MDCC - Privacidade de Dados - 2022

Privacidade Diferencial – Mecanismo de Laplace

Javam Machado

1 Objetivo

- Implementar o mecanismo de Laplace, segundo a privacidade diferencial. Realizar consultas sobre uma base de dados através do mecanismo de Laplace implementado e fornecer os resultados de forma correta e privada.

2 Especificação

- Carregue o conjunto de dados "Covid-CE.csv". Calcule a idade dos indivíduos representados em cada registro a partir da data de nascimento. Considere apenas datas de nascimento referentes aos séculos 20 e 21.
- Realizar um conjunto de três consultas como especificadas abaixo. As consultas deverão respeitar os seguintes $\varepsilon = \{0.1, 0.5, 1.0, 10\}$.
- Calcule a sensibilidade global para cada consulta. Lembre que a sensibilidade global independe do conjunto de dados. (Dica: faça premissas sobre os dados trabalhados a fim de calcular a sensibilidade quando não for possível.)
- Consultas a serem realizadas sobre a totalidade dos dados:
  1. Q_1: Média da idade dos indivíduos representados no dataset;
  2. Q_2: Número de exames positivos (atributo *resultadoFinalExame*);
  3. Q_3: Total de exames realizados por município (atributo *municipioCaso*;
- Para as consultas Q_1 e Q_2, mostre um gráfico com o resultado da consulta para cada ε comparado com o valor original.
- Para a consulta Q_3, defina bins para apresentação dos resultados, macro região por exemplo, e apresente um gráfico para cada ε. Cada gráfico vai mostrar a frequência original e a frequência perturbada referente a cada bin.

3 Requisitos

- Linguagens: C++ ou Python
- Meio de entrega: criar uma pasta zipada chamado "Trab_Priv_Diff_Laplace_<nome>" contendo código e dataset de entrada e arquivos "csv" com os resultados das consultas. Fazer o upload na plataforma Classroom da disciplina.
- Preparar uma Demo para explicar, mostrar o seu programa e os resultados durante a aula de entrega. Escreva um Readme.txt descrevendo o projeto.
- O trabalho deverá ser entregue até as 14h da **segunda-feira, 27/06/2022** e explicado durante as aulas da 2a feira 27/06 e 4a feira 29/06.

4 Avaliação

Na avaliação serão considerados os seguintes indicadores:

- **Corretude** do programa;
- **Precisão** pela comparação do dataset original com o dataset anonimizado;
- Clareza na **explicação** do programa durante a Demo;
- **Pontualidade** e **documentação/qualidade** do código-fonte.