Ethics in Natural Language Processing 2024 Homework 2

TECHNISCHE UNIVERSITÄT DARMSTADT

Due until Wednesday, 19.06.2024 at 11:59pm

Submission Guidelines for Homework

- This homework is worth 20 points
- Submit your code in a single .ipynb notebook. Submit subjective answers using the given latex template.
- Extra credit shall be given to well-structured submissions.
- In case of questions or remarks, please contact:
  - Aishik Mandal, aishik.mandal@tu-darmstadt.de

Before you start, make sure you read the Submission Guidelines instructions associated with this homework for important setup and submission information.

1 k-anonymity, l-diversity and t-closeness (8 points)

Consider the following non-anonymised table, with four personal attributes (Name, Age, Gender, City) and a sensitive attribute (Crime). The (id) attribute is not part of the data and is there to help you refer to the table rows, if needed.

(id)	Name	Age	Gender	City	Crime
1	Dwight	32	Male	Mainz	Murder
2	Meredith	25	Female	Munich	Murder
3	Pam	30	Female	Leverkusen	Robbery
4	Jim	34	Male	Mainz	Assault
5	Kelly	29	Female	Munich	Robbery
6	Angela	31	Female	Leverkusen	Murder
7	Michael	38	Male	Mainz	Parking
8	Ryan	18	Male	Frankfurt	Murder
9	Holly	33	Female	Leverkusen	Parking
10	Erin	20	Female	Munich	Speeding
11	Mose	19	Male	Frankfurt	Robbery
12	Kevin	35	Male	Dortmund	Assault
13	Oscar	39	Male	Dortmund	Speeding
14	Stanley	19	Male	Frankfurt	Speeding
15	Andy	34	Male	Dortmund	Murder

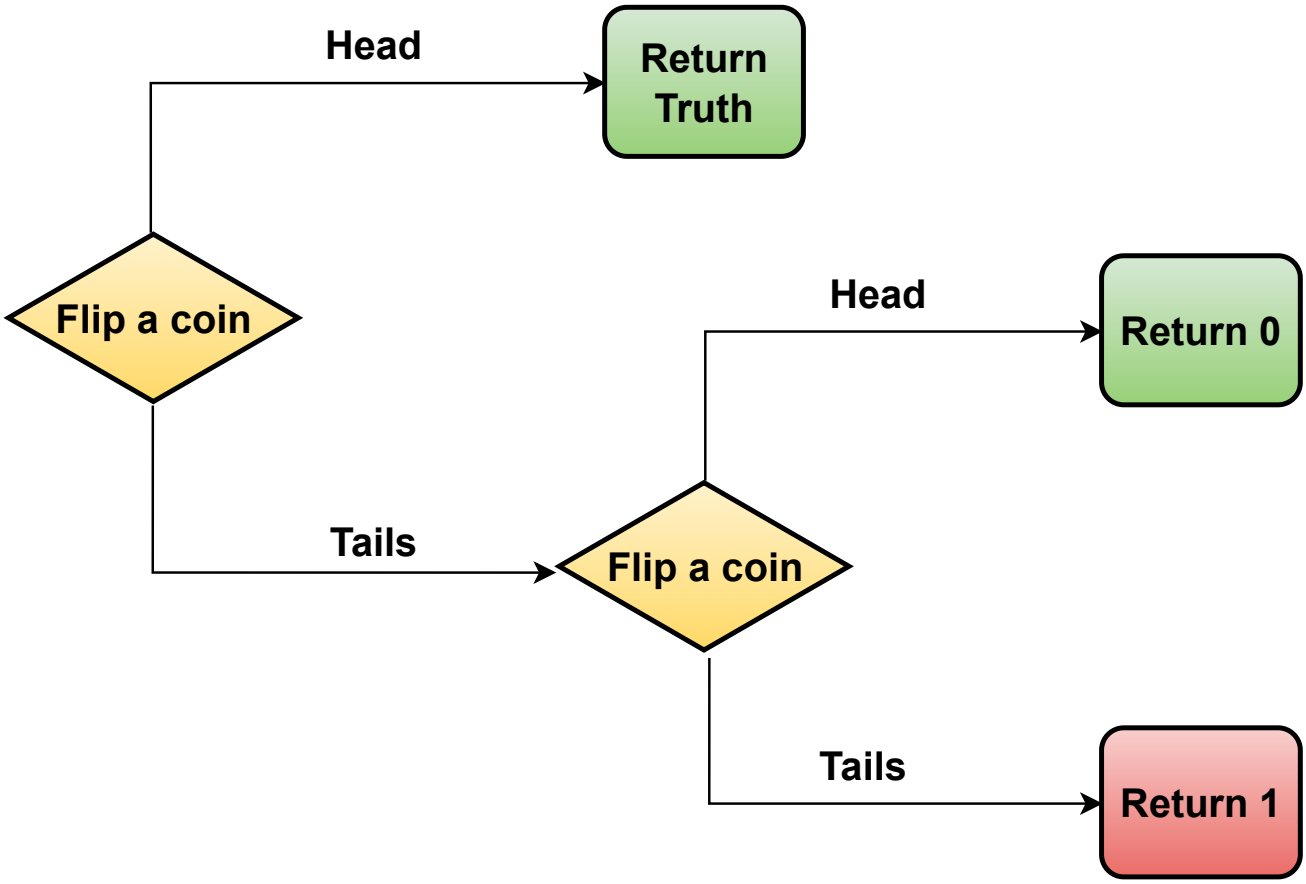
Table 1: A table showing crimes committed in different cities.

Your tasks are as follows:

- Anonymise the personal attribute values such that the dataset has 3-anonymity, that is, k-anonymity where k=3 (4 points). Do this by applying generalization and suppression to the data.
- Calculate the l-diversity of the modified table (2 points).
- Pick one equivalence class (a group of rows with the same non-sensitive attribute values). List the table rows that belong to this equivalence class. Calculate the t-closeness for this equivalence class (2 points).

2 Differential Privacy (6 points)

Consider the coin toss example you studied in class. We assumed that the coin was fair, i.e. returning heads and tails with a probability of 0.5. But it doesn't need to be. Now, consider an unfair one that gives heads with probability p and tails with probability $(1-p)$. Write a function that takes the actual (non-anonymized) distribution X and the "heads" probability p as inputs and returns a noisy distribution Y. Run this function for an actual distribution of $X = (0.8, 0.2)$, where 0.8 is the probability of $X = 0$ and 0.2 is the probability of $X = 1$ and $p \in 0, 0.2, 0.5, 0.8, 1$. For each of the p values, compare the true distribution X to the noisy, anonymized distribution Y. What can you say about the privacy and utility of the anonymized distribution depending on the value of p?



3 Text Anonymisation (6 points)

In the texts below, suppress the following personally identifiable information (PII) by replacing it with placeholders: [NAME] for names and usernames, [URL] for urls and mails addresses, [DATE] for dates, [PHONE] for phone numbers, [LOC] for locations. Even after the text is sanitized this way, there are still some elements in the language of these texts that could "leak" or allow us to guess something about the authors' location, demographics, etc. Can you list some of these elements (words, phrases)?

- Text 1: Had such a blast hanging out with @Sarah and @Bob in #Frankfurt today! We explored the city, went to the #FrankfurtZoo, and had an awesome time. It's always a good time catching up with old friends and making new memories together. Frankfurt has so much character – definitely need to come back soon! Thanks for the awesome day, @Sarah and @Bob. Let's do it again sometime! #Friends #Frankfurt #FB20 #TUDA

- Text 2:
  Date: 10.04.2024
  To: m.lakoff@cambridge.org, sonja@stanford.edu, jjjschmidt@uni-mannheim.de, max99999@gmail.com
  Title: RE: RE: RE: website
  Dear website team,
  here is a kick-off mail for the website. Format – a github page which lists some papers, between https://github.com/uclanlp/awesome-fairness-papers (too little info) and https://github.com/Hannibal046/Awesome-LLM (too much for now). For starters, we'd pick 50 papers from the list we already made https://docs.google.com/spreadsheets/d/1234567. Mike has kindly agreed to lead the work on this. Julia has agreed to develop the site. Max and Sonja have agreed to help with selecting papers and giving feedback. We need to have it running by 21.04 to link it in the paper. With that, I pass it to Mike.
  I guess the first step is setting up the repo and adding all people in this thread as Maintainers. My Github account is hello@johndoe.tu-darmstadt.de. If there are any questions, please let me know, if something urgent pops up, give me a call at +49 123 45678910.
  Best,