**SI 618 Fall 2020 HW 7 – Mturk**

In this homework, we will utilize Amazon Mechanical Turk (Mturk) to collect labels on content gathered through Lab5. Our goal here is to measure civility in news discussions through one specific dimension—personal attacks. Given the widespread

Data: si618\_lab7\_tweets\_uniqname1\_uniqname2.csv

**Step 1 – Mturk registration**

Each team needs to create an account on MTurk as a requester (<https://requester.mturk.com/>). Note that you need to provide a payment method for the account. You will get reimbursed.

**Step 2 – Set up Mturk task**

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

**1.** Enter 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 html template we shared for the description of the task

Keywords: Use “news discussions, tweets, classification”

Reward per payment: $0.1, number of workers per hit: 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 100 HITS (100 tweets \* one tweet per HIT).

Total amount of fee:$0.035 per worker \* 3 workers per Hit \* 100 HITS = $10.5 + $8.715 (Mturk overhead costs) is about **$20**.

Please do not exceed this amount, you will only be reimbursed for that amount.

**2.** Design Layout & Mturk input file preparation

Check out the template code, hw7\_html\_template.html, for insights. Pay attention to instructions provided in comment form (<!-- Instructions: … -->). I have provided the instructions in hw7\_html\_template.html. If you read the instructions provided, your questions and answer options should be based on that.  For example, the first question says there should be at least one answer (meaning there can be multiple) and the following questions should only have one answer as an option. What kind of buttons will you select? If you look at hw7\_example\_input\_data.csv, You will notice your si618\_lab7\_tweets\_uniqname1\_uniqname2.csv can be used as an input csv file for Mturk without changing anything if your arguments are well-specified.

**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 you generated in Lab7. 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. You can also look at hw7\_example\_batch\_result.csv as an example. Once you have the result, change the result csv file name to si618\_hw7\_batch\_result\_uniqname1\_uniqname2.csv.

Create si618\_ hw7\_compute\_uniqname1\_uniqname2.py. Compute which fraction of tweets were deemed related to trump. Then compute which fraction of tweets were deemed related to Biden.

Take a look at hw7\_example\_compute.txt as a reference.

**Step 5 – Reading Summary**

Read <https://hbr.org/2017/01/the-humans-working-behind-the-ai-curtain>

Please give a paragraph (5-6 sentences) summary and reflection of what the article is about and what kind of knowledge and insight it provided you about this industry. You and your partner should consolidate both of your paragraphs to one document. Please put your name and uniqname before your paragraph.

Format should look like:

Kevin Lee (kjunwonl)

<enter paragraph>

Ceren Budak (cbudak)

<enter paragraph>

**What to submit:**

Submit a zip file named si618\_hw7\_uniqname1\_uniqname2.zip containing

Mturk result CSV file: si618\_hw7\_batch\_result\_uniqname1\_uniqname2.csv

Compute result txt file: si618\_hw7\_compute\_uniqname1\_uniqname2.txt

Python file: si618\_hw7\_compute\_uniqname1\_uniqname2.py

PDF file: si618\_hw7\_summary\_reflection\_uniqname1\_uniqname2

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