Impacto da Redução de Dimensão e Seleção de Atributos na Generalização de Modelos de Detecção de Intrusão

Kelson Carvalho Santos[1,2], **Rodrigo Sanches Miani**[1]

[1]Universidade Federal de Uberlândia (UFU) - Uberlândia, MG - Brasil

[2]Instituto Federal do Piauí (IFPI) - Teresina, PI - Brasil

{kelson,miani}@ufu.br, kelson@ifpi.edu.br

Abstract. *Previous studies have presented few initiatives to improve the generalization of intrusion detection models in different network traffic scenarios. In the search for solutions to this issue, the frequent use of the Dimension Reduction technique stands out. However, comparative analyses between this technique and similar approaches are still scarce. This paper proposes the implementation and comparative analysis of Dimension Reduction and Feature Selection to evaluate their impacts as strategies to improve the generalization of machine learning-based intrusion detection models. The results obtained indicate that both can contribute to facing this challenge.*

Resumo. *Trabalhos anteriores apresentam poucas iniciativas voltadas à melhoria da generalização de modelos de detecção de intrusão em cenários distintos de tráfego de rede. Na busca por soluções para esse problema, destaca-se o uso frequente da técnica de Redução de Dimensão. No entanto, análises comparativas entre essa técnica e outras abordagens semelhantes ainda são escassas. Este trabalho propõe a implementação e a análise comparativa da Redução de Dimensão e da Seleção de Atributos, com o objetivo de avaliar seus impactos como estratégias para melhorar a generalização de modelos de detecção de intrusão baseados em aprendizado de máquina. Os resultados obtidos indicam que ambas podem contribuir no enfrentamento desse desafio.*

1. Introdução

Os Sistemas de Detecção de Intrusão *(IDS)* representam mecanismos de segurança cibernética projetados para monitorar, analisar e alertar contra comportamentos suspeitos ou maliciosos no tráfego de rede. De acordo com a metodologia de detecção, um *IDS* pode ser categorizado em duas abordagens principais: Detecção baseada em Assinaturas e Detecção baseada em Anomalias. O *IDS* baseado em Assinatura compara atividades em tempo real com padrões de ataques conhecidos e bancos de dados de comportamentos maliciosos. No entanto, esta abordagem é menos robusta contra ataques desconhecidos [Naseer et al. 2021]. Por sua vez, o *IDS* baseado em Anomalia envolve o estabelecimento de um perfil do comportamento normal da rede. Assim, o *IDS* monitora as atividades em busca de desvios significativos desta norma [Mahfouz et al. 2020].

Nesse contexto, o Aprendizado de Máquina *(ML)* é uma área bastante utilizada no desenvolvimento de modelos de detecção de intrusão, pois se concentra no desenvolvimento de modelos capazes de identificar padrões em dados e tomar decisões com base

no aprendizado adquirido [Mahesh 2020]. No entanto, o desempenho dos modelos depende da capacidade de generalização na detecção de intrusão em diferentes ambientes de tráfego de rede [Khraisat et al. 2019]. A generalização de um *IDS* refere-se à capacidade de um modelo em identificar e classificar com eficácia novos tipos de dados ou ataques desconhecidos [Sudyana et al. 2024]. Essa habilidade permite que o *IDS* aplique o conhecimento adquirido durante o treinamento a conjuntos de dados diferentes daqueles utilizados no processo de aprendizado.

Nos últimos cinco anos, a generalização de detecção de intrusão tem despertado crescente interesse na comunidade científica. O estudo de Verkerken et al. [Verkerken et al. 2021] demonstrou que modelos treinados em um conjunto de dados são eficazes na identificação de ataques dentro desse mesmo domínio. Porém, seu desempenho diminui quando avaliados em conjuntos de dados com características distintas de tráfego de rede. Essa limitação também foi observada por Layeghy e Portmann [Layeghy and Portmann 2022] e D'hooge et al. [D'hooge et al. 2020], que, embora tenham destacado o problema, apresentaram limitações na aplicação de técnicas destinadas a melhorar o desempenho dos modelos diante dos desafios de generalização em cenários variados de tráfego de rede. Apesar dos trabalhos citados anteriormente realizarem experimentos e demonstrarem a ineficácia da generalização de modelos de detecção de intrusão em diferentes cenários de tráfego de rede, eles não propõem ações para melhorar esse desempenho. No escopo deste artigo, essas ações são denominadas intervenções, definidas como um conjunto de medidas destinadas a melhorar a generalização e a facilitar o reconhecimento de padrões pelos modelos de detecção de intrusão em diferentes cenários.

Na busca por melhorias na generalização de modelos, a revisão da literatura revela que a Redução de Dimensão é frequentemente explorada como uma estratégia de intervenção nos conjuntos de dados. Esse método consiste em simplificar os dados por meio da remoção de atributos redundantes ou pouco relevantes para a classificação [Zebari et al. 2020]. Apesar disso, os trabalhos anteriores que aplicaram a Redução de Dimensão não esclareceram como os conjuntos de dados foram estruturados ou padronizados com os mesmos atributos para a realização de testes de generalização em diferentes cenários de tráfego de rede. Até o momento da realização deste estudo, não foram identificados trabalhos que tenham conduzido uma avaliação comparativa entre a Redução de Dimensão e outras técnicas similares, com o objetivo de identificar o impacto desse método na melhoria da generalização de detecção de intrusão. Isso evidencia uma lacuna importante na literatura. Assim, o problema abordado neste artigo consiste na implementação e análise comparativa da Redução de Dimensão e da Seleção de Atributos. A pesquisa baseia-se nas seguintes questões:

- **QP1:** A Redução de Dimensão e a Seleção de Atributos contribuem para melhorar a generalização de modelos de detecção de intrusão?
- **QP2:** Entre a Redução de Dimensão e a Seleção de Atributos, qual abordagem tem maior impacto na melhoria da generalização de modelos de detecção de intrusão em diferentes cenários de tráfego de rede?

A metodologia proposta neste trabalho envolve a implementação e análise comparativa da Redução de Dimensão e Seleção de Atributos, com o objetivo de melhorar a generalização de modelos de detecção de intrusão em diferentes cenários de tráfego de rede. O processo é estruturado em etapas que incluem a seleção dos conjuntos de dados, a

implementação das intervenções nos dados, o treinamento e experimentação dos modelos em diferentes cenários, a avaliação da generalização e a interpretação dos resultados para verificar o impacto das intervenções empregadas. O estudo considera dois cenários de validação: *Intraset*, no qual os modelos são testados dentro do mesmo domínio de dados, e *Interset*, onde os modelos são avaliados em domínios de dados distintos, analisando sua capacidade de generalização para novos ambientes.

Neste artigo, é possível delinear as seguintes contribuições: (i) Análise e comparação de intervenções em conjuntos de dados para o melhoramento da generalização de modelos de detecção de intrusão em diferentes cenários de tráfego de rede; e (ii) Disposição de conjuntos de dados modificados, com estruturação e padronização de atributos para testar a generalização de modelos de detecção de intrusão. O restante do artigo está organizado da seguinte forma: a Seção 2 apresenta uma revisão de trabalhos relacionados que estudaram a questão da generalização de detecção de intrusão; a Seção 3 detalha a abordagem metodológica; a Seção 4 apresenta os experimentos realizados e os resultados encontrados para responder às questões de pesquisa; e finalmente, a Seção 5 descreve as conclusões e os trabalhos futuros.

2. Trabalhos Relacionados

A generalização de modelos de detecção de intrusão tem sido investigada nos últimos cinco anos, com destaque para os desafios na identificação de ataques em cenários fora do domínio de treinamento. No entanto, poucos trabalhos propõem intervenções para melhorar a generalização desses modelos. Isso evidencia uma lacuna na literatura.

D'hooge et al. [D'hooge et al. 2020] examina a generalização de detecção de intrusão em cenários de domínios único e múltiplo de dados. Apesar dos resultados demonstrarem bons desempenhos em cenários de domínio único, os modelos falharam em generalizar nos cenários de domínios múltiplos. Um novo estudo é conduzido por D'hooge et al. [D'hooge et al. 2023], no qual os resultados indicaram que os modelos conseguiram uma boa generalização na identificação de ataques *DDoS* em cenários de domínios múltiplos de dados. Porém, os autores confirmaram a dificuldade de generalização nos demais tipos de ataques. Os autores concentram-se em avaliar a capacidade de generalização dos modelos, sem a aplicação de intervenções para enfrentar o desafio da generalização em diferentes cenários. Já no estudo de Verkerken et al. [Verkerken et al. 2021] a Redução de Dimensão é empregada, usando a Análise de Componentes Principais *(PCA)* [Jolliffe and Cadima 2016], como intervenção em conjuntos de dados. Embora os modelos tenham demonstrado bons desempenhos de generalização em cenários de domínio único de dados, esse desempenho é insatisfatório nos cenários de domínios múltiplos.

Marvi et al. [Marvi et al. 2021] aplica a Seleção de Atributos Integrados (IFS) [Khalid et al. 2014] para identificar ataques *DDoS*. Os resultados demonstraram melhorias em cerca de 20% na generalização. No entanto, o estudo é restrito a ataques *DDoS*, sendo uma limitação em cenários com maior diversidade de ataques. Algo semelhante é realizado por Rocha et al. [Rocha et al. 2023], que apesar de utilizar Análise de Grupos de Ataques [Cieslak et al. 2020] na melhoria da generalização, limita a análise a um conjunto de dados, sem avaliar o desempenho em domínios múltiplos de dados.

Layeghy e Portmann [Layeghy and Portmann 2022] avalia a generalização de modelos na identificação de ataques não vistos ou desconhecidos. Os resultados apontam

variações no desempenho dos modelos ao avaliar a generalização em domínios múltiplos de dados. Apesar disso, os modelos enfrentam limitações para generalizar diferentes ataques. Por fim, Sudyana et al. [Sudyana et al. 2024] aplica a Redução de Dimensão empregando o *PCA* para investigar a generalização utilizando um conjunto de dados baseado no ciclo de vida [Yudha 2023]. Os resultados demonstram melhorias na capacidade de generalização dos modelos, indicando que o método proposto é robusto e adaptável aos desafios inerentes à generalização de detecção de intrusão, em comparação com os métodos tradicionais que utilizam conjuntos de dados baseados em ataques.

Os trabalhos citados evidenciam desafios e limitações na generalização de detecção de intrusão, especialmente em cenários que envolvem múltiplos domínios de dados. Além disso, observa-se uma escassez de estudos que aplicam intervenções em conjuntos de dados com o objetivo de melhorar a generalização de modelos. A Tabela 1 resume os trabalhos discutidos, destacando suas contribuições e limitações.

Tabela 1. Resumo dos trabalhos relacionados.

[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
[D'hooge et al. 2020]	7 Tree-Based 2 neighborhood-based 2 vector-based 1 regression-based	ausente	ausente	CIC-IDS2017 CIC-IDS2018	ausente	sim	sim
[Verkerken et al. 2021]	Autoencoder oSVM iForest	Redução de Dimensão	ausente	CIC-IDS2017 CIC-IDS2018	sim	sim	sim
[Marvi et al. 2021]	LGBM	Redução de Dimensão	ausente	CIC-DDoS-2019	ausente	sim	ausente
[Layeghy and Portmann 2022]	Feed Forward LSTM Random Forest Extra-Tree iForest oSVM SGD-oSVM	ausente	ausente	UNSW-NB15 CIC-IDS2018 ToN-IoT Bot-IoT	ausente	sim	sim
[D'hooge et al. 2023]	7 Tree-Based 2 neighborhood-based 2 vector-based 1 regression-based	ausente	ausente	CIC-IDS2017 CIC-IDS2018 CIC-DDoS-2019	ausente	sim	sim
[Rocha et al. 2023]	Decision Tree Random Forest SVM Logistic Regression Extra-Tree	Análise de Grupos de Ataques	ausente	CIC-IDS2017	ausente	sim	ausente
[Sudyana et al. 2024]	CNN XGBoost	Redução de Dimensão	ausente	CREMEv2 CIC-IDS2017 CIC-IDS2018 CIC-DDOS-2019 CCU Mirai HTTP	ausente	sim	sim
Este trabalho	Random Forest kNN XGBoost MLP (ANN)	Redução de Dimensão Seleção de Atributos	sim	UNSW-NB15 CIC-IDS2017 CIC-IDS2018	sim	sim	sim

[1] Referência, [2] Algoritmos, [3] Implementação de Estratégias, [4] Comparação das Estratégias, [5] Conjunto de Dados utilizados,
[6] Descrição dos Atributos Utilizados nos Testes de Generalização, [7] Avaliação em Domínio Único de Dados,
[8] Avaliação em Domínio Múltiplo de Dados. Fonte: Elaborado pelo autor.

A Tabela 1 evidencia o uso da Redução de Dimensão como o principal método para melhorar a generalização de detecção de intrusão. Além disso, observa-se que os trabalhos discutidos não fornecem detalhes sobre como os atributos foram estruturados para garantir a uniformidade nos conjuntos de dados. Essa uniformidade é essencial para testes de generalização em domínios múltiplos de dados. Este estudo se diferencia ao comparar a robustez da Redução de Dimensão com outras técnicas, como a Seleção de Atributos, na melhoria da generalização de modelos de detecção de intrusão.

3. Metodologia

Neste trabalho, propõe-se a implementação e análise comparativa da Redução de Dimensão e Seleção de Atributos, visando identificar o impacto de ambas na melhoria da generalização de modelos de detecção de intrusão em diferentes cenários de tráfego de rede. Para isso, a metodologia adotada segue uma abordagem organizada em sete etapas, estruturadas em um fluxo de trabalho sequencial, conforme ilustrado na Figura 1. As principais contribuições deste trabalho estão destacadas nas etapas em cor azul.

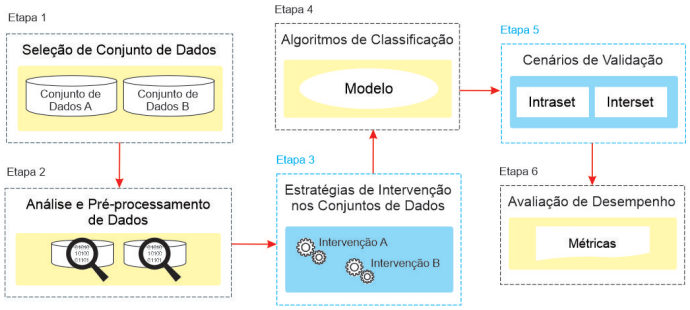

Figura 1. Visão Geral do Fluxo de Trabalho da Abordagem Proposta.
Fonte: Elaborado pelo autor.

Na Etapa 1, realiza-se a seleção dos conjuntos de dados para o desenvolvimento dos modelos de detecção de intrusão. Os dados contidos nesses conjuntos servirão como base para a criação dos padrões necessários à classificação. Por isso, é essencial que abranjam uma gama diversificada de informações que representem tanto atividades normais quanto maliciosas no tráfego da rede, garantindo maior robustez dos modelos na identificação desses padrões. Além disso, os conjuntos devem disponibilizar os dados brutos (arquivos *PCAP*) para extração dos fluxos de rede.

É comum os conjuntos de dados possuírem irregularidades que devem ser tratadas. Em sua maioria, por exemplo, é possível encontrar dados ausentes, inconsistentes ou redundantes. Esses e outros problemas podem influenciar na qualidade e na confiabilidade dos modelos [Kenyon et al. 2020]. Assim, na Etapa 2 é realizada a análise e o pré-processamento dos dados, que é empregada para a correção de problemas encontrados nos conjuntos de dados. Os métodos de limpeza de dados, codificação e normalização são algumas técnicas que podem ser aplicadas para o tratamento dos dados [Santos et al. 2024]. Em seguida, na Etapa 3 são implementadas as estratégias de intervenção nos conjuntos de dados com o objetivo de melhorar a capacidade de generalização dos modelos. A Tabela 2 apresenta um resumo das principais diferenças entre as intervenções utilizadas. Já na Etapa 4, realiza-se a seleção e aplicação dos algoritmos que definem o aprendizado dos modelos. A escolha desses algoritmos depende da natureza dos dados e dos requisitos específicos da tarefa de detecção de intrusão [Mahesh 2020]. Por sua vez, na Etapa 5, realizam-se experimentos em diferentes cenários de validação, nos quais as intervenções propostas são testadas. Esses cenários estão ilustrados na Figura 2.

No cenário *Intraset*, por exemplo, o subconjunto A1 é inicialmente submetido à primeira estratégia de intervenção. Em seguida, o subconjunto A1, já processado com a técnica implementada, é utilizado para treinar um algoritmo. Posteriormente, o modelo treinado é testado utilizando os demais subconjuntos (A2, A3, A4 e A5), com o objetivo de validar seu desempenho dentro do mesmo domínio. Já no cenário *Interset*,

Tabela 2. Comparação das Intervenções Utilizadas.

Aspecto	Redução de Dimensão	Seleção de Atributos
Definição	Simplificação dos dados por meio da criação de novos atributos derivados dos originais.	Identificação e seleção dos atributos mais relevantes do conjunto de dados original.
Método	Cria uma nova matriz de atributos reduzidos.	Mantém os atributos originais, descartando os irrelevantes ou redundantes.
Foco	Transformação dos dados.	Seleção direta de atributos existentes.
Objetivo	Remover redundância e irrelevância para melhorar a robustez e eficiência do modelo.	Reduzir a complexidade focando nos atributos mais significativos.
Impacto nos dados	Gera novos atributos que podem ser combinações ou transformações dos originais.	Preserva os atributos originais considerados relevantes.
Exemplo de aplicação	Principal Component Analysis (PCA).	Métodos como Chi-Square, Recursive Feature Elimination (RFE).

Fonte: Elaborado pelo autor.

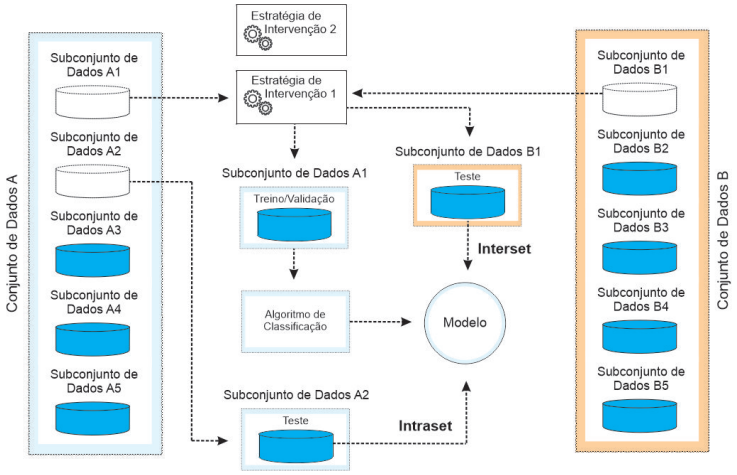

Figura 2. Cenários de Validação.
Fonte: Elaborado pelo autor.

o mesmo modelo treinado é testado utilizando subconjuntos (B1, B2, B3, B4 e B5) de outro conjunto de dados, previamente submetido à mesma estratégia de intervenção. Essa abordagem busca validar a capacidade de generalização do modelo em um domínio de dados diferente. Esse fluxo de trabalho é repetido até que todos os subconjuntos de A e B sejam utilizados como dados de treinamento e teste, respectivamente. Por fim, na Etapa 6, realiza-se a análise de desempenho utilizando métricas de avaliação. Essa avaliação não valida apenas a confiabilidade dos modelos, mas evidencia o impacto das estratégias de intervenção utilizadas na melhoria da generalização.

4. Experimentos e Resultados

Os experimentos foram conduzidos de acordo com o fluxo de trabalho ilustrado na Figura 1, apresentada na Seção 3, que detalha uma sequência de etapas. A execução dessas etapas e os resultados obtidos são apresentados nas subseções a seguir.

4.1. Conjuntos de Dados e Pré-Processamento

Foram selecionados três conjuntos de dados de referência na literatura: *CIC-IDS2018*, *CIC-IDS2017* [Sharafaldin et al. 2018] e *UNSW-NB15* [Moustafa and Slay 2015]. A seleção baseou-se nos critérios recomendados por Viegas, Santin e Oliveira [Viegas et al. 2017], que sugerem que conjuntos de dados para *IDS* atendam aos seguintes requisitos: (i) conter pacotes ou fluxos que permitam a classificação dos eventos como

normais ou maliciosos; (ii) representar as características reais de cenários de rede; (iii) possuir uma estrutura flexível que facilite sua atualização ou reprodução, permitindo comparações com outros conjuntos de dados; (iv) garantir a ausência de dados confidenciais que possam restringir seu compartilhamento; e (v) disponibilizar documentação detalhada, informando as limitações e os métodos empregados em sua construção.

Os conjuntos de dados selecionados são disponibilizados em nível de pacotes no formato *PCAP*. A partir desses arquivos, foi possível processar e extrair informações detalhadas sobre os fluxos de rede utilizando a ferramenta *NFStream* [Aouini and Pekar 2022]. Essa ferramenta permitiu estruturar e padronizar os atributos dos conjuntos de dados, possibilitando o teste de generalização dos modelos, mesmo em cenários distintos de tráfego de rede. O detalhamento de todos os atributos extraídos e a estruturação dos conjuntos de dados estão disponíveis como artefatos deste artigo, conforme descrito na Seção 5. Cada conjunto foi submetido a um pré-processamento que incluiu a remoção de valores nulos e duplicados, codificação de valores categóricos e rotulagem manual das instâncias de classe, seguindo os requisitos detalhados em Santos et al. [Santos et al. 2024]. Especificamente, os fluxos maliciosos de cada conjunto foram organizados em categorias de ataque, facilitando a análise das instâncias e a comparação entre diferentes tipos de ameaças. Para simplificar a identificação, os conjuntos modificados foram denominados *GenIDS-CIC18*, *GenIDS-CIC17* e *GenIDS-NB15*.

Os conjuntos de dados foram divididos em cinco subconjuntos para garantir uma validação cruzada. Essa divisão foi realizada com base na quantidade de instâncias de cada categoria de ataque, assegurando que a proporção de ataques fosse mantida próxima àquela observada nos conjuntos originais. O percentual de cada subconjunto consistiu em 80% de fluxos normais e 20% maliciosos, garantindo a representação de todas as categorias de ataque. Os conjuntos de dados modificados e os detalhes sobre sua estruturação estão disponibilizados como artefatos deste artigo, conforme descrito na Seção 5.

4.2. Estratégias de Intervenção

A execução e os resultados das estratégias de intervenção implementadas nos conjuntos de dados para melhorar a capacidade de generalização e facilitar o reconhecimento de padrões pelos modelos de detecção de intrusão são descritas nas seguintes subseções.

4.2.1. Redução de Dimensão

A Redução de Dimensão foi realizada com Análise de Componentes Principais *(PCA)* para processar os atributos dos conjuntos dados. Conforme a Seção 2, entre os poucos estudos que aplicaram intervenções para melhorar a generalização, o *PCA* foi o mais utilizado, justificando sua escolha.

O *PCA* reduziu o número de atributos em cada subconjunto do *GenIDS-CIC18* e *GenIDS-CIC17*, de 70 para 32, o que corresponde a uma redução de aproximadamente 54% no total de atributos. Além disso, a média da soma dos pesos atribuídos aos componentes principais em cada subconjunto foi de aproximadamente 99%. Esse resultado demonstra que as características mais significativas dos dados originais foram preservadas, garantindo a retenção da maior parte da variância presente nos dados. Já no *GenIDS-NB15*, o *PCA* reduziu o número de atributos de 70 para 28, o que corresponde a uma

redução de aproximadamente 60% no total de atributos em cada subconjunto. Além disso, a média da soma dos pesos atribuídos aos componentes principais alcançou cerca de 99%. Dessa forma, o *PCA* assegurou a preservação da maior parte da variância presente nos dados originais em todos os conjuntos analisados. As tabelas com os pesos de significância e as figuras dos mapas de calor resultantes da intervenção de Redução de Dimensão são disponibilizadas como artefatos deste artigo (ver Seção 5).

4.2.2. Seleção de Atributos

Na Seleção de Atributos, utilizou-se o método *Chi-square* [Thaseen et al. 2019]. Esse método permitiu identificar os atributos que mais contribuem para a classificação dos dados. Assim, observou-se que as diferenças nos pesos de significância estatística dos atributos tornam-se menos expressivas nos subconjuntos de cada conjunto de dados, a partir dos seguintes percentuais de seleção:

Nos conjuntos *GenIDS-CIC18* e *GenIDS-NB15*, foram selecionados 32 atributos (46% do total), enquanto no *GenIDS-CIC17*, 40 atributos (58% do total). Esses percentuais mostram a variação dos pesos de significância entre os subconjuntos. A análise desses pesos permitiu identificar os atributos mais relevantes para a classificação dos fluxos, utilizados no treinamento dos modelos. As tabelas e figuras da intervenção de Seleção de Atributos estão disponíveis como artefatos deste artigo (Seção 5).

4.3. Algoritmos de Aprendizado de Máquina

Neste trabalho, foram empregados quatro algoritmos supervisionados de aprendizado de máquina. Esses algoritmos são comumente descritos na literatura sobre modelos de detecção de intrusão, entre os quais incluem: *Random Forest*, baseado em árvores; *kNN*, baseado em vizinhança; *XGBoost*, um método baseado em *ensemble*; e *MLP (ANN)*, que utiliza aprendizado profundo *(deep learning)*.

Para a criação dos modelos, foi necessário particionar os dados em subconjuntos de treino e teste. Na literatura, não foram identificadas regras definidas para esse particionamento. Contudo, algumas recomendações podem ser encontradas para otimizar os dados durante o treinamento de modelos. Por exemplo, em Obaid et al. [Obaid et al. 2019] e Paulauskas et al. [Paulauskas and Auskalnis 2017], são sugeridas proporções como 80/20, 85/15, 70/30 ou 75/25 para os subconjuntos de treino e teste, respectivamente. Neste trabalho, foi adotada a divisão de 80/20 nos dados de treino e teste.

4.4. Análise das Questões de Pesquisa

A validação dos experimentos em diferentes cenários e a avaliação de desempenho dos modelos evidenciaram o impacto das intervenções na generalização, reduzindo a complexidade dos dados e aprimorando a identificação de padrões. Assim, os resultados ajudam a responder às questões de pesquisa (QP1 e QP2).

4.4.1. Cenários de Validação e Avaliação de Desempenho

Os experimentos foram realizados em dois cenários distintos para avaliar o impacto das estratégias de intervenção nos modelos. Cada intervenção foi aplicada a um sub-

conjunto específico dos conjuntos de dados *GenIDS-CIC18*, *GenIDS-CIC17* e *GenIDS-NB15*. Após a aplicação da intervenção, o subconjunto foi utilizado para treinar o modelo. Em seguida, o modelo treinado foi testado em cenários de domínio único *(Intraset)* e de múltiplos domínios de dados *(Interset)* para validação de desempenho.

Devido à limitação de páginas deste artigo, serão apresentados apenas os resultados da avaliação de desempenho do *GenIDS-CIC18*. Assim, no cenário *Intraset*, os modelos foram treinados com um subconjunto específico do *GenIDS-CIC18* e testados com os demais subconjuntos do mesmo conjunto. Por outro lado, no cenário *Interset*, os modelos foram treinados com um subconjunto específico do *GenIDS-CIC18* e testados com subconjuntos do *GenIDS-CIC17* e *GenIDS-NB15*.

Ambos os cenários são ilustrados na Figura 2 da Seção 3. Por exemplo, no cenário *Intraset*, o subconjunto A1 do *GenIDS-CIC18* foi inicialmente submetido à intervenção de Redução de Dimensão. Em seguida, o subconjunto A1, já processado com a técnica implementada, foi utilizado para treinar o modelo *kNN*. Posteriormente, o modelo treinado foi testado utilizando os demais subconjuntos (A2, A3, A4 e A5) do *GenIDS-CIC18*, a fim de validar seu desempenho dentro do mesmo domínio. Já no cenário *Interset*, o mesmo modelo treinado foi submetido a testes utilizando subconjuntos de outros dois conjuntos de dados (*GenIDS-CIC17* e *GenIDS-NB15*), previamente submetidos à mesma intervenção, com o objetivo de validar a capacidade de generalização do modelo em diferentes domínios de dados. Esse fluxo de trabalho foi repetido até que todos os subconjuntos do *GenIDS-CIC18* fossem utilizados como dados de treinamento. A implementação da intervenção de Seleção de Atributos seguiu o mesmo fluxo de trabalho.

4.4.2. Linha de Base

Para responder às questões de pesquisa (QP1 e QP2), foi necessário estabelecer uma linha de base para analisar o desempenho dos modelos antes da implementação das intervenções. Os experimentos foram conduzidos de acordo com os cenários descritos anteriormente, cujos resultados estão apresentados nas Tabelas 3 e 4.

Tabela 3. Cenário Intraset Sem a Implementação das Intervenções.

Conjunto de Dados	Algoritmos	Fluxos	Treino/Validação				Cenário Intraset			
			Precisão	SDEV* (+/−)	Revocação	SDEV* (+/−)	Precisão	SDEV* (+/−)	Revocação	SDEV* (+/−)
GenIDS-CIC18	Random Forest	Normal	95,30	0,26	99,02	0,29	95,18	0,27	99,00	0,32
		Malicioso	95,26	1,24	80,03	1,19	95,25	1,36	79,93	1,24
	kNN	Normal	95,47	0,14	96,74	0,32	95,25	0,19	96,85	0,27
		Malicioso	85,92	1,15	81,23	0,63	86,50	0,90	80,70	0,84
	XGBoost	Normal	96,02	0,10	97,55	0,17	95,91	0,15	97,50	0,12
		Malicioso	89,29	0,66	83,48	0,44	89,27	0,45	83,38	0,66
	MLP (ANN)	Normal	94,53	0,89	98,69	1,21	94,31	0,91	98,64	1,22
		Malicioso	93,87	4,75	76,58	4,20	93,77	4,72	76,11	4,24

*Desvio Padrão. Fonte: Elaborado pelo autor.

No cenário *Intraset*, os modelos alcançaram resultados próximos aos obtidos durante os treinamentos. Além disso, foram observadas variações nas métricas de Precisão e Revocação, tanto para a classificação de fluxos normais quanto para fluxos maliciosos. Por outro lado, no cenário *Interset*, os resultados indicam que os modelos apresentaram desempenhos variáveis, com valores distintos dependendo do conjunto de teste, do tipo de

Tabela 4. Cenário Interset Sem a Implementação das Intervenções.

Algoritmos	Fluxos	Treino/Validação (GenIDS-CIC18)				Cenário Interset							
						Teste (GenIDS-NB15)				Teste (GenIDS-CIC17)			
		Precisão	SDEV* (+/−)	Revocação	SDEV* (+/−)	Precisão	SDEV* (+/−)	Revocação	SDEV* (+/−)	Precisão	SDEV* (+/−)	Revocação	SDEV* (+/−)
Random Forest	Normal	95,30	0,26	99,02	0,29	79,36	0,13	96,15	0,76	81,16	0,53	99,21	0,27
	Malicioso	95,26	1,24	80,03	1,19	0,05	0,10	0,01	0,02	69,81	10,63	7,84	3,34
kNN	Normal	95,47	0,14	96,74	0,32	77,84	1,22	86,05	4,66	81,13	0,05	98,55	0,13
	Malicioso	85,92	1,15	81,23	0,63	4,43	4,32	2,19	2,05	59,08	2,40	8,33	0,21
XGBoost	Normal	96,02	0,10	97,55	0,17	79,20	0,49	93,48	1,40	82,48	1,02	98,78	0,41
	Malicioso	89,29	0,66	83,48	0,44	6,56	7,18	1,80	2,10	75,03	12,52	16,03	5,73
MLP (ANN)	Normal	94,53	0,89	98,69	1,21	77,34	0,91	79,30	3,87	82,42	1,14	98,25	0,59
	Malicioso	93,87	4,75	76,58	4,20	7,87	3,55	7,15	3,72	67,12	7,88	15,00	6,51

*Desvio Padrão. Fonte: Elaborado pelo autor.

fluxo (normal ou malicioso) e da métrica avaliada. Em ambos os cenários, os resultados fornecem uma base inicial sobre o potencial de generalização de cada modelo, servindo como referência para avaliar o impacto das intervenções realizadas.

4.4.3. Comparação das Intervenções

No cenário *Intraset* da linha de base (ver Tabela 3), os resultados obtidos demonstraram um bom desempenho dos modelos na classificação de fluxos normais e maliciosos. Assim, a análise dos impactos das intervenções na capacidade de generalização dos modelos foi direcionada exclusivamente para o cenário *Interset*. Assim, a Tabela 5 apresenta os resultados obtidos na linha de base com aqueles obtidos após a implementação das intervenções, utilizando o conjunto de dados *GenIDS-CIC18* para treino e os conjuntos *GenIDS-NB15* e *GenIDS-CIC17* para teste. Os índices destacados em azul indicam uma melhoria na generalização, em relação aos destacados em vermelho.

Tabela 5. Avaliação de Desempenho das Intervenções.

Modelo	Intervenção	Fluxo	Atributos**	Cenário Interset							
				GenIDS-NB15				GenIDS-CIC17			
				Precisão	SDEV*	Revocação	SDEV*	Precisão	SDEV*	Revocação	SDEV*
Random Forest	Baseline	Normal	70	79,36	0,13	96,15	0,76	81,16	0,53	99,21	0,27
		Malicioso		0,05	0,10	0,01	0,02	69,81	10,63	7,84	3,34
	Redução de Dimensão	Normal	32	79,92	0,13	98,98	0,81	81,84	0,87	99,74	0,15
		Malicioso		10,76	14,78	0,54	0,57	91,89	1,48	11,33	5,29
	Seleção de Atributos	Normal	32	79,54	0,45	96,07	0,99	81,82	1,19	99,25	0,13
		Malicioso		5,85	12,01	1,14	2,36	76,02	9,44	11,74	7,04
kNN	Baseline	Normal	70	77,84	1,22	86,05	4,66	81,13	0,05	98,55	0,13
		Malicioso		4,43	4,32	2,19	2,05	59,08	2,40	8,33	0,21
	Redução de Dimensão	Normal	32	79,67	0,44	94,11	3,17	81,25	0,38	98,24	0,86
		Malicioso		13,88	9,79	3,99	3,21	58,07	11,49	9,32	2,01
	Seleção de Atributos	Normal	32	79,54	1,34	93,12	2,55	81,23	0,20	98,79	0,33
		Malicioso		12,62	19,88	4,19	6,35	64,57	6,38	8,68	1,17
XGBoost	Baseline	Normal	70	79,20	0,49	93,48	1,40	82,48	1,02	98,78	0,41
		Malicioso		6,56	7,18	1,80	2,10	75,03	12,52	16,03	5,73
	Redução de Dimensão	Normal	32	79,81	0,23	96,37	1,62	84,58	3,23	99,08	0,40
		Malicioso		13,35	6,77	2,51	1,75	84,81	7,82	27,16	18,16
	Seleção de Atributos	Normal	32	79,71	0,50	94,57	0,97	82,79	0,23	98,84	0,27
		Malicioso		14,23	8,23	3,68	2,55	79,49	2,76	17,82	1,46
ANN	Baseline	Normal	70	77,34	0,91	79,30	3,87	82,42	1,14	98,25	0,59
		Malicioso		7,87	3,55	7,15	3,72	67,12	7,88	15,00	6,51
	Redução de Dimensão	Normal	32	77,55	2,03	82,56	10,79	81,46	0,35	98,29	0,91
		Malicioso		8,55	7,52	5,17	3,65	62,34	11,83	10,49	2,21
	Seleção de Atributos	Normal	32	78,93	1,88	91,59	9,16	81,16	0,11	98,32	0,66
		Malicioso		10,55	11,13	2,86	3,00	57,93	10,51	8,71	0,41

*Desvio Padrão. **Atributos: Baseline 100%, Redução de Dimensão ∼54%, Seleção de Atributos ∼46%. Fonte: Elaborado pelo autor.

Para a análise comparativa, foram considerados relevantes apenas os índices que apresentaram diferenças iguais ou superiores a 5% entre a linha de base e a implementação

de cada intervenção. Dessa forma, os resultados indicam que as intervenções geram variações nos índices, dependendo do conjunto de teste (*GenIDS-NB15* ou *GenIDS-CIC17*), do tipo de fluxo (normal ou malicioso) e do modelo *(Random Forest, kNN, XGBoost ou MLP (ANN))*, tanto na métrica de Precisão quanto na Revocação. Para facilitar a visualização e compreensão dos principais achados, a Tabela 6 sintetiza os impactos mais relevantes das intervenções na Precisão e Revocação dos fluxos maliciosos.

Tabela 6. Sumarização dos Impactos das Intervenções na Precisão e Revocação dos Fluxos Maliciosos.

Conjunto de Teste	Modelo	Intervenção	Impacto na Precisão	Diferença entre as internvenções na Precisão	Impacto na Revocação	Diferença entre as internvenções na Revocação	Observação
GenIDS-NB15	Random Forest	Redução de Dimensão Seleção de Atributos	+10,71% +5,80%	∼5%	Sem impacto Sem impacto	-	Redução de Dimensão com maior impacto na Precisão. Intervenções sem impactos na Revocação.
	kNN	Redução de Dimensão Seleção de Atributos	+9,45% +8,19%	∼1%	Sem impacto Sem impacto	-	Intervenções com impactos similares na Precisão. Intervenções sem impactos na Revocação.
	XGBoot	Redução de Dimensão Seleção de Atributos	+6,79% +7,67%	∼1%	Sem impacto Sem impacto	-	Intervenções com impactos similares na Precisão. Intervenções sem impactos na Revocação.
	MLP (ANN)	Redução de Dimensão Seleção de Atributos	Sem impacto Sem impacto	-	Sem impacto Sem impacto	-	Intervenções sem impactos na Precisão. Intervenções sem impactos na Revocação.
GenIDS-CIC17	Random Forest	Redução de Dimensão Seleção de Atributos	+22,08% +6,21%	∼15%	Sem impacto Sem impacto	-	Redução de Dimensão com maior impacto na Precisão. Intervenções sem impactos na Revocação.
	kNN	Redução de Dimensão Seleção de Atributos	Sem impacto +5,49%	∼6%	Sem impacto Sem impacto	-	Seleção de Atributos com maior impacto na Precisão. Intervenções sem impactos na Revocação.
	XGBoot	Redução de Dimensão Seleção de Atributos	+9,78% Sem impacto	∼5%	+11,13% Sem impacto	∼9%	Redução de Dimensão com maior impacto na Precisão. Redução de Dimensão com maior impacto na Revocação.
	MLP (ANN)	Redução de Dimensão Seleção de Atributos	Sem impacto -9,19%	∼4%	Sem impacto -6,29%	∼2%	Seleção de Atributos com impacto negativo na Precisão. Seleção de Atributos com impacto negativo na Revocação.

Fonte: Elaborado pelo autor.

Na Tabela 6, é possível identificar de forma mais clara as tendências observadas em relação às intervenções. A análise revela que, em certos cenários, a Redução de Dimensão demonstrou maior eficácia em comparação à Seleção de Atributos, especialmente na Precisão. Além disso, os impactos positivos e negativos das intervenções podem ser observados para cada modelo avaliado. A tabela apresenta as diferenças nos índices de Precisão e Revocação, evidenciando como essas variações impactaram na classificação dos fluxos maliciosos em relação à linha de base. Por fim, a comparação entre as intervenções permite visualizar de maneira mais detalhada quais técnicas apresentaram maior impacto. A seguir, é apresentada a Tabela 7, que sintetiza os impactos mais relevantes das intervenções na Precisão e Revocação dos fluxos normais.

Tabela 7. Sumarização dos Impactos das Intervenções na Precisão e Revocação dos Fluxos Normais.

Conjunto de Teste	Modelo	Intervenção	Impacto na Precisão	Diferença entre as internvenções na Precisão	Impacto na Revocação	Diferença entre as internvenções na Revocação	Observação
GenIDS-NB15	Random Forest	Redução de Dimensão Seleção de Atributos	Sem impacto Sem impacto	-	Sem impacto Sem impacto	-	Intervenções sem impactos na Precisão. Intervenções sem impactos na Revocação.
	kNN	Redução de Dimensão Seleção de Atributos	Sem impacto Sem impacto	-	+8,06% +7,07%	∼1%	Intervenções sem impactos na Precisão. Intervenções com impactos similares na Revocação.
	XGBoot	Redução de Dimensão Seleção de Atributos	Sem impacto Sem impacto	-	Sem impacto Sem impacto	-	Intervenções sem impactos na Precisão. Intervenções sem impactos na Revocação.
	MLP (ANN)	Redução de Dimensão Seleção de Atributos	Sem impacto Sem impacto	-	Sem impacto +12,29%	∼9%	Intervenções sem impactos na Precisão. Seleção de Atributos com maior impacto na Revocação.
GenIDS-CIC17	Random Forest	Redução de Dimensão Seleção de Atributos	Sem impacto Sem impacto	-	Sem impacto Sem impacto	-	Intervenções sem impactos na Precisão. Intervenções sem impactos na Revocação.
	kNN	Redução de Dimensão Seleção de Atributos	Sem impacto Sem impacto	-	Sem impacto Sem impacto	-	Intervenções sem impactos na Precisão. Intervenções sem impactos na Revocação.
	XGBoot	Redução de Dimensão Seleção de Atributos	Sem impacto Sem impacto	-	Sem impacto Sem impacto	-	Intervenções sem impactos na Precisão. Intervenções sem impactos na Revocação.
	MLP (ANN)	Redução de Dimensão Seleção de Atributos	Sem impacto Sem impacto	-	Sem impacto Sem impacto	-	Intervenções sem impactos na Precisão. Intervenções sem impactos na Revocação.

Fonte: Elaborado pelo autor.

Diferentemente dos fluxos maliciosos, na Tabela 7, observa-se que na maioria dos cenários, as intervenções não geraram alterações significativas na Precisão, independentemente do modelo avaliado. No entanto, em alguns casos, a Revocação foi impactada,

como no *kNN* e no *MLP (ANN)*, onde a Seleção de Atributos resultou em um aumento considerável na Revocação. Além disso, a Redução de Dimensão mostrou impacto positivo na Revocação do *kNN*, mas sem influenciar a Precisão. De maneira geral, a tabela evidencia que os efeitos das intervenções nos fluxos normais foram menos expressivos em comparação aos fluxos maliciosos, reforçando a importância de uma análise diferenciada entre esses dois tipos de tráfego.

4.4.4. Avaliação da QP1

Sim, tanto a Redução de Dimensão quanto a Seleção de Atributos contribuem para a melhoria da generalização dos modelos de detecção de intrusão, mas seus impactos variam conforme o tipo de tráfego. Nos fluxos maliciosos (ver Tabela 6), a Redução de Dimensão demonstrou maior eficácia, especialmente na Precisão, enquanto nos fluxos normais (ver Tabela 7), os efeitos das intervenções foram menos expressivos. No entanto, em alguns modelos específicos, como *kNN* e *MLP (ANN)*, a Seleção de Atributos resultou em um aumento na Revocação. Isso sugere que essas intervenções podem ser úteis para aprimorar a generalização, dependendo do modelo e do cenário de tráfego.

4.4.5. Avaliação da QP2

A Redução de Dimensão teve um impacto mais significativo na Precisão dos fluxos maliciosos, enquanto a Seleção de Atributos influenciou principalmente a Revocação em alguns modelos nos fluxos normais. Assim, em termos de impacto geral, a Redução de Dimensão parece ter maior influência na melhoria da generalização, especialmente para a detecção de fluxos maliciosos. No entanto, a Seleção de Atributos também pode ser vantajosa em determinados cenários, como na melhoria da Revocação para modelos como *kNN* e *MLP (ANN)*. Dessa forma, a escolha entre as duas intervenções deve considerar o objetivo principal da detecção, seja maximizar a Precisão ou otimizar a Revocação em diferentes cenários de tráfego de rede.

5. Conclusão e Trabalhos Futuros

Este trabalho analisou intervenções voltadas para a melhoria da generalização de modelos de detecção de intrusão, comparando o impacto da Redução de Dimensão e da Seleção de Atributos em diferentes cenários de tráfego de rede. Os resultados evidenciaram que a Redução de Dimensão teve um impacto mais expressivo na Precisão dos fluxos maliciosos, especialmente em alguns modelos como o *Random Forest*. Já a Seleção de Atributos influenciou a Revocação em determinados casos, como no *kNN* e no *MLP (ANN)*, principalmente nos fluxos normais. No entanto, o impacto dessas intervenções variou conforme o modelo e a métrica analisada, reforçando a necessidade de considerar diferentes abordagens para otimizar a generalização dos modelos.

Além disso, a estruturação dos atributos foi essencial para garantir a uniformidade dos conjuntos de dados utilizados, permitindo a avaliação de desempenho dos modelos em múltiplos domínios. Os achados reforçam que a escolha da intervenção deve estar alinhada ao objetivo principal da detecção – seja melhorar a Precisão na identificação de fluxos maliciosos ou otimizar a Revocação para minimizar falsos negativos.

Os resultados obtidos abrem caminho para novos experimentos, incluindo a aplicação da metodologia com outros algoritmos e conjuntos de dados. Para trabalhos futuros, propõe-se a investigação de novas técnicas de Redução de Dimensão e Seleção de Atributos, visando melhorar a capacidade de generalização de detecção de intrusão.

Disponibilidade de Artefatos

Todas as informações complementares deste artigo estão disponíveis em um repositório no *GitHub*, acessível pelo *link*: https://github.com/kelsonc/paper-sbrc2025.

Referências

Aouini, Z. and Pekar, A. (2022). Nfstream: A flexible network data analysis framework. *Computer Networks*, 204:1–8.

Cieslak, M. C., Castelfranco, A. M., Roncalli, V., Lenz, P. H., and Hartline, D. K. (2020). t-distributed stochastic neighbor embedding (t-sne): A tool for eco-physiological transcriptomic analysis. *Marine genomics*, 51:1–12.

D'hooge, L., Verkerken, M., Wauters, T., De Turck, F., and Volckaert, B. (2023). Investigating generalized performance of data-constrained supervised machine learning models on novel, related samples in intrusion detection. *Sensors*, 23(4):1–39.

D'hooge, L., Wauters, T., Volckaert, B., and De Turck, F. (2020). Inter-dataset generalization strength of supervised machine learning methods for intrusion detection. *Journal of Information Security and Applications*, 54:1–13.

Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):1–16.

Kenyon, A., Deka, L., and D., E. (2020). Are public intrusion datasets fit for purpose characterising the state of the art in intrusion event datasets. *Computers & Security*, 99:1–26.

Khalid, S., Khalil, T., and Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. Science and Information Conference, London, UK, 27-29 Aug 2014.

Khraisat, A., Gondal, I., Vamplew, P., and Kamruzzaman, J. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity*, 2(20):1–22.

Layeghy, S. and Portmann, M. (2022). On generalisability of machine learning-based network intrusion detection systems. *arXiv preprint arXiv:2205.04112*, [*s.n.*]:1–12.

Mahesh, B. (2020). Machine learning algorithms - a review. *International Journal of Science and Research (IJSR)*, 9(1):381–386.

Mahfouz, A., Abuhussein, A., Venugopal, D., and Shiva, S. (2020). Ensemble classifiers for network intrusion detection using a novel network attack dataset. *Future Internet*, 12(11):1–19.

Marvi, M., Arfeen, A., and Uddin, R. (2021). A generalized machine learning-based model for the detection of ddos attacks. *International Journal of Network Management*, 31(6):1–22.

Moustafa, N. and Slay, J. (2015). Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). Military Communications and Information Systems Conference (MilCIS), Canberra, ACT, Australia, 10-12 Nov 2015.

Naseer, M., Rusdi, J. F., Shanono, N. M., Salam, S., Muslim, Z. B., Abu, N. A., and Abadi, I. (2021). Malware detection: issues and challenges. *Cybersecurity*, 1807(1):1–6.

Obaid, H. S., Dheyab, S. A., and Sabry, S. S. (2019). The impact of data pre-processing techniques and dimensionality reduction on the accuracy of machine learning. 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON), Jaipur, India, 13-15 Mar 2019.

Paulauskas, N. and Auskalnis, J. (2017). Analysis of data pre-processing influence on intrusion detection using nsl-kdd dataset.

Rocha, M. S., Bernardo, G. D., Mundim, L., Zarpelão, B. B., and Miani, R. S. (2023). Supervised machine learning and detection of unknown attacks: An empirical evaluation. AINA 2023: International Conference on Advanced Information Networking and Applications, Juiz de Fora, MG, Brazil, 29-31 Mar 2023.

Santos, K. C., Miani, R. S., and de Oliveira Silva, F. (2024). Evaluating the impact of data preprocessing techniques on the performance of intrusion detection systems. *Journal of Network and Systems Management*, 32(36):1–54.

Sharafaldin, I., Lashkari, A. H., and Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. 4th International Conference on Information Systems Security and Privacy (ICISSp), Funchal, Madeira, Portugal, 22-24 Jan 2018.

Sudyana, D., Lin, Y., Verkerken, M., Hwang, R., Lai, Y., D'Hooge, L., Wauters, T., Volckaert, B., and De Turck, F. (2024). Improving generalization of ml-based ids with lifecycle-based dataset, auto-learning features, and deep learning. *IEEE Transactions on Machine Learning in Communications and Networking*, 2:645–662.

Thaseen, I. S., Kumar, C. A., and Ahmad, A. (2019). Integrated intrusion detection model using chi-square feature selection and ensemble of classifiers. *Arabian Journal for Science and Engineering*, 44:3357–3368.

Verkerken, M., D'hooge, L., Wauters, T., Volckaert, B., and De Turck, F. (2021). Towards model generalization for intrusion detection: Unsupervised machine learning techniques. *Journal of Network and Systems Management*, 30(1):1–25.

Viegas, E. K., Santin, A. O., and Oliveira, L. S. (2017). Toward a reliable anomaly-based intrusion detection in real-world environments. *Computer Networks*, 127:200–216.

Yudha, F. (2023). Cremev2: A toolchain of automatic dataset collection for machine learning in intrusion detection based on mitre att&ck. Disponível em: https://github.com/masjohncook/CREMEv2. Accessado em 29 Jun 2024.

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., and Saeed, J. (2020). A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, 1(2):56–70.