Defesa/Análise de projeto

Informações teóricas

Para este projeto usámos os modelos de privacidade:

- k-Anonymity
- l-Diversity
- t-Closeness

E a partir daí, foram avaliados os seus resultados, seja em relação a riscos ou na utilidade dos dados anonimizados.

Os atributos apresentados no dataset são categorizados em dois tipos:

- Quasi-Identifier (QID), i.e., não identificam alguém isoladamente, mas quando combinados entre si ou com fontes externas, podem reidentificar indivíduos (Sex, Age, Marital-status, education, native-country, workclass, occupation);
- Sensitive, i.e., contêm informação intrinsecamente delicada cujo acesso não autorizado pode resultar em danos ou discriminação (race, salary-class).

Porque é que esses atributos são classificados dessa forma?

- Sex → QID → Possui um baixo poder de distinção direta, visto que só existem dois valores possíveis, mas em conjunto com outros atributos aumenta a unicidade dos registos;
- Age → QID → Alta cardinalidade, ou seja, existem muitas faixas etárias e maior poder de distinção, mas a reidentificação só é normalmente conseguida combinando outros QIDs;
- Race → Sensitive → Intrinsecamente sensível, que não permite identificação direta. Exige certos modelos para evitar inferências discriminatórias;
- Marital-status → QID → contribui para a distinção de perfis e não é sensível em si, mas pode reidentificar em conjunto com outros QIDs;
- Education → QID → Cardinalidade moderada (níveis de escolaridade), reforçando a distinção quando combinado com outros QIDs;

- Native-country → QID → Alta cardinalidade (vários países) e maior poder de distinção, mas vulnerável a reidentificação quando combinado com outros QIDs;
- Workclass → QID → Várias classes laborais, não é sensível e contribui para reidentificação de registos em conjunto;
- Occupation → QID → Alta cardinalidade (várias profissões), reforçando a distinção quando combinado com outros QIDs;
- Salary-class → Sensitive → Revela informação delicada (faixa de rendimento), sendo sensível por poder causar discriminação. Deve se usar certos modelos de privacidade para evitar inferências discriminatórias.

O que é o modelo atacante do promotor?

Neste modelo, o invasor tem como alvo um indivíduo específico e presume-se que ele já saiba que dados sobre ele estão contidos no conjunto de dados. Tendo acesso a recursos legais e informações complementares, tem alta probabilidade de ligar registos do dataset e indivíduos em mais de metade dos casos (68%). O risco máximo inicial deste modelo é de 100%, e a sua taxa de sucesso na reidentificação de registos é de 55%.

O que é o modelo atacante jornalístico?

Neste modelo, o invasor tem como alvo um indivíduo específico, mas não se espera que ele tenha conhecimento prévio sobre a filiação. Por exemplo, um jornalista da área de investigação possui técnicas de correlação de dados, alcançando níveis de risco semelhantes aos do promotor. Assim, destaca-se a vulnerabilidade das informações pessoais.

O que é o modelo atacante de marketing?

Aqui, o invasor não tem como um alvo um indivíduo específico mas sim a reidentificação de um grande número de indivíduos. Um ataque só pode ser considerado bem-sucedido se uma fração grande dos registos puder ser reidentificada, possuindo uma taxa de sucesso de 55%.

O que são limiares de risco?
São limiares pelos quais são delimitados valores, sendo que os

que superarem esse limite, são considerados perigosos e aumentam o risco de reidentificação.

O que é um modelo de privacidade?

Conjunto de princípios, regras ou técnicas, usado para proteger dados pessoais contra reidentificação ou exposição indevida durante a partilha ou análise dos mesmos.

Como funciona o modelo k-Anonymity?

Técnica de anonimização de dados usada para proteger a privacidade de indivíduos numa base de dados. Envolve generalização de dados e data masking, i.e., a substituição de dados (PII) por dados falsos (psudónimos) mas realistas. Uma desvantagem é a sua vulnerabilidade a ataques homogéneos e de background knowledge.

O que são ataques homogéneos?

Ataque de privacidade que pode ser aplicado a dados que são anonimizados usando uma simples técnica de generalização se os dados partilharem os mesmos valores dos QIDs e tiverem os mesmos valores nos atributos sensíveis. Se os grupos não contiverem valores diferentes, um atacante pode revelar informação sensível encontrando o grupo ao qual o indivíduo pertence. Nesses casos, os dados tornam-se vulneráveis, e é possível proteger dados destes ataques usando o modelo l-Diversity.

Como assim, background knowledge?

Informações que o atacante possui, que estão para além da informação protegida. Essa informação pode inferir dados sensíveis, mesmo após a sua anonimização.

Como funciona o modelo l-Diversity?

É uma extensão do modelo anterior, desenvolvida para reduzir a granulação/granularidade de dados numa base de dados, i.e., para reduzir o nível de detalhe ou especificidade com que os dados foram armazenados. Este modelo assegura que nenhuma informação de um indivíduo pode ser identificada pelo menos por L outros indivíduos da base de dados, baseando-se num atributo sensível e protegendo atributos sensíveis como gerais. A sua desvantagem é a sua vulnerabilidade a ataques skewness e de similaridade.

O que são ataques de skewness ou skewing attacks?

Os atacantes procuram falsificar dados de análise, encorajando o indivíduo a tomar decisões erradas de acordo com a informação manipulada.

O que são ataques de similaridade?

Quando valores de um atributo sensível numa classe equivalente são distintos mas semanticamente/estruturalmente semelhantes, um atacante pode aprender informação importante sobre indivíduos e grupos.

Como funciona o modelo t-Closeness?

É um refinamento/upgrade de l-Diversity, pois trata valores de um atributo de forma distinta, tendo em conta a distribuição dos mesmos. Uma classe de equivalência possui t-Closeness se a distância entre a distribuição de um atributo sensível nessa classe e a distribuição do atributo na tabela completa não excede t.

Neste contexto, o que é uma classe de equivalência?

Grupo de registos que partilham os mesmos valores para certos atributos QID, i.e., é o agrupamento de dados que são tratados da mesma forma em termos de proteção de privacidade. Este método é usado para classificar dados em diferentes categorias baseado no seu nível de sensibilidade ou importância. Agrupando registos em classes de equivalência, torna-se mais difícil conectar informações sensíveis de um indivíduo aos seus QIDs. Classes de equivalência são usadas em vários modelos de privacidade. Por exemplo, em k-Anonymity, cada classe de equivalência deve conter pelo menos k registos, tornando estatisticamente improvável de reidentificar um indivíduo dentro dessa classe. Outros modelos, como l-Diversity e t-Closeness, consideram a distribuição de atributos sensíveis dentro de cada classe para mitigar riscos de privacidade. Imaginemos um dataset com registos sobre as idades, géneros e salários de indivíduos. Uma classe de equivalência deve ser formada por todos os registos com a mesma idade e género. O salário é um atributo sensível.

O que é uma análise de risco?

São avaliados os modelos de atacante Promotor, Jornalista, Marketing. Dentro delas: Records at Risk/registos em risco, Highest Risk/Maior registo, Sucess rate/taxa de sucesso.

O que é Records at Risk?

Indica a fração de registos cuja probabilidade de reidentificação excede um limiar. É calculado estimando, para cada registo i, a sua probabilidade de ser reidentificado (pi). Usando um limiar t, conta-se quantas pi's são iguais ou superiores a esse limiar, dividindo esse valor pelo número total de registos. Após isso é usado outro limiar, derivado do anterior, para estabelecer o limite entre algo perigoso e o contrário.

O que é Highest Risk?

Captura o pior caso de risco dentro do dataset, i.e., num conjunto de probabilidades de reidentificação excedentes do limiar de Records at Risk, este será o seu valor máximo.

O que é o Success Rate?

Mede a efetividade prática do adversário a reidentificar registos ou atributos. Usa um limiar derivado.

O sucess Rate de Reidentificação global é a proporção de reidentificação feitas pelo atacante que acertam o verdadeiro ID ou valor sensível.

O sucess rate em Top-k é a proporção de vezes que, num cenário onde o atacante produz um ranking de k candidatos possíveis para cada registo, o verdadeiro alvo está dentro dos k primeiros resultados. Ou seja, é um risco teórico num desempenho simulado de ataque, sendo o indicador mais prático de vulnerabilidade.

O que é uma análise de utilidade?

Avalia o quanto a informação original é preservada após a aplicação de técnicas de privacidade, tendo como objetivo a garantia de que os dados continuam úteis para análises estatísticas, mesmo após a proteção de privacidade.

Como é avaliada a qualidade ao nível de Atributo?

Avalia o impacto da anonimização em atributos específicos.

- Missings/Supressões: nr de valores ausentes ou suprimidos após a anonimização
- Intensidade de generalização: lvl de generalização aplicado aos dados, i.e., mede o quão largo foi recodificar os atributos → quanto maior a

intensidade, maior a perda de detalhe

- Granularidade: mede lvl de detalhe mantido nos dados após a generalização
- Entropia (normalizada): avalia diversidade de valores num atributo → menor entropia, maior perda de informação
- Erro quadrático ao nível do atributo: erro médio quadrático entre os valores originais e anonimizados de um atributo numérico

Como é avaliada a qualidade ao nível do dataset?

Avalia o impacto da anonimização no conjunto como um todo.

- Discernibilidade: mede a capacidade de distinguir registos após a anonimização, i.e., pede a penalidade infligida às classes de equivalência → valores mais altos, menor utilidade
- tamanho médio de classes de equivalência: média de nrs de registos em cada grupo indistiguível. Proporção dessa média em relação ao tamanho ideal k → tamanho maior, maior perda de especificidade
- erro quadrático ao nível do registo: erro quadrático entre registos originais e anonimizados individualmente
- erro quadrático específico de agregação: avalia erro introduzido por técnicas de agregação específicas aplicadas durante a anonimização
- erro quadrático no geral serve para medir distorção entre dados originais e transformados → menor o erro, mais os dados preservam propriedades originais

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2$$

sendo x_i o valor original e $\hat{x}_i$ o valor transformado.

Análise de Resultados

Começando com k-Anonymity + l-Diversity, foram escolhidos 3 parâmetros diferentes para cada, numa análise dos mesmos. Após isso, foi selecionado um parâmetro para cada, e foram variados outros parâmetros.

Parâmetros (k,l) da primeira análise:

- (3,2)
- (5,4)
- (10,2)

Dados adicionais: medida de utilidade Loss, limite de supressão 100%, pesos de atributos 0.5

Pode-se concluir que, no contexto de privacidade do dataset e mitigação do risco de reidentificação de indivíduos a partir do mesmo, é completamente anulado na configuração (5,4), sendo o mais seguro na parte de anonimização. A partir da análise de utilidade percebe-se que essa configuração não é útil, pois apaga todos os dados do dataset. Assim, a configuração que nos dá mais equilíbrio risco x utilidade é (10,2). Em relação à qualidade dos atributos, mais uma vez a configuração (10,2) é a melhor, visto que, mesmo que haja perda de informação e utilidade reduzida nos atributos mais generalizados, dá uma melhor cobertura.

Parâmetros da segunda análise:

- Limites de supressão 10, 50, 100%
- Medidas de utilidade Discernibilidade, tamanho médio de classes de Equivalência, Loss
- Peso de atributos padrão (0.5), age = 0.8, sex = 0.2, age = 0.2, sex = 0.8

Usando a configuração (5,2)

LIMITES DE SUPRESSÃO

Em nenhum dos casos há registos em risco. O pior caso de reidentificação é de 20%, i.e., a menor classe de equivalência com k = 5 registos. A taxa de sucesso é de 0.31%.

Em relação à utilidade, tem bons resultados.

MEDIDAS DE UTILIDADE

Em nenhum dos casos há registos em risco. O pior caso de reidentificação é de 20%, i.e., a menor classe de equivalência com k = 5 registos. A medida com maior taxa de sucesso é avg class size. A mais baixa é discernibilidade.

Em relação à utilidade:

- A discernibilidade é a melhor escolha se for necessária máxima fidelidade aos valores originais
- O tamanho médio de classes é a melhor escolha se o essencial é ter sempre k registos por grupo
- Loss é a melhor escolha se quisermos um compromisso geral entre precisão local e global

PESO DE ATRIBUTOS

Em nenhum dos casos há registos em risco. O pior caso de reidentificação é de 20%, i.e., a menor classe de equivalência com k = 5 registos. A medida com maior taxa de sucesso é age = 0.8, sex = 0.2. As restantes são iguais. Há mais chance de acerto quando o algoritmo se preocupa mais em preservar informação etária. Esse mesmo peso não traz benefícios na utilidade, sendo que o cenário padrão é o melhor e mais equilibrado.

Começando com k-Anonymity + t-Closeness, foram escolhidos 3 parâmetros diferentes para cada, numa análise dos mesmos. Após isso, foi selecionado um parâmetro para cada, e foram variados outros parâmetros.

Parâmetros (k,t) da primeira análise:

- (3,0.15)
- (5,0.2)
- (10,0.15)

Dados adicionais: medida de utilidade Loss, limite de supressão 100%, pesos de atributos 0.5

Com as configurações (5,0.2) e (10, 0.15), não há registos em risco de serem reidentificados. O maior risco atinge o seu valor ótimo, ou mínimo, em (10,0.15), e máximo em (3,0.15), diminuindo consoante k aumenta. A taxa de sucesso atinge, mais uma vez, o seu ponto ótimo em (10,0.15), e o seu ponto máximo em (5,0.2). Em relação à utilidade, a configuração (5,0.2) possui uma fraca utilidade, visto que os dados estão mais distorcidos. A configuração (3,0.15) oferece melhor utilidade mas pior privacidade. A melhor configuração é, assim, a (10,0.15).

Em relação à qualidade de atributos, os valores são iguais entre (3, 0.15) e (10,0.15).

Alerta, há um erro de numeração nesta parte do relatório entregue, com a config (5,0.15) em vez de (3,0.15)!!!

Parâmetros da segunda análise:

- Limites de supressão 10, 50, 100%
- Medidas de utilidade Discernibilidade, tamanho médio de classes de Equivalência, Loss
- Peso de atributos padrão (0.5), age = 0.8, sex = 0.2, age = 0.2, sex = 0.8

Usando a configuração (5,0.15)

LIMITES DE SUPRESSÃO

Em nenhum dos casos há registos em risco. O pior caso de reidentificação é de 0.058%. A taxa de sucesso é de 0.0.17%.

Em relação à utilidade, tem bons resultados.

MEDIDAS DE UTILIDADE

Em nenhum dos casos há registos em risco. A medida com risco máximo mais baixo é a discernibilidade. Mais alto é avg class size, nao muito distante de Loss. A medida com maior taxa de sucesso é avg class size. A mais baixa é discernibilidade.

Em relação à utilidade:

- A discernibilidade apresenta baixa distorção e boa preservação de dados, sendo ideal para alta utilidade com algum proteção
- O tamanho médio de classes garante o anonimato mais forte com classes muito grandes
- Loss aplica generalização intensa, o que compromete utilidade individual dos dados, mas mantém a consistência das análises agregadas

O melhor caso seria a discernibilidade.

PESO DE ATRIBUTOS

Em nenhum dos casos há registos em risco. O pior caso de reidentificação é de 20%, i.e., a menor classe de equivalência com k = 5 registos. A medida com maior taxa de sucesso é age = 0.8, sex = 0.2. As restantes são iguais. Há mais chance de acerto quando o algoritmo se preocupa mais em preservar informação etária. Esse mesmo peso não traz benefícios na utilidade, sendo que o cenário padrão é o melhor e mais equilibrado.

Alerta, há um erro nesta parte do relatorio entregue: diz que a terceira config é igual à segunda, mas é igual à primeira!!