CS4.407: Online Privacy

Assignment- 2 (20 Marks) Deadline: Oct 22, 2021 23:59

Instructions:

- Jupyter Notebooks / Google Colab are strongly recommended for doing the coding part of the assignment and showing any analysis, graphs and code.
- Please cite any sources that you might use in the process.
- Please write your own writeup and code. All writeup and code will be tested for plagiarism and if found, institute policy will be followed.
- *Do this assignment individually.*

Q1 [10 marks])

Identity Resolution in Online Social Media is a major problem because it allows the aggregation of data on users across multiple platforms. Given in the .zip file is another ground truth dataset of 324 users with their usernames across three different social networks, in the following format:

Name	Twitter Username	Facebook Username	Instagram Username

Use the dataset to answer the following questions:

a. Use two commonly used distance metrics to study whether their usernames across different pairs of platforms depict some similarity. What is the best metric to use for
    a. Twitter-Facebook Identity Resolution
    b. Facebook-Instagram Identity Resolution
    c. Twitter-Instagram Identity Resolution
b. Plot the distribution for both the metrics across all three pairs of platforms from (a). What can you infer?

Q2 [10 marks])

Please look at all the cookies stored in your machine. Analyze them for various parameters like which site has the greatest number of cookies, what is the duration of cookies, the naming convention around cookies etc.

Please prepare some graphs / tables with the data and provide some insights and takeaways from the analysis.

Submission instructions:

<roll_number>/

|_ q1/

|_<roll_number>_q1.ipynb

|_ <roll_number>_q1.pdf

|_ q2/

|_<roll_number>_q2.ipynb

|_ <roll_number>_q2.pdf

Zip the <roll_number>/ folder and submit it on the portal as <roll_number>.zip

1. The qx folder should contain a jupyter notebook <roll_number>_qx.ipynb having all your code for Qx. Download this same .ipynb as a pdf and submit the same as <roll_number>_q2.pdf. The pdf file should include all the plots that have been plotted.
2. Any additional scripts etc that you may use should be included in the relevant folder.
3. Any insights/inferences asked to be shared must be done through markdown cells.
4. Include relevant markdown for both questions within the notebook itself that serves as an explanation of your approach. **We will not award any marks for the questions if you do not explain your approach via markdown.**