Anonimização de Dados para Inteligência Artificial usando o Algoritmo da Tropa dos Gorilas

Ivo A. Pimenta[1], Ramon S. Araújo [1], Renann L. Rodrigues[1], Matheus M. Silveira[2], Rafael L. Gomes[1]

[1]Universidade Estadual do Ceará (UECE)

{aguiar.pimenta,ramon.araujo,renann.rodrigues}@aluno.uece.br

rafa.lopes@uece.br

[2]Uber Technologies Inc.

matheusmsil@uber.com

Abstract. *The collection of data from the environment and individuals through the Internet of Things (IoT) is a reality, where such data is utilized by innovative solutions based on Artificial Intelligence (AI). However, particularly in the healthcare domain, user data must comply with privacy laws. Thus, there is a challenge in understanding the utility of data used in AI solutions while adhering to legal requirements, for instance, by anonymizing the data. Traditional anonymization methods compromise the effectiveness of AI models, reducing their performance. In this context, this article proposes the* $GOK - Privacy$ *algorithm, which combines a metaheuristic inspired by gorilla behavior with clustering techniques, enabling privacy preservation without sacrificing the performance of analytical models. Experiments conducted using real healthcare data demonstrate the effectiveness of the proposed solution in real-world scenarios.*

Resumo. *A coleta de dados do ambiente e das pessoas através da Internet das Coisas (IoT) é uma realidade, onde esses dados são usados por soluções inovadoras baseadas em Inteligência Artificial (IA). Contudo, especialmente na área de saúde, esses dados de usuários precisam atender às definições das Leis de Privacidade. Desta forma, há o desafio de entender a utilidade dos dados usados em soluções de IA enquanto cumpre os aspectos legais, por exemplo, anonimizando os dados. Métodos tradicionais de anonimização comprometem a eficácia dos modelos de IA, reduzindo a eficácia dos mesmos. Dentro deste contexto, este artigo propõe o algoritmo* $GOK - Privacy$*, que combina uma meta-heurística inspirada no comportamento de gorilas com técnicas de agrupamento, permitindo alcançar a preservação de privacidade sem sacrificar o desempenho dos modelos analíticos. Os experimentos realizados usando dados reais de saúde mostram a eficácia da proposta em cenários reais.*

1. Introdução

A Internet das Coisas (IoT) é um paradigma de comunicação no qual um conjunto de dispositivos heterogêneos se conecta com a Internet e comunica-se entre si [Portela et al. 2024, Gomes et al. 2014b]. Esses dispositivos de comunicação desempenham um papel essencial na sociedade moderna. As aplicações usuais de sistemas IoT

estão relacionadas ao monitoramento de usuários, equipamentos tecnológicos e do ambiente físico [Silva et al. 2022, Portela et al. 2023]. Um exemplo de contexto é o aumento dos ataques cibernéticos, como Negação de Serviço Distribuída (DDoS) e Softwares Maliciosos (Malwares), também ressalta a necessidade de medidas robustas para proteger usuários e dispositivos, sendo que há uma tendência da aplicação de soluções baseadas em IA [Gomes et al. 2014a, Silveira et al. 2023c, Souza et al. 2024].

Neste cenário, o uso de técnicas de IA junto com dados coletados via IoT traz preocupações em relação à privacidade e segurança dos dados, especialmente considerando que os dados coletados podem ser sensíveis e expor usuários a riscos [Seh et al. 2020, Ferreira et al. 2024]. Adicionalmente, leis de privacidade, como o GDPR na Europa e a LGPD no Brasil, demandam que informações sensíveis sejam protegidas contra acesso não autorizado [Silveira et al. 2023b, Silva et al. 2023].

Uma abordagem existente para lidar com esses aspectos de privacidade é o uso de técnicas de anonimização, pois permitem que dados sensíveis sejam utilizados de maneira não identificável, preservando a privacidade dos usuários enquanto possibilitam análises baseadas em Aprendizado de Máquina (ML) [El Mestari et al. 2024, Yuan and Wu 2022, Portela et al. 2024]. As técnicas de anonimização, embora cruciais para atender às regulamentações de privacidade, apresentam desafios, visto que essas técnicas tradicionais podem comprometer a qualidade dos dados, impactando negativamente a eficácia de modelos de ML[Pimenta et al. 2024, Silveira et al. 2023a].

Dentro deste contexto, este artigo propõe o algoritmo GOK-Privacy, que combina uma meta-heurística inspirada no comportamento de gorilas com técnicas de agrupamento com lógica fuzzy, permitindo alcançar a preservação de privacidade dos dados enquanto mantém a eficácia dos modelos de ML para classificar os dados de forma correta. A técnica de comportamento de gorilas, que compõe a meta-heurística base do GOK-Privacy, permite a preservação das propriedades matemáticas críticas dos dados, oferecendo um equilíbrio entre privacidade e usabilidade [Abdollahzadeh et al. 2021]. Desta forma, a proposta visa demonstrar que é possível preservar a privacidade dos dados anonimizados sem comprometer sua utilidade para a construção de soluções baseadas em ML. A principal contribuição da proposta é avançar o estado da arte em relação ao suporte à privacidade e inovação tecnológica com IA através da anonimização de dados.

Os experimentos foram realizados com um conjunto de dados reais disponíveis para a comunidade científica [Hussain et al. 2021], o qual possui dados de ambientes de IoT na área da saúde dentro do contexto de detecção de ataques cibernéticos. Os resultados mostram a eficácia da solução, diante de outras técnicas existentes na literatura ($K - Mondrian$ e $Top - Down$), em atender aspectos de privacidade e acurácia, alcançando uma acurácia média de 95% (contra 80% de outras técnicas de anonimização).

O restante deste trabalho está organizado da seguinte forma. Na Seção 2 serão discutidos os trabalhos relacionados. Na Seção 3 será detalhada a proposta deste artigo (GOK-Privacy), enquanto na Seção 4 serão apresentados os resultados dos experimentos realizados. Por fim, a Seção 5 trará as considerações finais e trabalhos futuros.

2. Trabalhos Relacionados

Slijepčević et al. [Slijepčević et al. 2021] exploram o impacto de algoritmos de anonimização baseados no critério k-anonymity em modelos de aprendizado de máquina.

O estudo investiga como técnicas de generalização e supressão afetam o desempenho de classificadores em diferentes conjuntos de dados reais. Os autores utilizam quatro algoritmos de anonimização (Mondrian, OLA, Top-Down Greedy e k-NN Clustering) e avaliam seu impacto em classificadores populares. Os resultados mostram que a força da restrição de k-anonymity geralmente degrada o desempenho de classificação, mas a magnitude dessa degradação varia conforme o conjunto de dados e o método de anonimização utilizado. Entre os algoritmos analisados, o estudo enfatiza a necessidade de considerar o impacto da anonimização em aplicações de aprendizado de máquina, destacando a importância da escolha do método de anonimização para minimizar a perda de informação e viabilizar análises robustas.

Ni et al. [Ni et al. 2022] apresentam um framework para avaliar técnicas de anonimização de dados no contexto de Big Data e IoT, abordando questões críticas como privacidade, utilidade dos dados e perda de informações. Os autores exploram cinco algoritmos principais de anonimização: ofuscação, permutação, k-anonimato, l-diversidade e privacidade diferencial. O trabalho introduz novas métricas para medir o impacto da anonimização, como penalidade de certeza global (GCP), entropia condicional e utilidade dos dados. Além disso, propõe um mecanismo de otimização baseado em árvores de decisão para equilibrar a privacidade e a utilidade dos dados. Os experimentos foram realizados em conjuntos de dados públicos, avaliando o desempenho de cada técnica em termos de tempo de execução, uso de memória e eficácia na preservação da privacidade. Os resultados destacam que, embora técnicas como k-anonimato e l-diversidade sejam amplamente utilizadas, apresentam compromissos significativos entre privacidade e utilidade, dependendo dos parâmetros selecionados e do contexto do uso.

Choudhury et al. [Choudhury et al. 2020] descrevem uma abordagem inovadora de anonimização sintática para aprendizado federado (FL) com foco na preservação da privacidade em contextos de dados sensíveis, como o setor de saúde. Diferentemente de métodos baseados em privacidade diferencial, a abordagem proposta busca maximizar a utilidade dos dados e o desempenho do modelo, atendendo simultaneamente a regulamentações como o GDPR e HIPAA. O método emprega anonimização baseada no modelo k-anonymity, aplicada tanto em atributos relacionais quanto transacionais, garantindo privacidade interpretável e defensável. Os autores avaliam a eficácia do método em dois casos reais: predição de reações adversas a medicamentos e taxas de mortalidade hospitalar, utilizando grandes conjuntos de dados de saúde. Os resultados mostram que a abordagem alcança melhor preservação da utilidade dos dados e desempenho do modelo em comparação com métodos baseados em privacidade diferencial, destacando sua viabilidade em cenários de aprendizado federado com dados distribuídos.

Kacha et al. [Kacha et al. 2021] propõem o KAB, uma abordagem inovadora para k-anonimização baseada no algoritmo de buraco negro (Black Hole Algorithm - BHA). A proposta busca superar limitações de técnicas tradicionais ao minimizar a perda de informação, melhorando a qualidade dos dados anonimizados. O BHA é usado para encontrar soluções ótimas para o problema de k-anonimidade, representando dados anonimizados como ”estrelas”e a melhor solução como o ”buraco negro”. Os experimentos mostram que a abordagem reduz significativamente a perda de informação em comparação com métodos de k-anonimização baseados em clustering e hierarquias.

Langari et al. [Langari et al. 2020] definem o KFCFA, uma abordagem híbrida

que combina clustering fuzzy com o algoritmo Firefly para garantir a privacidade em redes sociais. A técnica busca proteger dados contra ataques de identidade, atributos e similaridade, enquanto minimiza a perda de informação. A metodologia utiliza um algoritmo fuzzy c-means modificado para criar clusters balanceados, seguido pela otimização com o algoritmo Firefly. Experimentos com bancos de dados de redes sociais (Facebook, Twitter, YouTube e Google+) demonstram a eficiência do KFCFA em satisfazer k-anonimidade, l-diversidade e t-closeness com menor distorção dos dados.

Kumar et al. [Kumar et al. 2024] apresentam uma nova abordagem que combina k-anonymity com técnicas de *Machine Learning* (ML) para proteger dados sensíveis de saúde enquanto mantém sua utilidade para análise preditiva. Os autores propõem um framework em duas etapas que primeiro aplica k-anonymity aos dados e depois utiliza técnicas de aprendizado de máquina adaptadas para trabalhar com dados anonimizados. O método foi testado em um conjunto de dados de registros médicos eletrônicos com mais de 100.000 registros. Os resultados demonstram que é possível preservar a privacidade dos pacientes sem comprometer significativamente o desempenho dos modelos de ML.

Com base na revisão da literatura, essas propostas existentes não atendem por completo aos aspectos de privacidade e adequação ao processo de treinamento de modelos de IA, considerando técnicas de anonimização como habilitador para cumprir esses pontos. Sendo assim, a proposta avança o estado da arte em relação ao suporte à privacidade e inovação tecnológica com IA através da anonimização de dados.

3. Proposta

O algoritmo proposto combina as capacidades exploratórias do Algoritmo de Otimização dos Gorilas (*Gorilla Optimization Algorithm* - GOA) com o particionamento flexível do agrupamento fuzzy para alcançar a anonimização de dados com preservação de propriedades matemáticas [Abdollahzadeh et al. 2021], as quais são usadas pelos modelos de ML para classificação dos dados. Esta abordagem enfrenta o desafio de equilibrar a utilidade dos dados e a privacidade através de uma estrutura de otimização que considera tanto identificadores quase-identificadores numéricos quanto categóricos, preservando os padrões subjacentes para aplicações de aprendizado de máquina [El Mestari et al. 2024]. Uma visão geral da estrutura da proposta é apresentada na Figura 1.

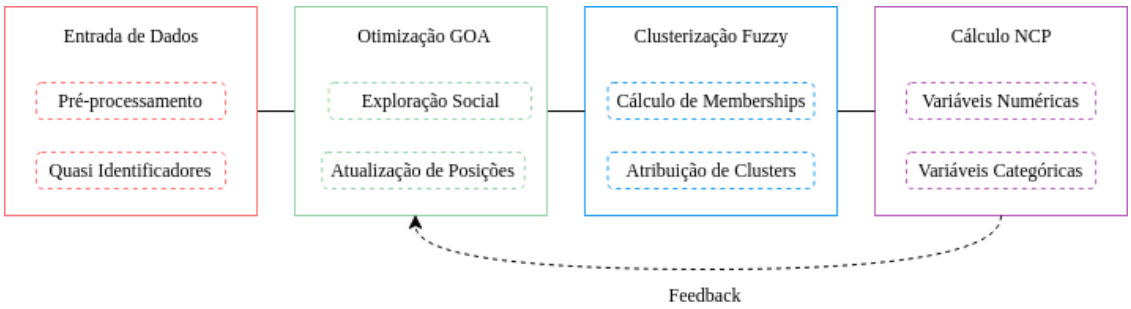

Figura 1. Visão Geral da Proposta.

Como pode ser observado na Figura 1, a proposta possui quatro etapas principais que realizam as seguintes funções:

1. Processamento de Entrada de Dados: Lida com a preparação inicial dos quase-identificadores e normalização dos dados;

2. Otimização GOA: Implementa o processo de otimização inspirado no comportamento social dos gorilas;
3. Agrupamento Fuzzy: Gerencia a atribuição flexível de registros a grupos de privacidade;
4. Cálculo Penalidade de Certeza Normalizada (*Normalized Certainty Penalty* - NCP): Avalia a qualidade da anonimização considerando a preservação de padrões.
5. Feedback: Após o cálculo do NCP , é retornado um valor de feedback que indica o quão bem os clusters gerados preservam os padrões originais dos dados. Esse feedback é utilizado como critério de parada. Se o valor de feedback não puder ser melhorado de uma iteração para a outra, significa que o algoritmo convergiu para uma solução ótima, e a execução é finalizada.

3.1. Entrada de Dados

A etapa de entrada de dados é essencial para o processo de anonimização e consiste em duas funções principais: pré-processamento e gerenciamento de quase-identificadores. Quase-identificadores são atributos em um conjunto de dados que, quando combinados, podem potencialmente identificar um indivíduo de maneira única, mesmo que nenhum desses atributos individualmente seja capaz de identificá-lo.Por exemplo, em um conjunto de dados de pacientes de um hospital, atributos como idade, código postal e data de admissão podem ser considerados quase-identificadores.

Na função de pré-processamento, removemos dados faltantes e tratamos as inconsistências, caso possuam algumas. O algoritmo gerencia quase-identificadores numéricos e categóricos de forma distinta, aplicando generalizações específicas conforme o tipo de dado. Essas estratégias garantem a preservação do significado semântico dos valores originais ao longo do processo.

3.2. Otimização dos Gorilas

A meta-heurística de otimização dos gorilas (GOA) foi desenvolvida para simular o comportamento social dos gorilas na natureza, aplicando suas estratégias de liderança e interação para resolver problemas de otimização complexos [Abdollahzadeh et al. 2021]. Essa abordagem se baseia nas características como a hierarquia de liderança e os movimentos estratégicos dos gorilas.

3.2.1. Abordagem do GOA

A população inicial de N gorilas é distribuída aleatoriamente dentro do espaço de busca $\mathcal{D}$:

$$\mathbf{x}_i = \mathbf{x}_{\text{min}} + \text{rand}(0,1) \cdot (\mathbf{x}_{\text{max}} - \mathbf{x}_{\text{min}}), \quad i = 1, 2, \ldots, N, \tag{1}$$

onde $\mathbf{x}_{\text{min}}$ e $\mathbf{x}_{\text{max}}$ são os limites inferiores e superiores do espaço de busca, e $\text{rand}(0,1)$ gera um número aleatório em $[0,1]$.

Cada solução candidata $\mathbf{x}_i$ é avaliada usando a função objetivo $f(\mathbf{x}_i)$. O gorila com a melhor aptidão é designado como o líder, $\mathbf{x}_{\text{leader}}$. As posições dos gorilas são atualizadas com base em estratégias de exploração e exploração, regidas por comportamentos probabilísticos.

Na fase de exploração, os gorilas se movem para regiões desconhecidas do espaço de busca. A atualização da posição é definida como:

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + \alpha \cdot (\mathbf{x}_{\text{rand}} - \mathbf{x}_i^{(t)}), \tag{2}$$

onde α é um fator de escala, e $\mathbf{x}_{\text{rand}}$ é uma posição selecionada aleatoriamente no espaço de busca.

Na fase de exploração local, os gorilas seguem o líder para refinar suas posições. A atualização da posição é dada por:

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + \beta \cdot (\mathbf{x}_{\text{leader}} - \mathbf{x}_i^{(t)}), \tag{3}$$

onde β é um fator de escala que controla a influência do líder.

O algoritmo itera até que um critério de convergência seja atingido, como: Um número máximo de iterações $T_{\max}$; e uma mudança mínima no valor da função objetivo entre iterações.

3.2.2. Algoritmo GOA

Durante a exploração, a GOA executa três mecanismos principais de movimentação dos gorilas. O primeiro deles é a migração para uma posição desconhecida, onde os gorilas se movem para áreas que ainda não conhecem, ampliando a exploração do espaço de busca. O Algoritmo GOA é apresentado no Algoritmo 1.

Algorithm 1 Algoritmo de Otimização dos Gorilas (GOA)

Inicialize a população de N gorilas aleatoriamente dentro de $\mathcal{D}$.
Avalie a aptidão de cada gorila.
Identifique o líder $\mathbf{x}_{\text{leader}}$.
for $t = 1$ até $T_{\max}$ **do**
    for cada gorila i **do**
        Atualize a posição $\mathbf{x}_i$ usando a fase de exploração ou exploração local.
        Avalie a nova aptidão $f(\mathbf{x}_i)$.
    end for
    Atualize $\mathbf{x}_{\text{leader}}$ se uma solução melhor for encontrada.
end for
Retorne $\mathbf{x}_{\text{leader}}$ como a melhor solução.

O algoritmo começa com a inicialização de uma população de ”gorilas”(soluções candidatas) de forma aleatória. Cada solução candidata representa um conjunto de centros de clusters para as variáveis a serem anonimizadas. O grupo de gorilas é governado por um *silverback* que tem a capacidade de realizar todas as ações. Além disso, os gorilas machos mais jovens, conhecidos como *blackbacks*, agem como defensores de backup para o grupo [Abdollahzadeh et al. 2021].

Em seguida, o algoritmo entra em um laço , onde a cada iteração ele calcula a adequação de cada solução candidata. Quando o algoritmo atualiza a posição de cada

gorila (solução candidata), ele utiliza uma combinação de forças sociais e forças de exploração. A exploração social é calculada como a diferença entre a posição do gorila atual e a posição do melhor gorila. Isso simula o comportamento de seguir o líder e aprender com ele, similar ao que acontece na natureza com os gorilas mais jovens em relação ao *silverback*. Durante o laço, o algoritmo mantém controle da melhor solução encontrada até o momento, atualizando as posições, que podem ser consideradas análogas ao papel do gorila *silverback* (líder) na hierarquia social. A adequação é calculada com base no NCP. O objetivo é minimizar essa função de adequação, encontrando a melhor configuração de clusters que atenda aos requisitos de anonimização.

3.3. Agrupamento com Lógica Fuzzy

O componente fuzzy do GOK-Privacy implementa uma versão similar à do Fuzzy C-Means [Chiu and Tsai 2007], para fazer o cálculo de *Membership* adaptada para lidar com requisitos de k-anonimidade. A Equação 4 apresenta a definição do grau de pertinência μ_{ij} de um registro i ao cluster j, onde d_{ij} é a distância euclidiana entre o registro i e o centroide j, m é o fator de fuzzificação e c é o número de clusters.

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} (\frac{d_{ij}}{d_{ik}})^{\frac{2}{m-1}}} \tag{4}$$

Assim, baseado nessa equação, verificamos a pertinência fuzzy de cada cluster ao invés de uma associação rígida tradicional. O grau de pertinência $u_{i,j}$ como mostrado na equação, determina quando cada registro *i* pertence ao cluster j, usando a distância euclidiana $d_{i,j}$ entre os registros e o centroide do cluster, ponderada pelo fator de fuzificação m, onde cada gorila tem um determinado grau de pertencimento a cada cluster, ao invés de ser simplesmente atribuído a um cluster único. Quanto menor a distância de um gorila a um determinado centro, maior será o seu grau de pertencimento a esse cluster.

A atribuição dos clusters é feita da seguinte forma , após calculado os graus de pertencimento o algoritmo atribui cada gorila a um cluster específico. Ao atribuir os gorilas aos clusters, o algoritmo garante que cada cluster tenha pelo menos um determinado número mínimo de gorilas (representado pelo parâmetro k). Isso evita a formação de clusters muito pequenos, o que poderia comprometer a qualidade da solução final.

3.4. Cálculo do NCP

He et al. [He et al. 2012] propuseram um algoritmo de anonimização k baseado em agrupamento. O algoritmo é um processo iterativo. Em cada rodada, o conjunto de dados é particionado em dois subconjuntos, G_1 e G_2, com base no NCP como medida de distância. Se o tamanho de um dos subconjuntos for menor que k, assumindo que $|G_1| < k$, os tamanhos dos dois subconjuntos são ajustados, emprestando $k - |G_1|$ tuplas de G_2 para garantir que G_1 tenha uma cardinalidade igual ou maior que k. O ajuste continua até que cada subconjunto contenha pelo menos k tuplas. A proposta deste trabalho aplica a mesma abordagem a fim de maximizar o processo de aprimoramento de permutações e manutenção de propriedades matemáticas, aplicando o NCP como critério de iteração das etapas da proposta.

Desta forma, a NCP mede a perda de informação em um processo de generalização de dados. Para um atributo numérico A, o NCP é calculado conforme apresentado na

Equação 5, onde A é o conjunto de valores do atributo numérico no cluster atual, D_A é o domínio completo do atributo considerando todos os dados, $max(A)$ é o valor máximo no cluster, $min(A)$ é o mínimo dentro do cluster, $max(D_A)$ é o valor máximo no domínio completo, $min(D_A)$ é mínimo dentro do domínio completo, σ_A é o desvio padrão dos valores no cluster e σ_{D_A} é o desvio padrão considerando todo o domínio.

$$NCP_{num}(A) = \begin{cases} \frac{max(A)-min(A)}{max(D_A)-min(D_A)} \times \frac{\sigma_A}{\sigma_{D_A}}, & \text{se } max(D_A) - min(D_A) \neq 0 \\ 0, & \text{caso contrário} \end{cases} \quad (5)$$

Para atributos categóricos, o NCP é definido como descrito na Equação 6, onde $|distinct(A)|$ é a cardinalidade do conjunto de valores únicos no cluster, $|distinct(D_A)|$ é a cardinalidade do conjunto de valores únicos no domínio completo, $|A|$ é o número total de registros no cluster e $|D_A|$ é o número total de registros no domínio completo.

$$NCP_{cat}(A) = \begin{cases} \frac{|distinct(A)|-1}{|distinct(D_A)|-1} \times \frac{|A|}{|D_A|}, & \text{se } |distinct(D_A)| > 1 \\ 0, & \text{caso contrário} \end{cases} \quad (6)$$

4. Experimentos

Nesta seção serão apresentados os experimentos deste trabalho, bem como discutidos os resultados alcançados a partir dos experimentos. Na Seção 4.1 são apresentadas as informações sobre o processo de avaliação, enquanto que na Seção 4.2 são discutidos os resultados e o comportamento de cada técnica analisada. É válido ressaltar que o código desenvolvido, bem como os dados utilizados nos experimentos, estão disponíveis no repositório do projeto[1] e com as instruções necessárias para reprodutibilidade.

4.1. Configuração dos Experimentos

Nesta seção iremos descrever a configuração dos experimentos e o processo de avaliação das técnicas analisadas. Uma visão geral da estrutura dos experimentos para avaliação é ilustrada na Figura 2, onde percebe-se o fluxo de execução das técnicas de anonimização. A seguir serão detalhadas as informações sobre o conjunto de dados usado, o processo de anonimização, treinamento dos Modelos de ML e a avaliação da eficiência das técnicas.

4.1.1. Conjunto de Dados

O conjunto de dados desenvolvido por [Hussain et al. 2021] foi criado para abordar a detecção de tráfego malicioso em ambientes de IoT na área da saúde. Ele foi gerado utilizando a ferramenta IoT-Flock, que permitiu a simulação de tráfego normal e malicioso envolvendo dispositivos de monitoramento ambiental (como temperatura e umidade) e dispositivos de monitoramento de pacientes (como oxímetros e monitores de glicemia). Esses dispositivos foram configurados para transmitir dados baseados em perfis

[1] https://github.com/IvoAP/GOK_SBRC_2025/

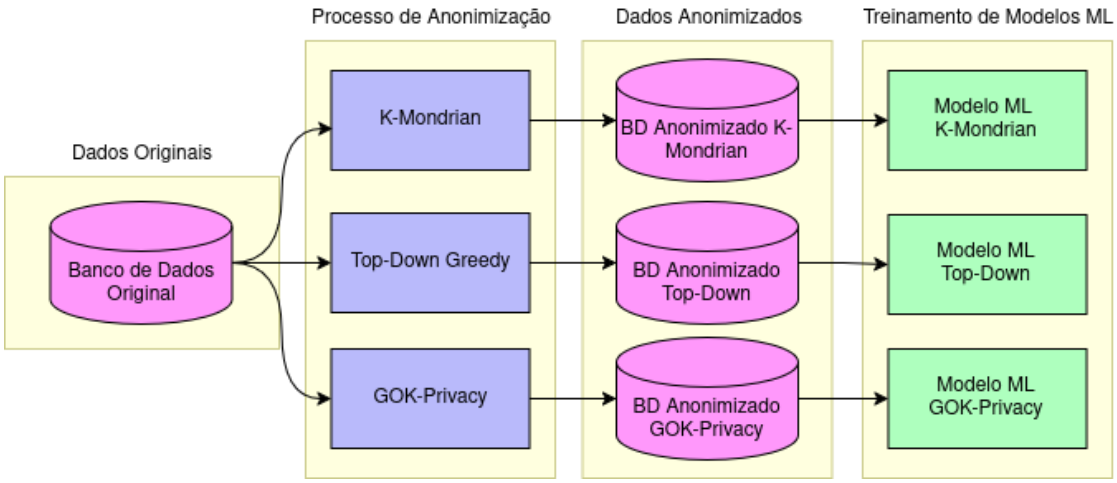

Figura 2. Organização dos experimentos realizados.

específicos de tipo e intervalo de tempo, refletindo condições reais de operação. Os dispositivos são classificados em duas categorias: Dispositivos de monitoramento do ambiente e dispositivos de monitoramento do paciente. O conjunto de dados contém tanto tráfego normal quanto malicioso, incluindo ataques como DDoS e *publish flood*, e foi projetado para treinar modelos de aprendizado de máquina voltados à segurança em IoT.

4.1.2. Processo de anonimização

Durante os experimentos, além da proposta deste trabalho ($GOK - Privacy$), foram analisadas as técnicas $K - Mondrian$ e $Top - DownGreddy$. Todas as técnicas aplicam uma abordagem de definição de um valor k como base de ações. Assim, os experimentos foram realizados testando diferentes valores de k (variando de 5 a 50, assim como outras referências da literatura, tais como [LeFevre et al. 2006, Xu et al. 2006]).

O $K - Mondrian$, proposto por LeFevre et al. [LeFevre et al. 2006], é um algoritmo de aproximação gulosa desenvolvido para alcançar a k-anonimidade. Ele opera dividindo o espaço de domínio em regiões multidimensionais, permitindo uma partição eficiente e estruturada dos dados. O $K - Mondrian$ utiliza uma maior generalização dos quasi-identificadores como ponto de partida e especializa recursivamente as partições aplicando divisões multidimensionais até que não sejam possíveis mais divisões. Cada iteração do algoritmo precisa escolher uma dimensão (atributo) para realizar a divisão. Em seguida, o valor de divisão é determinado utilizando a partição mediana, e o corte é realizado de acordo com esse valor de divisão.

Por outro lado, Xu et al. [Xu et al. 2006] propuseram um método de generalização local, denominado $Top - DownGreddy$, baseado em uma abordagem gulosa top-down para anonimização. O algoritmo recebe uma tabela contendo os dados como entrada e a particiona recursivamente em classes de equivalência cada vez mais locais. Para isso, utiliza a divisão binária em combinação com uma heurística para bi-particionar os dados em cada iteração. A métrica de NCP incorpora tanto a perda de informação causada pela anonimização quanto a importância dos atributos. Além disso, mede a incerteza dos valores dos atributos do registro generalizado, comparando-os com os valores originais e atribuindo-lhes pesos correspondentes.

4.1.3. Treinamento e Avaliação dos Modelos de ML

As técnicas de aprendizado de máquina utilizadas neste trabalho foram: Árvore de Decisão (*Decision Tree) - DT*, Floresta Aleatória (*Random Forest) - RF*, Perceptron Multicamadas (*Multi-Layer Perceptron - MLP)*, K-Nearest Neighbors (*KNN)* e Naive Bayes Gaussiano (NB). As técnicas de aprendizado de máquina utilizadas representam uma gama diversificada de abordagens que permitem avaliar os dados sob diferentes perspectivas analíticas. Cada método oferece características específicas que o tornam adequado para determinados tipos de problemas, proporcionando uma avaliação abrangente e robusta. Essa diversidade metodológica é fundamental na comparação do impacto das técnicas de anonimização e na identificação de quais técnicas são mais eficazes para o contexto de treinamento de modelos com dados anonimizados.

Para avaliar o desempenho dos modelos de classificação utilizados, foram consideradas as métricas de Acurácia e F1-Score, amplamente utilizadas no contexto de IA. A Acurácia mede a proporção de classificações corretas em relação ao total de instâncias. Por outro lado, o F1-Score combina de forma harmônica a Precisão e o Recall, oferecendo uma visão mais detalhada do equilíbrio entre os falsos positivos e falsos negativos. Essa métrica é especialmente valiosa em contextos onde o impacto de erros em classes específicas pode ser crítico. Assim, o uso combinado dessas métricas possibilita uma análise mais completa, equilibrando simplicidade e sensibilidade às características do problema.

4.2. Resultados

Nesta seção, apresentaremos de forma detalhada os resultados obtidos nos experimentos realizados com diferentes modelos de classificação em aprendizado de máquina. O objetivo principal é avaliar o desempenho de cada modelo em termos de acurácia e vermos como $GOK - Privacy$ se comporta comparado às outras técnicas de anonimização.

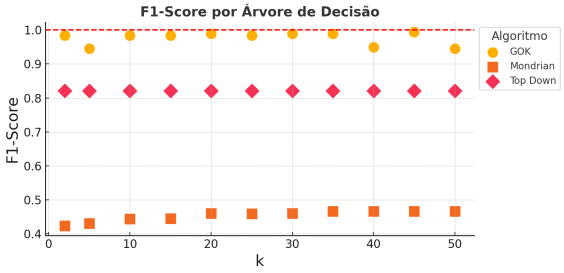

Figura 3. F1-Score do DT.

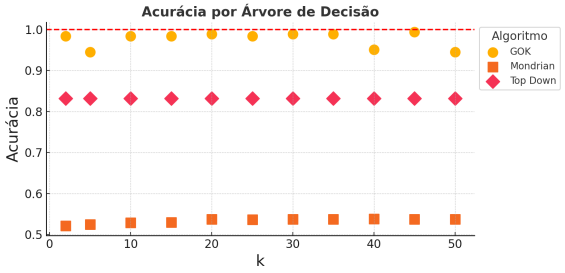

Figura 4. Acurácia do DT.

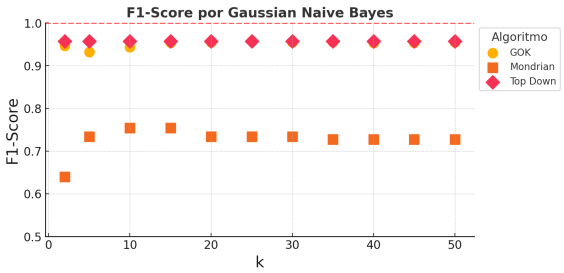

Figura 5. F1-Score do NB.

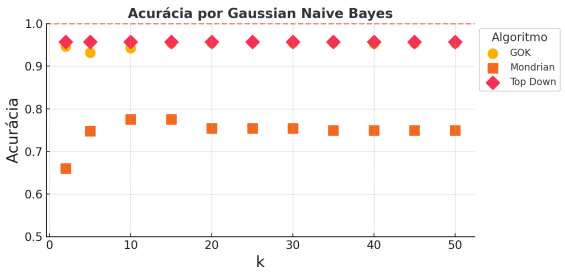

Figura 6. Acurácia do NB.

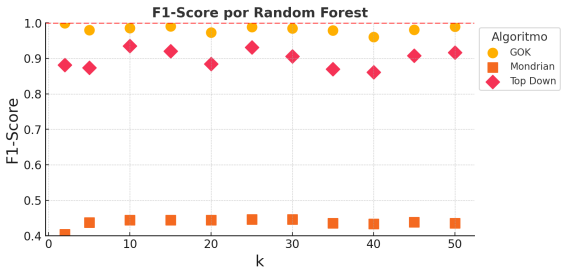

Figura 7. F1-Score do RF.

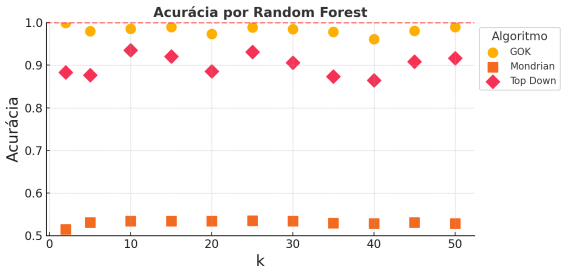

Figura 8. Acurácia do RF.

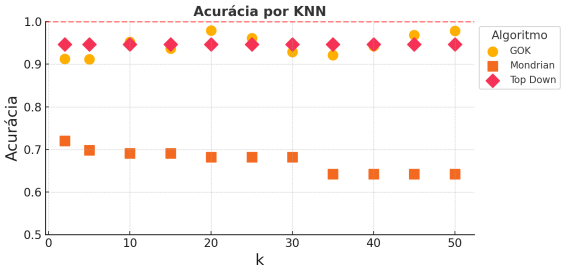

Figura 9. Acurácia do KNN.

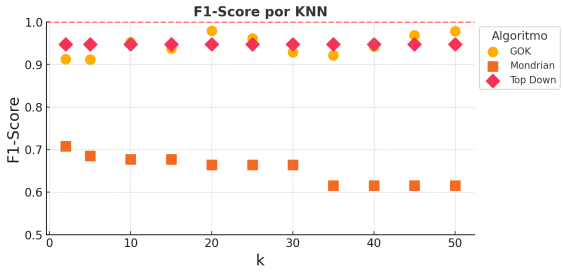

Figura 10. F1-Score do KNN.

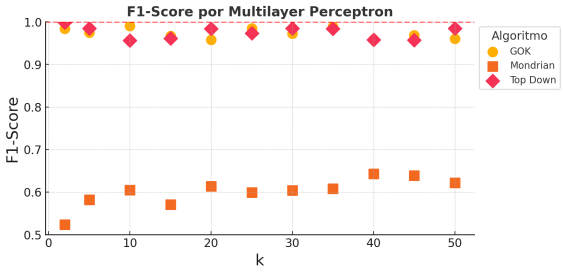

Figura 11. F1-Score do MLP.

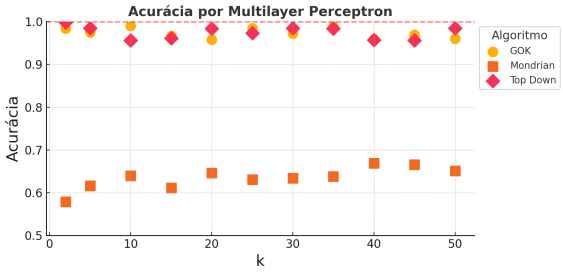

Figura 12. Acurácia do MLP.

Com base nos resultados apresentados nos gráficos para diferentes modelos de *machine learning* (*Decision Tree*, *KNN*, *Gaussian Naive Bayes*, *MLP* e *Random Forest*), é possível observar um padrão consistente no desempenho dos três algoritmos de anonimização avaliados, considerando tanto a acurácia quanto o *F1-score*.

O *GOK-Privacy* demonstrou superioridade em ambos os indicadores, mantendo acurácias elevadas entre 90% e 95% e *F1-scores* proporcionalmente altos na maioria dos modelos. Esta performance robusta sugere que o método consegue preservar eficientemente as relações importantes nos dados enquanto mantém a privacidade, sendo particularmente eficaz em modelos mais complexos, como *MLP* e *Random Forest*, onde os valores de *F1-score* confirmam a alta precisão e a capacidade de generalização dos modelos.

O *Top Down* apresentou uma performance intermediária, mas competitiva, especialmente nos modelos *KNN*, *Gaussian Naive Bayes* e *MLP*, onde tanto a acurácia quanto o *F1-score* foram próximos aos resultados obtidos pelo *GOK-Privacy*. Apesar disso, algumas limitações foram notadas em modelos como *Decision Tree*, onde sua acurácia e *F1-score* ficaram em torno de 65%. A estabilidade do *Top Down* em diferentes valores de k demonstra que o algoritmo é robusto às variações no nível de anonimização, o que o torna uma alternativa viável ao *GOK-Privacy*, embora menos eficaz.

O *Mondrian*, por outro lado, apresentou consistentemente os piores desempenhos entre os algoritmos avaliados. Suas acurácias variaram entre 55% e 75%, enquanto os *F1-scores* foram proporcionalmente baixos, especialmente em modelos como *Decision Tree* e *Random Forest*, onde os valores ficaram próximos ao limite inferior do intervalo. Mesmo em cenários mais favoráveis, como com o modelo *KNN*, o *Mondrian* não conseguiu competir em igualdade com os outros algoritmos. Esses resultados reafirmam que o *GOK-Privacy* é a melhor escolha entre os três métodos, oferecendo o melhor compromisso entre privacidade e utilidade dos dados, com o *Top Down* como uma alternativa razoável e o *Mondrian* apresentando limitações significativas em tarefas de ML.

5. Conclusão

A coleta massiva de dados revolucionou a criação de soluções inovadoras baseadas em IA. No entanto o uso desses dados exige a conformidade com leis de privacidade. O desafio reside em conciliar a necessidade de anonimizar dados para proteger a privacidade com a manutenção da eficácia dos modelos de IA, uma vez que métodos tradicionais de anonimização podem prejudicar o desempenho desses modelos.

A partir dessa realidade, este trabalho propôs o *GOK-Privacy* que demonstrou ser uma solução eficaz para o desafio de equilibrar privacidade e utilidade em dados de IoT na área da saúde, particularmente em UTIs. A combinação do GOA com técnicas de clustering fuzzy permitiu preservar k-anonimidade sem comprometer significativamente o desempenho dos modelos de aprendizado de máquina. Os resultados experimentais indicaram que o *GOK-Privacy* supera abordagens tradicionais, como *Mondrian* e *Top-Down*, em termos de precisão, mesmo em cenários com requisitos elevados de anonimização. Além disso, a avaliação detalhada revelou a capacidade do *GOK-Privacy* de preservar padrões críticos nos dados enquanto atende aos requisitos regulatórios impostos por legislações como GDPR e LGPD. Esses resultados reforçam a viabilidade da abordagem proposta para cenários sensíveis, como o setor de saúde, abrindo caminho para futuras aplicações e melhorias, incluindo sua adaptação a diferentes contextos de privacidade e aprendizado de máquina em redes IoT.

Como trabalhos futuros, pretende-se avaliar a proposta com outras abordagens de otimização e avaliá-la considerando conjuntos de dados de diversos contextos, como saúde, segurança, documentos judiciais, dentre outros.

Agradecimentos

Pesquisa parcialmente financiada pelo CNPq (Processos 405940/2022-0 e 303877/2021-9) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 88887.954253/2024-00.

Referências

Abdollahzadeh, B., Gharehchopogh, F. S., and Mirjalili, S. (2021). A novel metaheuristic optimization algorithm inspired by gorilla troops' behaviors. *Expert Systems with Applications*, 182:115083.

Chiu, C. C. and Tsai, C. Y. (2007). Weighted feature c-means clustering algorithm for data mining in intelligent transportation systems. *Expert Systems with Applications*, 33(1).

Choudhury, O., Gkoulalas-Divanis, A., Salonidis, T., Sylla, I., Park, Y., Hsu, G., and Das, A. (2020). Anonymizing data for privacy-preserving federated learning. *arXiv preprint*, arXiv:2002.09096.

El Mestari, S. Z., Lenzini, G., and Demirci, H. (2024). Preserving data privacy in machine learning systems. *Computers Security*, 137:103605.

Ferreira, M. C., Ribeiro, S. E., Nobre, F. V., Linhares, M. L., Araújo, T. P., and Gomes, R. L. (2024). Mitigating measurement failures in throughput performance forecasting. In *2024 20th International Conference on Network and Service Management (CNSM)*, pages 1–7.

Gomes, R. L., Bittencourt, L. F., and Madeira, E. R. M. (2014a). A similarity model for virtual networks negotiation. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, SAC '14, page 489–494, New York, NY, USA. Association for Computing Machinery.

Gomes, R. L., Bittencourt, L. F., Madeira, E. R. M., Cerqueira, E., and Gerla, M. (2014b). An architecture for dynamic resource adjustment in vsdns based on traffic demand. In *2014 IEEE Global Communications Conference*, pages 2005–2010.

He, X., Chen, H., Chen, Y., Dong, Y., Wang, P., and Huang, Z. (2012). Clustering-based k-anonymity. In *Advances in Knowledge Discovery and Data Mining: 16th Pacific-Asia Conference, PAKDD 2012, Kuala Lumpur, Malaysia, May 29-June 1, 2012, Proceedings, Part I 16*, pages 405–417. Springer.

Hussain, F., Abbas, S. G., Shah, G. A., Pires, I. M., Fayyaz, U. U., Shahzad, F., Garcia, N. M., and Zdravevski, E. (2021). A framework for malicious traffic detection in iot healthcare environment. *Sensors*, 21(9):3025.

Kacha, L., Zitouni, A., and Djoudi, M. (2021). Kab: A new k-anonymity approach based on black hole algorithm. *Journal of King Saud University - Computer and Information Sciences*.

Kumar, R., Chen, W., and Smith, S. (2024). Privacy-preserving machine learning through k-anonymity: A novel approach for healthcare data protection. *Journal of Medical Systems*, 48(1):1–15.

Langari, R. K., Sardar, S., Mousavi, S. A. A., and Radfar, R. (2020). Combined fuzzy clustering and firefly algorithm for privacy preserving in social networks. *Expert Systems With Applications*, 141:112968.

LeFevre, K., DeWitt, D. J., and Ramakrishnan, R. (2006). Mondrian multidimensional k-anonymity. In *22nd International Conference on Data Engineering (ICDE'06)*, pages 25–25. IEEE.

Ni, C., Cang, L. S., Gope, P., and Min, G. (2022). Data anonymization evaluation for big data and iot environment. *Information Sciences*, 605:381–392.

Pimenta, I. A., Silva, D. A., Moura, E. S., Silveira, M. M., and Gomes, R. L. (2024). Impact of data anonymization in machine learning models. In *13th Latin-American Symposium on Dependable and Secure Computing (LADC 2024)*, pages 1–4, Recife, Brazil.

Portela, A. L., Menezes, R. A., Costa, W. L., Silveira, M. M., Bittecnourt, L. F., and Gomes, R. L. (2023). Detection of iot devices and network anomalies based on anonymized network traffic. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, pages 1–6.

Portela, A. L. C., Ribeiro, S. E. S. B., Menezes, R. A., de Araujo, T., and Gomes, R. L. (2024). T-for: An adaptable forecasting model for throughput performance. *IEEE Transactions on Network and Service Management*, pages 1–1.

Seh, A. H., Zarour, M., Alenezi, M., Sarkar, A. K., Agrawal, A., Kumar, R., and Khan, R. A. (2020). Healthcare data breaches: Insights and implications. *Healthcare*, 8(2):133.

Silva, M., Ribeiro, S., Carvalho, V., Cardoso, F., and Gomes, R. L. (2023). Scalable detection of sql injection in cyber physical systems. In *Proceedings of the 12th Latin-American Symposium on Dependable and Secure Computing*, LADC '23, page 220–225, New York, NY, USA. Association for Computing Machinery.

Silva, M. V., Mosca, E. E., and Gomes, R. L. (2022). Green industrial internet of things through data compression. *International Journal of Embedded Systems*, 15(6):457–466.

Silveira, M., Santos, D., Souza, M., Silva, D., Mesquita, M., Neto, J., and Gome, R. L. (2023a). An anonymization service for privacy in data mining. In *Proceedings of the 12th Latin-American Symposium on Dependable and Secure Computing*, LADC '23, page 214–219, New York, NY, USA. Association for Computing Machinery.

Silveira, M. M., Portela, A. L., Menezes, R. A., Souza, M. S., Silva, D. S., Mesquita, M. C., and Gomes, R. L. (2023b). Data protection based on searchable encryption and anonymization techniques. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, pages 1–5.

Silveira, M. M., Silva, D. S., Rodriguez, S. J. R., and Gomes, R. L. (2023c). Searchable symmetric encryption for private data protection in cloud environments. In *Proceedings of the 11th Latin-American Symposium on Dependable Computing*, LADC '22, page 95–98, New York, NY, USA. Association for Computing Machinery.

Slijepčević, D., Henzl, M., Klausner, L. D., Dam, T., Kieseberg, P., and Zeppelzauer, M. (2021). k-anonymity in practice: How generalisation and suppression affect machine learning classifiers. *Computers & Security*, 111:102488.

Souza, M. S., Ribeiro, S. E. S. B., Lima, V. C., Cardoso, F. J., and Gomes, R. L. (2024). Combining regular expressions and machine learning for sql injection detection in urban computing. *Journal of Internet Services and Applications*, 15(1):103–111.

Xu, J., Wang, W., Pei, J., Wang, X., Shi, B., and Fu, A. W.-C. (2006). Utility-based anonymization using local recoding. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–790. ACM.

Yuan, S. and Wu, X. (2022). Trustworthy anomaly detection: A survey. *arXiv preprint*, arXiv:2202.07787.