab>YY<tab>neutral-
majority<tab>ZZ<tab>no_majority<tab>AA

<month3><year3>

supports-lgbtq-movements-majority<tab>XX<tab>opposes-some-lgbtq-movements-majority <tab>YY<tab>neutral-
majority<tab>ZZ<tab>no_majority<tab>AA

targeted_ supports-lgbtq-movements-majority<tab>XX<tab>targeted_ opposes-some-lgbtq-movements-
majority<tab>YY<tab>targeted_ neutral-majority<tab>ZZ

For Guns:

Computation Tasks:

You can consider a tweet to be a supports-gun-control-majority/supports-gun-rights-majority/neutral-majority tweet if at least two workers agreed that it supports-gun-control/supports-gun-rights/neutral.

1. Compute the fraction of tweets that were labeled as supports-gun-control, supports-gun-rights and neutral about guns by at least one, at least two and three workers.
2. Compute the ratio of supports-gun-control-majority, supports-gun-rights-majority, neutral-majority, no-majority tweets for each month (Hint: use the tweet's timestamp). Please **sort the results by ascending order of months** before printing out to the results file.
3. Compute the fraction of supports-gun-control-majority tweets that are also labeled as targeted at least once. Do the same for supports-gun-rights-majority and neutral/unclear-majority tweets.

Result Format:

Write the results to a `username1_username2_si618_hw4_computations_guns.txt` file in the following format.

```
supports-gun-control1<tab>XX<tab>supports-gun-rights1<tab>XX<tab>neutral-guns1<tab>XX<tab>
supports-gun-control2<tab>XX<tab>supports-gun-rights2<tab>XX<tab>neutral-guns2<tab>XX<tab>
supports-gun-control3<tab>XX<tab>supports-gun-rights3<tab>XX<tab>neutral-guns3<tab>XX<tab>
```

```
<month1><year1>
supports-gun-control-majority<tab>XX<tab>supports-gun-rights-majority <tab>YY<tab>neutral-
majority<tab>ZZ<tab>no_majority<tab>AA
<month2><year2>
supports-gun-control-majority<tab>XX<tab>supports-gun-rights-majority <tab>YY<tab>neutral-
majority<tab>ZZ<tab>no_majority<tab>AA
<month3><year3>
supports-gun-control-majority<tab>XX<tab>supports-gun-rights-majority <tab>YY<tab>neutral-
majority<tab>ZZ<tab>no_majority<tab>AA
```

```
targeted_ supports-gun-control-majority<tab>XX<tab>targeted_ supports-gun-rights-majority<tab>YY<tab>targeted_ neutral-
majority<tab>ZZ
```

What to submit:

- Data Sampling Python file: `username1_username2_si618_hw4_data_sampling_<topic>.py`
- Mturk form HTML file: `username1_username2_si618_hw4_crowd_form_design_<topic>.html`
- Mturk result CSV file: `username1_username2_si618_hw4_batch_result__<topic>.csv`
- Computation results text file: `username1_username2_si618_hw4_computations_<topic>.txt`
- Computation code Python file: `username1_username2_si618_hw4_compute_<topic>.py`

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