Security and Privacy

Lab Assignment Worksheet

João Vilela - Manuel E. Correia

2024/2025

— Assignment #3: Anonymization of a Dataset with Utility and Risk Analysis

---

Due date: June 1st, 23:59

Grading: Assignment #3 is worth **2 points**

TO BE DONE IN **GROUPS OF THREE (MANDATORY)**

The goal of this assignment is to perform the anonymization of a larger dataset. You should resort to an example project/dataset provided by the authors of the tool here: https://arx.deidentifier.org/downloads/.

After opening the project, you should:

1. Attributes are by default all set as quasi-identifying. Analyze the attributes and classify them into: Identifying, QIDs, Sensitive or Insensitive. Justify your choice.

   This should take into consideration the values for distinction and separation.

2. You should now characterize/analyze the privacy risks of the dataset in original form.

   Re-identification risk of the original Dataset (ARX: Analyze risk > Attacker models)

3. You should apply **at least two privacy models** to your dataset and conduct an analysis of the performance of each privacy model applied to the dataset.

   For the analysis you should consider:

   - The re-identification risk of the anonymized dataset vs the original dataset (ARX: Analyze risk > Attacker models)
   - The utility level, measured through appropriate utility metrics (ARX: Analyze utility)

4. You should assess the effect of varying parameters of the privacy models (e.g. suppression limit, coding model, attribute weights, etc) on the level of privacy and utility achieved. For that, you should create plots of privacy risk and utility levels (y axis) with varying privacy model parameters (x axis). You need to provide a description of the plots and appropriately define the axis and metrics, for a standalone understanding of the plots.

** Submit in moodle (1) your ARX project with anonymization models applied and a (2) document explaining (2-a) the classification of attributes, and (2-b) the above plots and corresponding analysis. **

Evaluation Criteria

- Classification of attributes and justification [30%]
- Privacy models: utility and risk assessment [70%]
  - Analysis of the utility and privacy (attacker models and re-identification risk) levels (before and after anonymization)
  - Analysis of privacy models' results according to varying privacy model's parameters