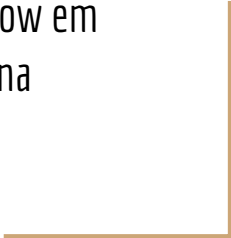




Projeto Orientado em Computação 2: Pitch Parcial

Relações entre Fairness, Privacidade e
Quantitative Information Flow em
Aprendizado de Máquina



Nome: Artur Gaspar da Silva
Orientador: Mário Sérgio Alvim

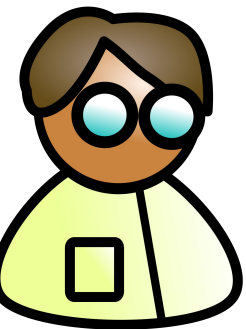
Objetivos do POC2

- Estudar mais a fundo conceitos de Privacidade Diferencial e Fairness
- Explorar conexões com Quantitative Information Flow
- Explorar impacto de métodos de obfuscagem em fairness
- Explorar como budget de privacidade (Local Differential Privacy, LDP) pode ser dividido entre variáveis de importância diferente

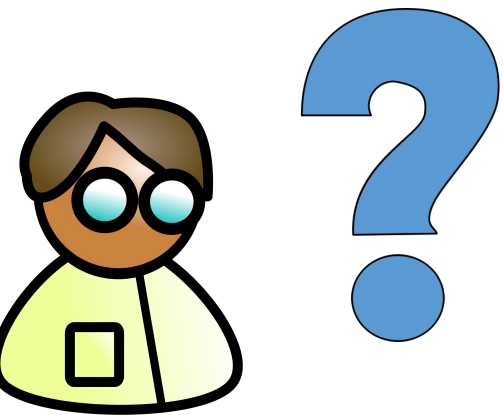
Etapas concluídas até agora e etapas seguintes

- Concluído até agora:
 - Exploração de conceitos chave de LDP e QIF
 - Tentativas iniciais de modelar (delta, epsilon)-LDP com QIF (especialmente o delta)
 - Discussão inicial de métodos de ofuscagem relacionados a privacidade em fairness
- Etapas seguintes no POC2:
 - Tentar outras abordagens de demonstração para introduzir a modelagem do parâmetro delta em (delta,epsilon)-LDP
 - Formalizar a relação de métodos de ofuscagem de privacidade com fairness, com foco no efeito da etapa de reversão do ruído
 - Explorar como um budget de privacidade pode ser melhor distribuído entre variáveis com níveis de importância diferentes

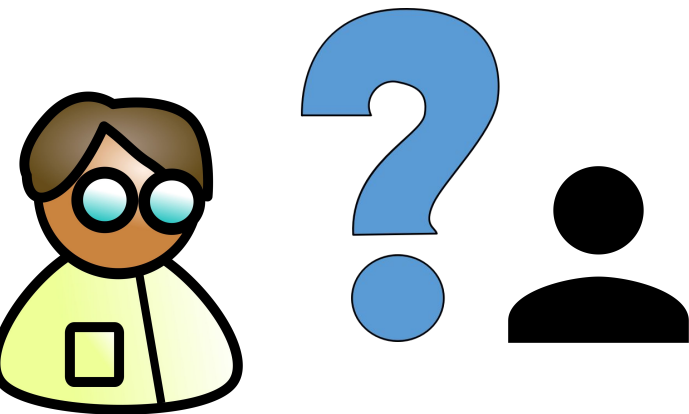
Local Differential Privacy



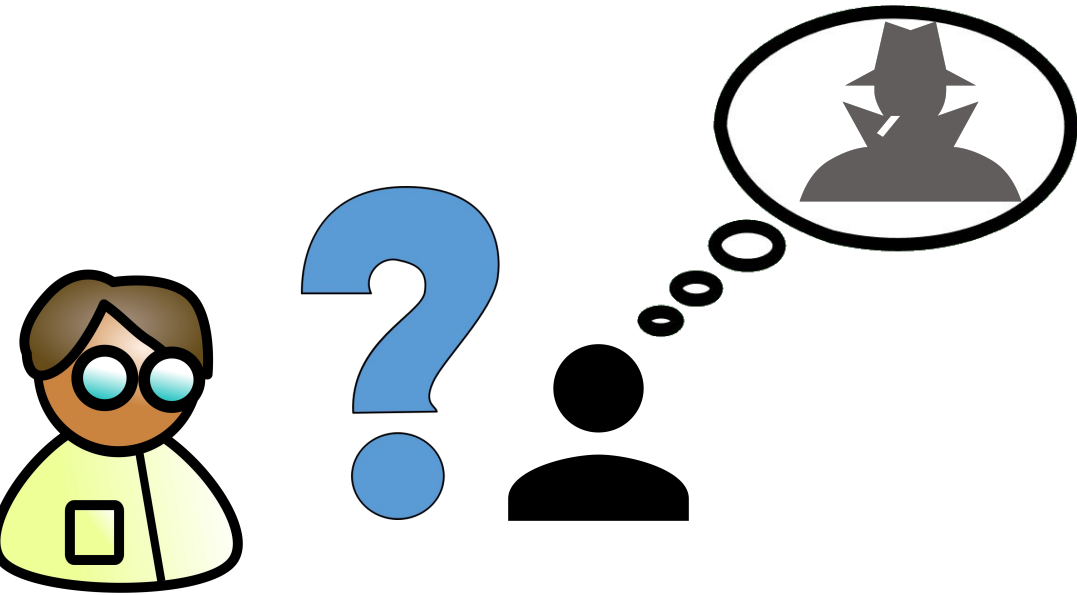
Local Differential Privacy



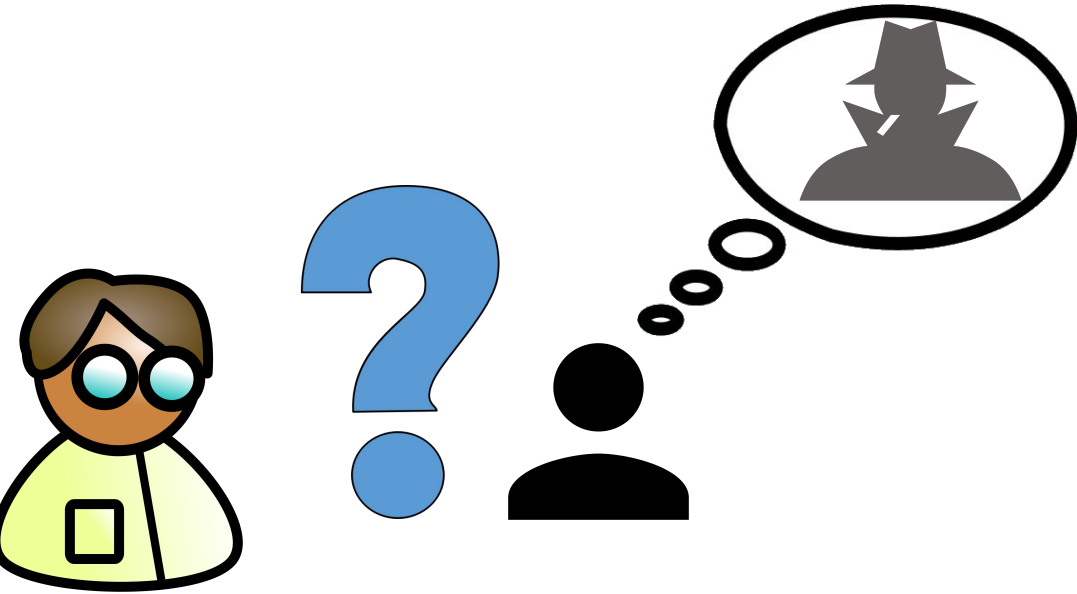
Local Differential Privacy



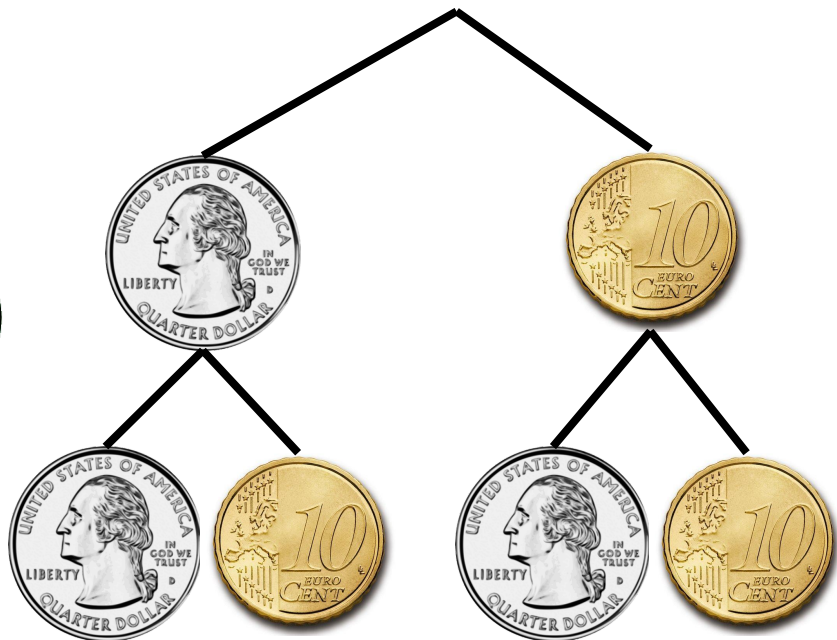
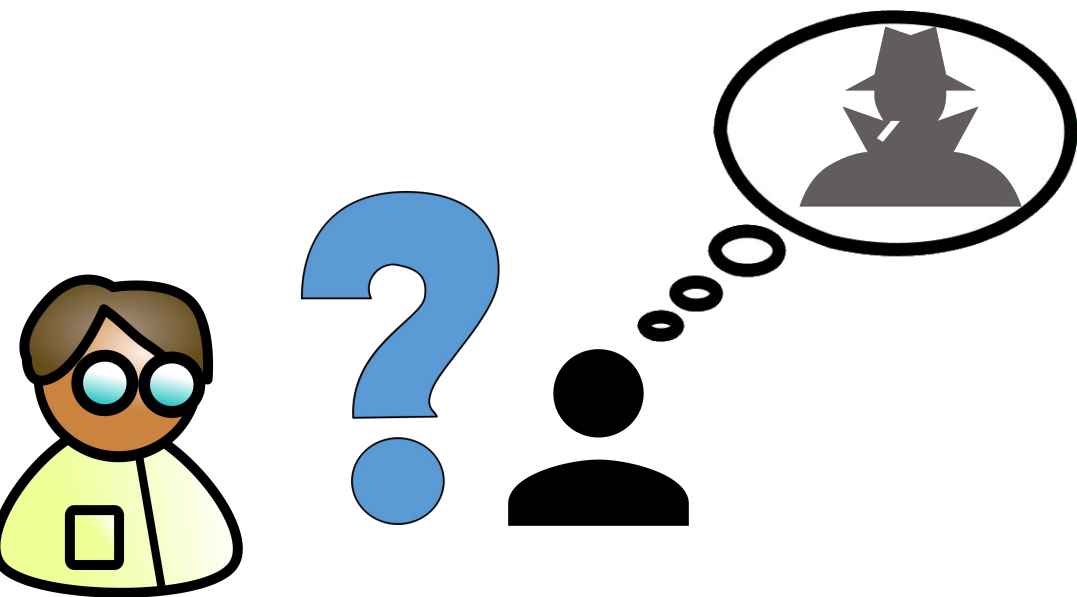
Local Differential Privacy



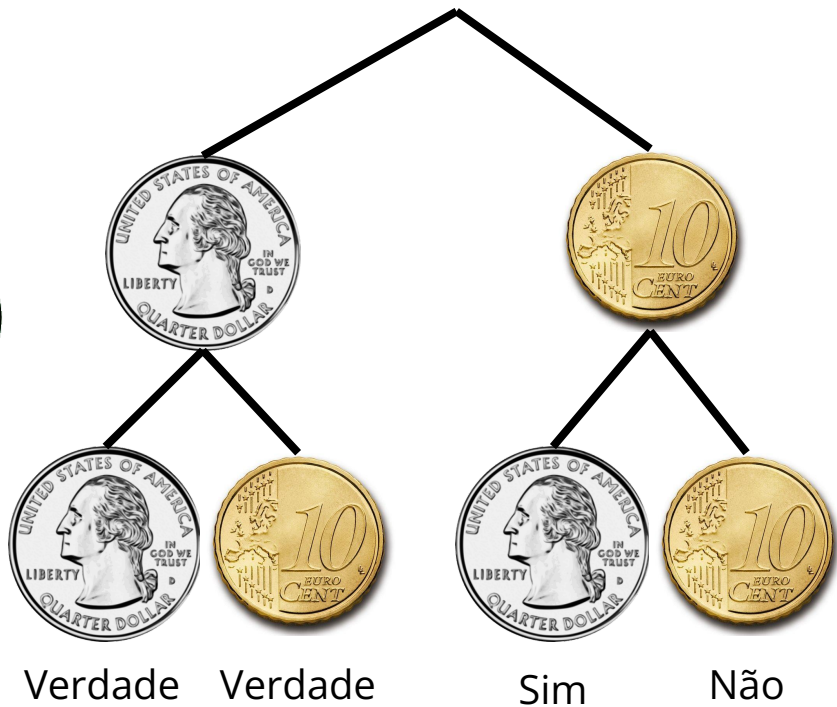
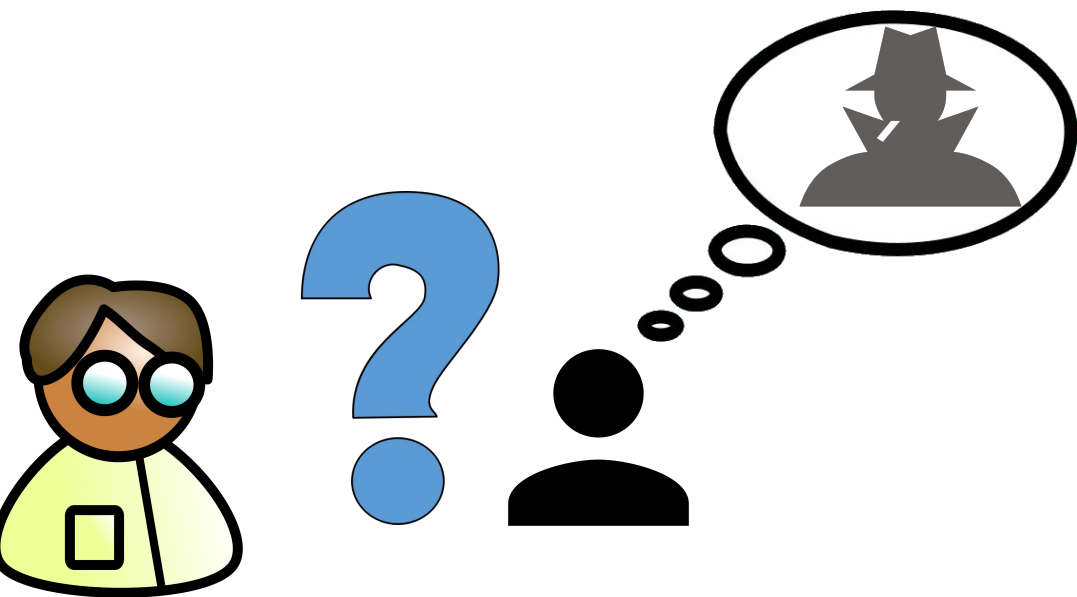
Local Differential Privacy



Local Differential Privacy



Local Differential Privacy



Fairness: (Conditional) Statistical Disparity



Fairness: Equal Opportunity Difference

$$P(\hat{Y}=1 | A=1, Y=1) - P(\hat{Y}=1 | A=0, Y=1)$$

Parâmetros epsilon e delta

- Epsilon representa o quanto de incerteza do adversário toleramos, no pior caso, entre o valor real e qualquer outro valor da informação secreta.
- Delta representa, de certa forma, qual a margem tolerável de erro para o valor de epsilon real estar errado
- Um artigo recente de 2024 modelou epsilon usando Max-QIF, que considera cenários com probabilidade não nula
- A ideia é considerar $\delta_{\text{Max-QIF}}$, cenários com probabilidade pelo menos delta, pois cenários dentro da margem seriam automaticamente aceitos

0 efeito em fairness sem reversão de ruído

- Artigos recentes consideram como obfuscar variáveis sensíveis pode automaticamente melhorar métricas de fairness
- No entanto, esses artigos consideram que a etapa de reversão de ruído em LDP não é realizada, o que não condiz com a realidade
- Problemas de privacidade surgem se LDP for aplicada apenas à variável sensível, que acaba não sendo protegida
- Os resultados desses artigos **não condizem com a realidade** se a reversão de ruído é realizada

Distribuição do budget de privacidade

- Simplesmente aplicar ruído a variáveis sensíveis não é o bastante dum ponto de vista de privacidade
- Aplicar ruído a tudo pode ser overkill
- Possivelmente é mais eficiente aplicar ruído às variáveis sensíveis e a variáveis correlacionadas
- Verificar a viabilidade de fazer isso