

Trabalho 1 - k -anonimato

Javam Machado

Setembro 2019

1 Objetivo:

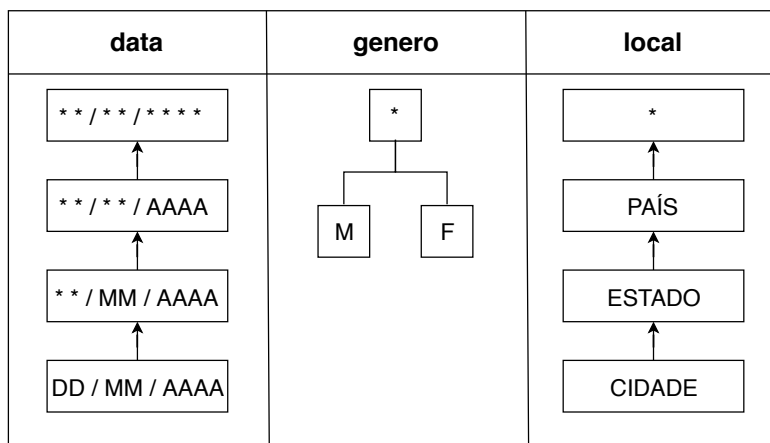
- Implementar um algoritmo que anonimize um conjunto de dados de tal forma que seja atendido o modelo de privacidade k -anonimato.

2 Especificação:

- Carregue o conjunto de dados “salario.csv”. Os atributos são:
 - *Identificadores Explícitos*: id;
 - *Semi-identificadores*: país, estado, cidade, gênero e data;
 - *Sensíveis*: salário.
- O valor de k deve variar no conjunto $k = \{2, 4, 8, 16\}$.
- Para cada valor de k , o conjunto de dados deve ser anonimizado de tal forma que seja atendido o modelo k -anonimato.
- A métrica de utilidade a ser adotada deve ser a **Precisão**.

3 Requisitos

- Linguagem: Python
- Meio de entrega: criar um repositório chamado “disciplina_privacidade_2019” no Github e compartilhar com os seguintes e-mails: {andre.luis, iago.chaves, israel.vidal, javam.machado}@lsbd.ufc.br. **Todos os trabalhos da disciplina serão entregues através desse repositório.**
- Criar uma pasta “ k -anonimato” no repositório “disciplina_privacidade_2019”.
- Equipes de até 2 pessoas.
- Somente um repositório deve ser criado por equipe.
- O arquivo “README.md” no repositório deve conter os componentes da equipe.
- Utilize as seguintes hierarquias de generalização:



4 Avaliação

- O algoritmo implementado irá anonimizar um novo conjunto de dados “salario_2.csv”, que terá os mesmo atributos do arquivo “salario.csv”, mas **não necessariamente o mesmo número de linhas**.
- Na avaliação será considerada a **corretude** e a **precisão** do algoritmo implementado.

5 Entrega

- 17 de Setembro de 2019. *Commit* do código até 13:59.