

Residência em Segurança da Informação

Trabalho – Técnicas de Anonimização – Generalização

Prof. Javam Machado

1 Objetivo:

- Implementar a técnica de generalização em um conjunto de dados.

2 Especificação:

- Carregue os conjuntos de dados “covid.csv” e “regioes.csv” que lhe serão fornecidos. O primeiro dataset já estará pseudoanonimizado, enquanto o segundo deverá ser utilizado para generalização de um atributo. Você deve implementar um programa para realizar a generalização do dataset de COVID:
 - Defina uma árvore de generalização com 2 níveis para o atributo “Município” considerando as mesorregiões do Estados do Ceará contidas no dataset “regioes.csv”, e uma outra árvore para data de nascimento com 3 níveis: dado original (aaaa/mm/dd), data sem dia (aaaa/mm), data sem dia e mês (aaaa). Considere o nível 0, o nível dos dados originais.
 - A partir do dataset “covid.csv” e tendo como base as árvores de generalização enviadas, o programa deve realizar:
 1. A generalização de *domínio completo* dos valores do atributo “Município” para o nível 1 e gravar um arquivo “covid_anon01.csv” anonimizado;
 2. A generalização de *domínio completo* dos valores do atributo “Nascimento” para o nível 1 e gravar um arquivo “covid_anon02.csv” anonimizado.
 3. A generalização de sub-árvore dos valores do atributo “Nascimento” para o nível 2 e gravar um arquivo “covid_anon03.csv” anonimizado. Para tal, aplicar a anonimização apenas nos registros cujo indivíduo tenha nascido na década de 1950 (anos 1950 a 1959).
 4. A generalização de *domínio completo* dos valores dos atributos “Município” (nível 1) e “Nascimento” (nível 2) e gravar um arquivo “covid_anon04.csv” anonimizado.

- O programa deve mostrar na tela as duas hierarquias de generalização definidas a partir do dataset.
- Programe uma interface para seleção e visualização dos datasets gerados para cada saída, onde o usuário possa visualizar o dataset anonimizado selecionado.

3 Requisitos

- Linguagem: C/C++ ou Python
- Meio de entrega: Empacotar os artefatos de software (programa, arquivos de entrada e de saída) para um arquivo “<SeuNome>.zip” e fazer o *upload* pelo classroom.

4 Deadline

- *Upload* pelo classroom até as 23:59 do dia 28/08 (domingo).