Privacidade Diferencial Adaptativa por Meio de Informação Mútua para Preservação da Utilidade em Aprendizado de Máquina

Ivo A. Pimenta[1], **Rafael L. Gomes**[1] **(Orientador)**

[1]Universidade Estadual do Ceará (UECE)

aguiar.pimenta@aluno.uece.br, rafa.lopes@uece.br

Abstract. *The increasing use of machine learning on sensitive data has amplified the need for privacy preservation mechanisms. However, traditional Differential Privacy mechanisms apply uniform perturbation to all attributes, disregarding differences in predictive relevance and statistical dependencies between features. Within this context, the dissertation presented in this document proposed Mutual Information Adaptive Differential Privacy (MIADP), a model-independent anonymization mechanism that adaptively distributes the privacy budget using Mutual Information to measure the relevance of attributes and employs correlation-sensitive noise injection to preserve relevant statistical structures. Experimental evaluations performed on different real-world datasets demonstrate that MIADP consistently outperforms traditional mechanisms, especially under severe privacy constraints.*

Resumo. *A crescente utilização de aprendizado de máquina sobre dados sensíveis tem ampliado a necessidade de mecanismos de preservação de privacidade. Contudo, mecanismos tradicionais de Privacidade Diferencial aplicam perturbação uniforme a todos os atributos, desconsiderando diferenças de relevância preditiva e dependências estatísticas entre características. Dentro desse contexto, a dissertação apresentada neste documento propôs o Mutual Information Adaptive Differential Privacy (MIADP), um mecanismo de anonimização independente de modelo que distribui o orçamento de privacidade de forma adaptativa utilizando Informação Mútua para mensurar a relevância dos atributos e emprega injeção de ruído sensível à correlação para preservar estruturas estatísticas relevantes. Avaliações experimentais realizadas em diferentes conjuntos de dados reais demonstram que o MIADP supera consistentemente mecanismos tradicionais, especialmente sob restrições severas de privacidade.*

1. Introdução

O crescimento acelerado de aplicações baseadas em dados, especialmente em domínios como saúde, cidades inteligentes, Internet das Coisas e sistemas distribuídos, tem ampliado a exposição de informações sensíveis [Kamalov et al. 2023]. Embora a Privacidade Diferencial seja atualmente considerada um dos mecanismos mais robustos para proteção de dados, sua aplicação tradicional baseia-se na adição uniforme de ruído a todos os atributos do conjunto de dados [Hyrup et al. 2025]. Entretanto, atributos possuem relevâncias distintas para tarefas de aprendizado de máquina [Gong et al. 2022].

A aplicação uniforme de ruído frequentemente degrada atributos altamente informativos e reduz significativamente o desempenho dos modelos treinados sobre dados anonimizados. Além disso, mecanismos convencionais ignoram dependências estatísticas existentes entre atributos correlacionados, comprometendo a preservação da estrutura dos dados [Silveira et al. 2023]. Dessa forma, surge a necessidade de mecanismos adaptativos capazes de considerar simultaneamente a relevância preditiva dos atributos e suas correlações estatísticas, promovendo um equilíbrio mais eficiente entre proteção da privacidade e preservação da utilidade dos dados.

1.1. Motivação

Embora mecanismos de anonimização sejam amplamente empregados para reduzir riscos de privacidade, abordagens tradicionais frequentemente comprometem a utilidade dos dados ao aplicar níveis uniformes de perturbação, desconsiderando a relevância dos atributos para as tarefas analíticas. Esse problema torna-se ainda mais evidente em cenários de computação distribuída e de borda, nos quais restrições de processamento coexistem com a necessidade de análises precisas e preservação de privacidade. Nesse contexto, surge a motivação para investigar mecanismos adaptativos capazes de equilibrar de forma mais eficiente a proteção dos dados e a manutenção de sua utilidade, explorando informações sobre a importância e a correlação entre atributos para orientar o processo de anonimização.

1.2. Objetivos

Dentro do contexto em questão, a dissertação apresentada neste documento tem como objetivo desenvolver e avaliar um mecanismo adaptativo de anonimização baseado em Privacidade Diferencial capaz de melhorar o equilíbrio entre privacidade e utilidade em aplicações de aprendizado de máquina. Para atingir esse objetivo, o trabalho investiga inicialmente as limitações dos mecanismos tradicionais de Privacidade Diferencial, que distribuem uniformemente o orçamento de privacidade entre todos os atributos do conjunto de dados, desconsiderando diferenças de relevância preditiva e dependências estatísticas existentes entre as características.

A partir dessa análise, propõe-se a utilização da Informação Mútua como mecanismo para quantificar a relevância de cada atributo em relação à variável alvo, permitindo que o orçamento de privacidade seja distribuído de forma adaptativa. Além disso, o trabalho busca preservar relações estatísticas importantes entre atributos por meio da identificação de grupos de características correlacionadas e da aplicação de uma estratégia de perturbação sensível à correlação.

1.3. Inovação e Contribuições

A principal originalidade da proposta reside na integração de dois elementos que normalmente são tratados de forma independente na literatura: a relevância informacional dos atributos e sua estrutura de correlação. Enquanto abordagens tradicionais distribuem uniformemente o orçamento de privacidade ou consideram apenas aspectos relacionados ao treinamento dos modelos, o MIADP utiliza Informação Mútua para orientar a alocação do orçamento e incorpora mecanismos de preservação da estrutura estatística durante a perturbação dos dados. Essa combinação permite reduzir perdas de utilidade sem comprometer as garantias formais de privacidade.

Como contribuições científicas, este trabalho propõe um mecanismo inédito de anonimização adaptativa baseado em Informação Mútua, desenvolve uma estratégia de distribuição diferenciada do orçamento de privacidade em nível de atributo, introduz uma abordagem de perturbação sensível à correlação para preservação de dependências estatísticas e realiza uma extensa avaliação experimental envolvendo diferentes conjuntos de dados, algoritmos de aprendizado de máquina e cenários de privacidade. Os resultados demonstram que o MIADP supera consistentemente mecanismos tradicionais baseados no mecanismo de Laplace, especialmente sob orçamentos de privacidade mais restritivos, evidenciando que a utilização conjunta de relevância informacional e correlação estatística constitui uma alternativa promissora para o desenvolvimento de mecanismos de anonimização orientados à utilidade.

2. Proposta e Contribuições

A Figura 1 apresenta uma visão geral do *Mutual Information Adaptive Differential Privacy* (MIADP), mecanismo proposto nesta dissertação para anonimização de dados em cenários de aprendizado de máquina. O objetivo central da proposta é superar uma das principais limitações dos mecanismos tradicionais de Privacidade Diferencial (DP): a aplicação uniforme de ruído sobre todos os atributos do conjunto de dados, independentemente de sua relevância para a tarefa analítica. Embora essa estratégia forneça garantias formais de privacidade, ela frequentemente degrada a utilidade dos dados, especialmente em aplicações de aprendizado de máquina nas quais diferentes atributos possuem contribuições preditivas distintas.

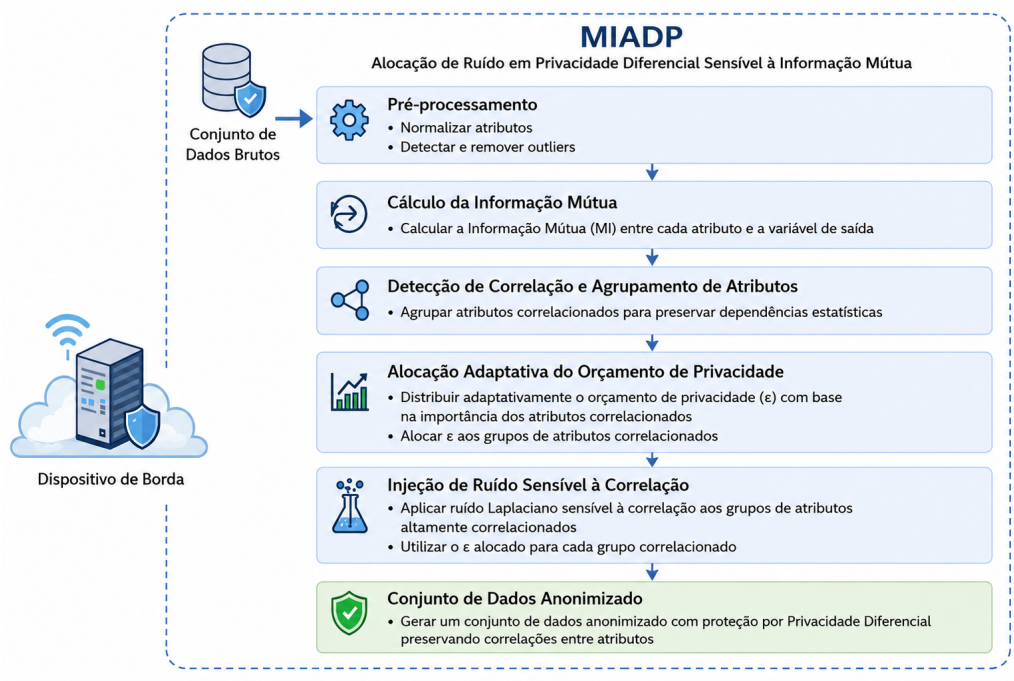

Figura 1. visão geral da proposta.

O fluxo do MIADP inicia-se com uma etapa de pré-processamento, responsável pela normalização dos atributos e remoção de valores discrepantes. Em seguida, o mecanismo calcula a Informação Mútua (*Mutual Information – MI*) entre cada atributo e a

variável alvo. Diferentemente de métricas puramente lineares, a Informação Mútua é capaz de capturar relações lineares e não lineares, fornecendo uma medida mais robusta da relevância informacional de cada característica para o processo de classificação. Os valores obtidos são posteriormente normalizados e utilizados como base para a distribuição adaptativa do orçamento de privacidade.

Após a etapa de cálculo da Informação Mútua, o mecanismo realiza a detecção de correlações entre atributos. Características que apresentam forte dependência estatística são agrupadas em conjuntos correlacionados, permitindo que o processo de anonimização considere explicitamente a estrutura dos dados. Essa etapa busca minimizar um problema recorrente em mecanismos tradicionais de DP, nos quais atributos fortemente relacionados são perturbados de forma independente, ocasionando a destruição de dependências estatísticas relevantes para tarefas de aprendizado de máquina.

Com base nos escores de Informação Mútua e nos grupos de correlação identificados, o MIADP executa a etapa de alocação adaptativa do orçamento de privacidade. Nesse processo, atributos mais relevantes recebem uma parcela maior do orçamento de privacidade, sofrendo menor perturbação, enquanto atributos menos informativos recebem níveis mais elevados de ruído. Para atributos pertencentes ao mesmo grupo correlacionado, o orçamento é ajustado de forma conjunta, preservando a coerência estatística entre as variáveis.

Na etapa seguinte, denominada Injeção de ruído sensível a correlação, é realizado o processo de anonimização propriamente dito. Diferentemente da aplicação independente do mecanismo de Laplace tradicional, o MIADP injeta ruído considerando simultaneamente a relevância dos atributos e suas dependências estatísticas. Como resultado, o conjunto de dados anonimizado preserva de forma mais fiel as propriedades analíticas do conjunto original, mantendo as garantias formais de Privacidade Diferencial.

3. Resultados

3.1. Configuração Experimental

A avaliação experimental do MIADP foi conduzida com o objetivo de analisar simultaneamente a utilidade dos dados anonimizados e a efetividade da preservação de privacidade. Para isso, foram utilizados quatro conjuntos de dados amplamente empregados na literatura de mineração de dados e aprendizado de máquina: *Adult Income*, *Contraceptive Method Choice* (CMC), *Heart Disease* e *Mammographic Mass Dataset* (MGM). Esses conjuntos apresentam diferentes características em termos de dimensionalidade, distribuição dos atributos e complexidade de classificação, permitindo uma avaliação abrangente do comportamento do mecanismo proposto.

Os experimentos compararam o MIADP com o mecanismo tradicional de Privacidade Diferencial baseado no mecanismo de Laplace considerados tanto valores mínimos quanto máximos sob diferentes valores do orçamento de privacidade (ϵ). Após a anonimização dos dados, foram treinados cinco algoritmos de aprendizado de máquina amplamente utilizados na literatura: Random Forest (RF), AdaBoost (ADB), Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN) e Gaussian Naive Bayes (GNB). Para garantir uma comparação justa, os modelos tiveram seus hiperparâmetros ajustados por meio de otimização bayesiana associada à validação cruzada.

A utilidade dos dados foi avaliada por meio do *F1-score*, métrica amplamente utilizada em problemas de classificação por combinar precisão e revocação em uma única medida. Adicionalmente, a efetividade da preservação de privacidade foi analisada utilizando ataques de inferência de pertencimento (*Membership Inference Attacks* – MIA), cuja eficácia foi medida por meio da acurácia do atacante. Estas métricas são usualmente analisadas em contextos de dados sensíveis e garantias de privacidade [Gong et al. 2022, Hyrup et al. 2025].

As Figuras 2 a 5 apresentam os resultados de desempenho dos classificadores treinados sobre os conjuntos de dados anonimizados. Em todos os cenários avaliados observa-se o comportamento esperado de aumento do F1-score à medida que o orçamento de privacidade se torna menos restritivo. Entretanto, mesmo para valores reduzidos de ϵ, nos quais uma quantidade maior de ruído é inserida nos dados, o MIADP manteve níveis de desempenho superiores aos obtidos pelo mecanismo tradicional baseado em Laplace.

No conjunto Adult (Figura 2), os ganhos obtidos pelo MIADP foram consistentes para todos os classificadores avaliados, sendo mais evidentes nos cenários de privacidade mais rigorosos. Esse comportamento indica que a estratégia de alocação adaptativa baseada em Informação Mútua foi capaz de preservar com maior eficiência os atributos mais relevantes para a tarefa de classificação. Tendência semelhante foi observada nos conjuntos MGM e CMC (Figuras 3 e 4), nos quais as diferenças entre o MIADP e o mecanismo de Laplace tornam-se ainda mais pronunciadas, sugerindo que a proposta é particularmente vantajosa em bases de dados que apresentam atributos com relevâncias distintas para a variável alvo. No conjunto Heart (Figura 5), embora os ganhos absolutos sejam menores, o MIADP manteve desempenho superior ou equivalente ao mecanismo tradicional em praticamente todas as configurações avaliadas.

A Figura 6 apresenta os resultados da avaliação de preservação de privacidade por meio de ataques de inferência de pertencimento (*Membership Inference Attacks* – MIA). Nesse tipo de ataque, valores de acurácia próximos de 50% indicam que o atacante possui desempenho semelhante ao acaso, representando maior proteção à privacidade. Em contrapartida, valores mais elevados indicam maior capacidade de identificar se uma determinada amostra participou do treinamento do modelo.

À medida que o valor de ϵ aumenta, observa-se uma tendência de crescimento da acurácia do atacante para ambos os mecanismos, comportamento esperado devido à redução da quantidade de ruído inserida nos dados. Entretanto, mesmo quando o MIADP apresenta valores ligeiramente superiores em alguns conjuntos de dados, essas diferenças permanecem pequenas quando comparadas aos ganhos de utilidade observados nas avaliações de F1-score. De forma geral, os resultados demonstram que a proposta mantém níveis de proteção comparáveis aos da Privacidade Diferencial tradicional, evidenciando que os ganhos de desempenho obtidos pelo MIADP não são alcançados à custa da redução das garantias de privacidade.

4. Produção Científica e Tecnológica

Com relação à produção científica, têm-se os seguintes artigos diretamente associados à dissertação: [Pimenta et al. 2024b, Pimenta et al. 2024c, Pimenta et al. 2024a, Pimenta et al. 2025a, Pimenta et al. 2026c, Pimenta et al. 2025b, Pimenta et al. 2026a, Pimenta et al. 2026b]. Adicionalmente, tem-se a se-

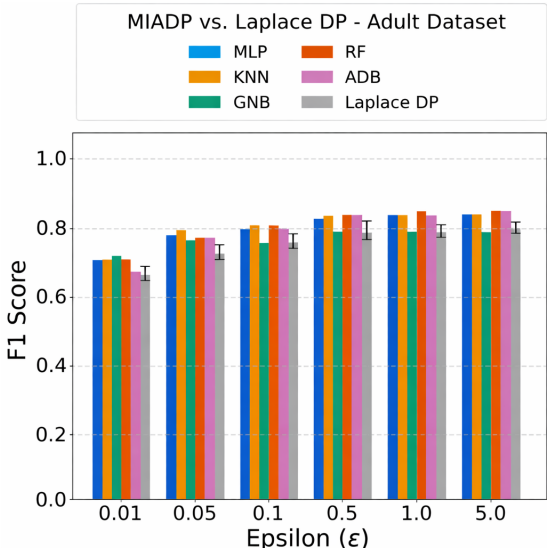

Figura 2. Classification Performance (Adult)

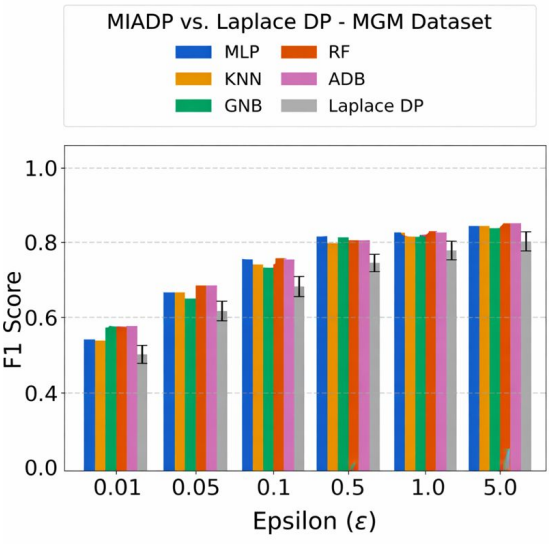

Figura 3. Classification Performance (MGM)

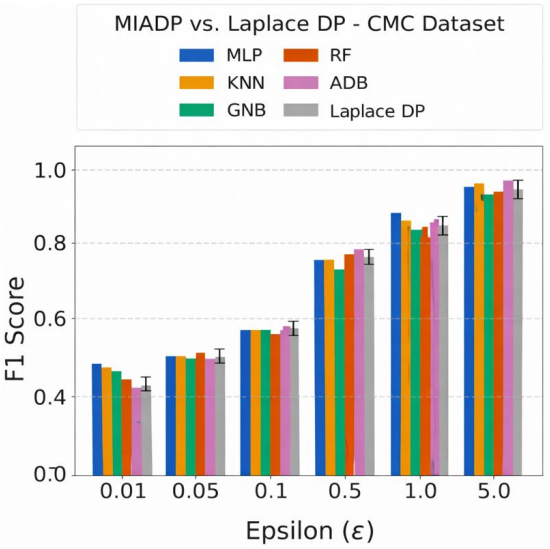

Figura 4. Classification Performance (CMC)

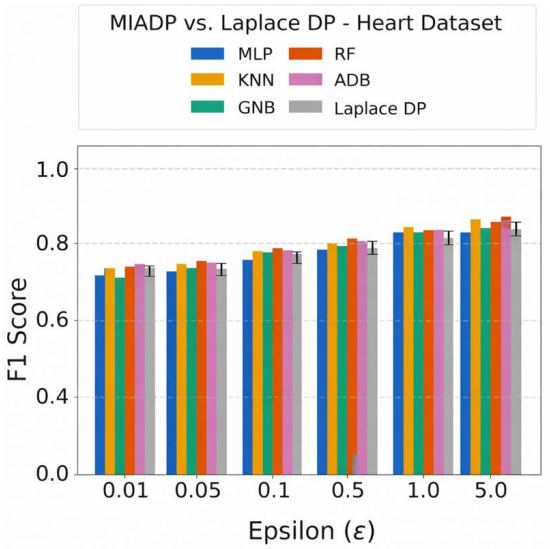

Figura 5. Classification Performance (Heart)

guinte produção científica complementar desenvolvida durante o mestrado: [Urbano et al. 2024, Gomes et al. 2024, Nascimento et al. 2026a, Mesquita et al. 2026, Souza et al. 2024, Nascimento et al. 2026b].

Registros de software oriundos da proposta desenvolvida:

- **ATGK: Algoritmo da Tropa dos Gorilas Integrado com K-Anonimato**. Registro de software concedido pelo INPI(Instituto Nacional de Propriedade Industrial), sobre o número de processo: 512025005950-7
- **MIADP: Ferramenta de Anonimização Diferencial Adaptativa Baseada em Informação Mútua**. Registro de software concedido pelo INPI(Instituto Nacional de Propriedade Industrial), sobre o número de processo: 512026004244-5

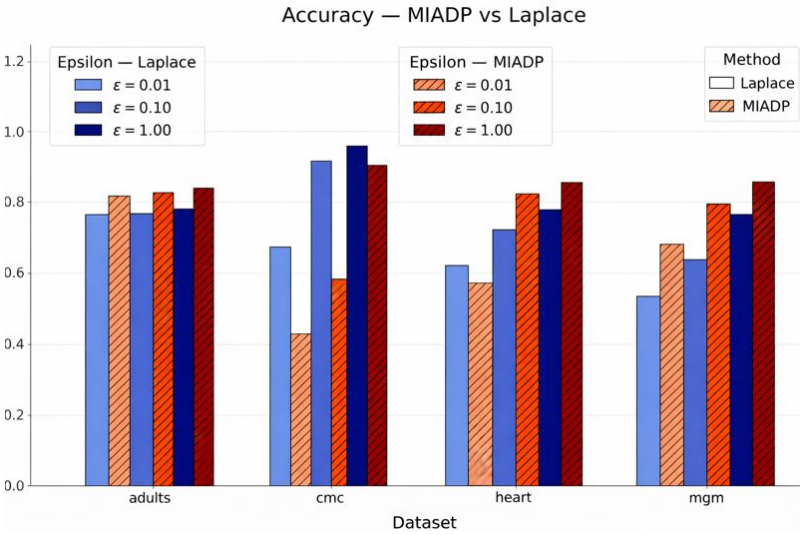

Figura 6. Privacy Preservation Evaluation

Referências

Gomes, R., Mosca, E., Aquino, C., Braga Neto, A., and Pimenta, I. (2024). Agricultura de precisão integrando de rede em malha iot e computação em nuvem. *Revista de Ciência da Computação*, 6(1):5–12.

Gong, X., Chen, Y., Wang, Q., Wang, M., and Li, S. (2022). Private data inference attacks against cloud: Model, technologies, and research directions. *IEEE Communications Magazine*, 60(9):46–52.

Hyrup, T., Lautrup, A. D., Zimek, A., and Schneider-Kamp, P. (2025). A systematic review of privacy-preserving techniques for synthetic tabular health data. *Discover Data*, 3(1):1–32.

Kamalov, F., Pourghebleh, B., Gheisari, M., Liu, Y., and Moussa, S. (2023). Internet of medical things privacy and security: Challenges, solutions, and future trends from a new perspective. *Sustainability*, 15(4).

Mesquita, M., Linhares, M., Portela, A., Pimenta, I., Araújo, T., and Gomes, R. (2026). Predição iterativa de vazão de rede utilizando uma abordagem de janelas deslizantes combinada com imputação de dados dinâmica. In *Anais do XLIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 1164–1177, Porto Alegre, RS, Brasil. SBC.

Nascimento, E., Pimenta, I., Lee, M., Santos, J., Araújo, T., and Gomes, R. (2026a). Usando representações não supervisionadas para extração de características em grafos em dados de redes sociais. In *Anais do XLIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 1499–1512, Porto Alegre, RS, Brasil. SBC.

Nascimento, E. S., Pimenta, I. A., Lee, M. H., de Araujo, T., and Gomes, R. L. (2026b). A decoupled embedding-based framework for efficient node classification on social networks. *Journal of Internet Services and Applications*, 17(1):214–223.

Pimenta, I., Araújo, R., Rodrigues, R., Silveira, M., and Gomes, R. (2025a). Anonimização de dados para inteligência artificial usando o algoritmo da tropa dos gorilas. In *Anais do XLIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 448–461, Porto Alegre, RS, Brasil. SBC.

Pimenta, I., Braga Neto, A., Moura, E., Freitas, K., Lee, M., and Gomes, R. (2025b). Gerenciamento adaptÁvel de dispositivos iot heterogÊneos em casas inteligentes. *Revista de Ciência da Computação*, 7(1):26–32.

Pimenta, I., Freitas, K., Moura, E., Nascimento, E., Faria, F., and Gomes, R. (2026a). Perturbação controlada de autovetores em pca para preservação de privacidade em dados sensíveis. In *Anais do XLIV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 1136–1149, Porto Alegre, RS, Brasil. SBC.

Pimenta, I., Lee, M., Moura, E., Nascimento, E., Filho, G. R., and Gomes, R. (2026b). Preservando a utilidade de dados urbanos via alocação adaptativa de ruído em privacidade diferencial. In *Anais do X Workshop de Computação Urbana*, pages 239–252, Porto Alegre, RS, Brasil. SBC.

Pimenta, I., Silva, D., Moura, E., Silveira, M., and Gomes, R. L. (2024a). Impact of data anonymization in machine learning models. In *Proceedings of the 13th Latin-American Symposium on Dependable and Secure Computing*, LADC '24, page 188–191, New York, NY, USA. Association for Computing Machinery.

Pimenta, I., Souza, M., Aquino, C., Portela, A., and Gomes, R. (2024b). Predição de qualidade de vídeo em serviços de streaming utilizando redes 5g. In *Anais do VIII Workshop de Computação Urbana*, pages 43–56, Porto Alegre, RS, Brasil. SBC.

Pimenta, I. A., Aquino, C. A., Almeida, Y. O., Lima, V. C., and Gomes, R. L. (2024c). Prediction of multimedia quality over 5g networks in urban environments. In *2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*, pages 748–755.

Pimenta, I. A., Lee, M. H., Bittencourt, L. F., and Gomes, R. L. (2026c). Adaptive privacy based on mutual information for machine learning in edge–cloud environments. *IEEE Networking Letters*, 8:49–53.

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

Souza, M., Ribeiro, S., Pimenta, I., Almeida, Y., Cardoso, F., and Gomes, R. (2024). Detecção inteligente de injeção de sql integrando ambientes de nuvem e borda. In *Anais do XLII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 435–448, Porto Alegre, RS, Brasil. SBC.

Urbano, A., Costa, Y., Paula, M., Portela, A., Pimenta, I., and Gomes, R. (2024). Detecção inteligente de ips maliciosos através do monitoramento de ameaças. In *Anais do XVI Simpósio Brasileiro de Computação Ubíqua e Pervasiva*, pages 21–30, Porto Alegre, RS, Brasil. SBC.