Revista ft

www.revistaft.com.br

# Avaliação da eficácia de algoritmos de machine learning na detecção de tráfego malicioso em redes corporativas

## Evaluating the effectiveness of machine learning algorithms in detecting malicious traffic on corporate networks

Cleilson Lopes Monteiro [1]
Viviane dos Santos Almeida [2]
Thiago Amaral Guarnieri [3]
Dhiodines Fabrício Souza da Costa [4]
Carlos Eduardo de Souza Santos[5]
Renato do Nascimento[6]
Cristino Corrêa Jordão[7]

[1] Discente do Curso de Especialização em Redes e Computação Distribuída do Instituto Federal de Ciência e Tecnologia do Mato Grosso, *Campus* Cuiabá, e-mail: cleilsonmontteiro@gmail.com, lattes: https://lattes.cnpq.br/4881886811613821 | ORCID: https://orcid.org/0009-0007-1240-6445
[2] Docente do Curso de Especialização em Redes e Computação Distribuída do Instituto Federal de Ciência e Tecnologia do Mato Grosso, *Campus* Cuiabá, e-mail: vivianealmeida.edu@gmail.com, lattes: https://lattes.cnpq.br/6234559329439208 | ORCID: https://orcid.org/0009-0000-9897-4417
[3] Discente do Curso de Especialização em Redes e Computação Distribuída do Instituto Federal de Ciência e Tecnologia do Mato Grosso, *Campus* Cuiabá, e-mail: thiago.guarnieri@gmail.com, lattes: https://lattes.cnpq.br/4819463147196353 | ORCID: https://orcid.org/0009-0006-6689-7406
[4] Discente do Curso de Especialização em Redes e Computação Distribuída do Instituto Federal de Ciência e Tecnologia do Mato Grosso, *Campus* Cuiabá, e-mail: dh.iodines@hotmail.com, lattes: https://lattes.cnpq.br/9404784729905383 | ORCID: https://orcid.org/0009-0004-4141-0150
[5] Discente do Curso de Especialização em Redes e Computação Distribuída do Instituto Federal de Ciência e Tecnologia do Mato Grosso, *Campus* Cuiabá, e-mail: eduardopires99@gmail.com, lattes: https://lattes.cnpq.br/4081573005681428 | ORCID: https://orcid.org/0009-0003-0521-2504
[6] Discente do Curso de Especialização em Redes e Computação Distribuída do Instituto Federal de Ciência e Tecnologia do Mato Grosso, *Campus* Cuiabá, e-mail: renatodonascimento@live.com, lattes: https://lattes.cnpq.br/9873039236611224 | ORCID: https://orcid.org/0009-0007-4972-3567
[7] Discente do Curso de Especialização em Redes e Computação Distribuída do Instituto Federal de Ciência e Tecnologia do Mato Grosso, *Campus* Cuiabá, e-mail: cristino07jordao@gmail.com, lattes: https://lattes.cnpq.br/1881454363188440 | ORCID: https://orcid.org/0009-0002-7241-4192

## Resumo

A crescente complexidade das ameaças cibernéticas impõe desafios significativos às organizações que dependem de infraestruturas digitais para suas operações. Métodos tradicionais de detecção de intrusão baseados em assinaturas têm se mostrado limitados diante de ataques sofisticados e dinâmicos, especialmente aqueles potencializados por técnicas automatizadas. Nesse contexto, algoritmos de Machine Learning emergem como alternativa promissora para a identificação de padrões anômalos em tráfego de rede. O presente artigo tem como objetivo avaliar a eficácia de algoritmos de Machine Learning na detecção de tráfego malicioso em redes corporativas, por meio de pesquisa bibliográfica qualitativa fundamentada na literatura científica dos últimos vinte anos. A análise contempla estudos comparativos sobre desempenho de modelos supervisionados e não supervisionados, métricas de avaliação e limitações práticas de implementação. Os resultados indicam que modelos baseados em aprendizado supervisionado apresentam alto desempenho preditivo, porém enfrentam desafios relacionados à escalabilidade, explicabilidade e adaptação a ataques do tipo zero-day. Conclui-se que a eficácia desses algoritmos depende não apenas da arquitetura computacional adotada, mas também da maturidade organizacional e da governança de dados corporativa.



**Palavras-chave:** Machine Learning. Tráfego malicioso. Segurança de redes. Redes corporativas. Detecção de intrusão.



## Abstract

The growing complexity of cyber threats imposes significant challenges on organizations that rely on digital infrastructures for their operations. Traditional signature-based intrusion detection methods have proven limited against sophisticated and dynamic attacks, especially those enhanced by automated techniques. In this context, Machine Learning algorithms emerge as a promising alternative for identifying anomalous patterns in network traffic. This article aims to evaluate the effectiveness of Machine Learning algorithms in detecting malicious traffic in corporate networks through a qualitative bibliographic research grounded in scientific literature from the last twenty years. The analysis encompasses comparative studies on the performance of supervised and unsupervised models, evaluation metrics, and practical implementation limitations. Results indicate that models based on supervised learning exhibit high predictive performance; however, they face challenges related to scalability, explainability, and adaptation to zero-day attacks. It is concluded that the effectiveness of these algorithms depends not only on the computational architecture adopted but also on organizational maturity and corporate data governance.



**Keywords:** Machine Learning. Malicious traffic. Network security. Corporate networks. Intrusion detection.


# 1 INTRODUÇÃO

A transformação digital das organizações ampliou significativamente a superfície de ataque das infraestruturas corporativas. Redes empresariais tornaram-se ambientes altamente dinâmicos, caracterizados por grande volume de dados, dispositivos heterogêneos e interconectividade constante. Nesse cenário, a segurança da informação assume papel estratégico, especialmente no que se refere à detecção de tráfego malicioso.

Historicamente, sistemas de detecção de intrusão (*Intrusion Detection Systems – IDS*) foram estruturados com base em assinaturas previamente conhecidas, operando por meio de regras estáticas para identificação de padrões de ataque (STALLINGS, 2017). Embora

eficazes para ameaças catalogadas, tais sistemas apresentam limitações frente a ataques emergentes e técnicas de evasão sofisticadas. Conforme argumentam SOMMER e PAXSON (2010), a dependência exclusiva de assinaturas restringe a capacidade adaptativa dos mecanismos de defesa.

Paralelamente, avanços na área de Inteligência Artificial, especialmente no campo do *Machine Learning*, possibilitaram o desenvolvimento de modelos capazes de aprender padrões complexos a partir de grandes volumes de dados (GOODFELLOW; BENGIO; COURVILLE, 2016). Segundo BUCZAK e GUVEN (2016), técnicas de mineração de dados e aprendizado supervisionado vêm sendo aplicadas com resultados promissores na identificação de anomalias em tráfego de rede.

O uso de algoritmos como *Random Forest, Support Vector Machines* e Redes Neurais Profundas tem sido amplamente explorado na literatura recente (KHAN et al., 2019; VINAYAKUMAR et al., 2019). Esses modelos apresentam elevada taxa de detecção e capacidade de generalização, especialmente em ambientes de alto volume de dados. Contudo, persistem questionamentos quanto à sua eficácia real em redes corporativas, considerando fatores como custo computacional, necessidade de dados rotulados, suscetibilidade a ataques adversariais e desafios de explicabilidade (BIGGIO; ROLI, 2018).

Além disso, o próprio avanço da IA tem sido instrumentalizado para a criação de ataques automatizados, ampliando o desafio da segurança cibernética. Essa dualidade — IA como ferramenta de defesa e como vetor de ataque — torna imperativa a análise crítica da aplicabilidade desses algoritmos no contexto organizacional.

Diante desse panorama, este artigo busca responder à seguinte questão de pesquisa:

Os algoritmos de *Machine Learning* apresentam eficácia consistente na detecção de tráfego malicioso em redes corporativas, considerando os desafios técnicos e organizacionais contemporâneos?

O objetivo geral consiste em avaliar a eficácia de algoritmos de *Machine Learning* na identificação de tráfego malicioso em redes corporativas. Como objetivos específicos, pretende-se:

(a) analisar os principais modelos utilizados na literatura recente;

(b) discutir métricas de desempenho empregadas na avaliação desses modelos;

(c) examinar limitações técnicas e organizacionais associadas à sua implementação.

A relevância da pesquisa reside na necessidade de fundamentar decisões estratégicas de adoção tecnológica em ambientes corporativos. Do ponto de vista teórico, o estudo contribui para o aprofundamento da discussão sobre o estado da arte em detecção de intrusão baseada em aprendizado de máquina. Sob a perspectiva prática, oferece subsídios para gestores de tecnologia da informação quanto à viabilidade e aos riscos associados à implementação desses sistemas.

Assim, o objeto de estudo delimita-se à análise qualitativa da eficácia de algoritmos de *Machine Learning* aplicados à detecção de tráfego malicioso em redes corporativas, a partir da literatura científica publicada nos últimos vinte anos.

# 2 FUNDAMENTAÇÃO TEÓRICA OU REVISÃO DA LITERATURA

## 2.1 Evolução dos Sistemas de Detecção de Intrusão

A segurança de redes corporativas historicamente fundamentou-se em mecanismos de prevenção e detecção baseados em assinaturas. Esses sistemas, conhecidos como Intrusion Detection Systems (IDS), operam por meio da comparação entre padrões de tráfego observados e bases previamente catalogadas de ataques conhecidos (STALLINGS, 2017). Embora eficazes na identificação de ameaças previamente mapeadas, tais sistemas demonstram baixa capacidade adaptativa diante de ataques inéditos ou modificados.

Segundo SOMMER e PAXSON (2010), a principal limitação dos IDS tradicionais reside no chamado “problema do mundo fechado”, no qual assume-se que todas as ameaças possíveis são previamente conhecidas. Em ambientes corporativos dinâmicos, essa premissa mostra-se inviável, especialmente diante da emergência de ataques do tipo zero-day, caracterizados por explorarem vulnerabilidades ainda não documentadas.

A necessidade de superar essas limitações impulsionou a incorporação de técnicas de mineração de dados e aprendizado de máquina na segurança cibernética. Diferentemente dos métodos baseados em regras fixas, algoritmos de Machine Learning são capazes de aprender padrões a partir de dados históricos, identificando comportamentos anômalos mesmo na ausência de assinaturas previamente definidas.

## 2.2 Fundamentos de Machine Learning Aplicado à Segurança de Redes

O campo do Machine Learning insere-se no escopo mais amplo da Inteligência Artificial e pode ser definido como o conjunto de técnicas que permite a sistemas computacionais aprenderem padrões a partir de dados e realizarem previsões ou classificações

sem programação explícita para cada cenário (GOODFELLOW; BENGIO; COURVILLE, 2016).

Na detecção de tráfego malicioso, os modelos de aprendizado podem ser classificados em três categorias principais:

1. Aprendizado supervisionado – utiliza dados previamente rotulados para treinar classificadores capazes de distinguir tráfego benigno de tráfego malicioso.
2. Aprendizado não supervisionado – identifica padrões anômalos sem necessidade de rótulos prévios.
3. Aprendizado semi-supervisionado e híbrido – combina ambas as abordagens.

Conforme BUCZAK e GUVEN (2016), modelos supervisionados como Support Vector Machines (SVM), Random Forest e Redes Neurais Artificiais apresentam elevados índices de acurácia quando aplicados a conjuntos de dados bem estruturados. Já métodos não supervisionados, como k-means e técnicas de detecção de anomalias, mostram-se úteis em contextos onde há escassez de dados rotulados.

Estudos recentes indicam que modelos baseados em Deep Learning ampliaram significativamente a capacidade de detecção em ambientes de grande volume de dados (big data), especialmente por meio de arquiteturas como Redes Neurais Convolucionais e Redes Recorrentes (VINAYAKUMAR et al., 2019). Tais modelos conseguem extrair características complexas de pacotes de rede, reduzindo a dependência de engenharia manual de atributos.

## 2.3 Métricas de Avaliação e Desempenho

A avaliação da eficácia de algoritmos de Machine Learning na detecção de tráfego malicioso requer o uso de métricas estatísticas apropriadas. Entre as principais destacam-se:

Acurácia – proporção de classificações corretas.

Precisão – proporção de verdadeiros positivos em relação às detecções positivas.

Recall (sensibilidade) – capacidade do modelo de identificar efetivamente os ataques.

F1-score – média harmônica entre precisão e recall.

Taxa de falsos positivos – indicador crítico em ambientes corporativos.

KHAN et al. (2019) destacam que a taxa de falsos positivos é um dos principais desafios práticos na adoção de modelos de ML em redes empresariais, pois alertas excessivos podem comprometer a eficiência operacional das equipes de segurança.

Além disso, VERKERKEN et al. (2022) apontam que a escalabilidade e o tempo de processamento são variáveis determinantes em ambientes de alto desempenho, nos quais o tráfego de rede é contínuo e volumoso. Assim, a eficácia não deve ser medida apenas em termos estatísticos, mas também sob perspectiva operacional.

### 2.4 Desafios Contemporâneos: Ataques Adversariais e Robustez dos Modelos

Embora os algoritmos de ML apresentem desempenho elevado em ambientes controlados, sua robustez em cenários reais ainda é objeto de debate. BIGGIO e ROLI (2018) demonstram que modelos de aprendizado podem ser manipulados por meio de ataques adversariais, nos quais pequenas perturbações nos dados de entrada induzem classificações incorretas.

Esse fenômeno é particularmente relevante na segurança de redes, pois agentes maliciosos podem adaptar seus comportamentos para contornar mecanismos automatizados de detecção. Tal dinâmica gera um ciclo contínuo de adaptação entre atacantes e sistemas defensivos.

Além disso, a questão da explicabilidade dos modelos (*explainable AI)* emerge como preocupação central em ambientes corporativos. Modelos complexos de Deep Learning frequentemente operam como “caixas-pretas”, dificultando a interpretação das decisões tomadas pelo sistema. Essa limitação pode impactar requisitos de governança, auditoria e conformidade regulatória.

### 2.5 Viabilidade Organizacional e Implementação em Redes Corporativas

A adoção de algoritmos de ML em redes corporativas não depende exclusivamente de desempenho técnico. Aspectos organizacionais, como maturidade digital, infraestrutura computacional e governança de dados, influenciam diretamente a eficácia prática da implementação.

SOMMER e PAXSON (2010) argumentam que muitos estudos acadêmicos utilizam bases de dados experimentais que não refletem a complexidade do tráfego real corporativo. Já BUCZAK e GUVEN (2016) ressaltam a necessidade de integração entre especialistas em segurança e cientistas de dados para maximizar o potencial dessas ferramentas.

Outro fator relevante refere-se à atualização constante dos modelos. Redes corporativas são ambientes dinâmicos, nos quais padrões legítimos de tráfego podem mudar rapidamente. Modelos estáticos tendem a perder desempenho ao longo do tempo, exigindo mecanismos de re-treinamento contínuo.

Nesse contexto, observa-se tendência ao uso de arquiteturas híbridas, combinando sistemas baseados em assinaturas com modelos de ML, buscando equilibrar previsibilidade, desempenho e adaptabilidade.

### 2.6 Síntese do Estado da Arte

A literatura dos últimos vinte anos converge para o entendimento de que algoritmos de Machine Learning apresentam desempenho superior aos métodos tradicionais na detecção de tráfego malicioso, especialmente em relação a ameaças desconhecidas. Entretanto, desafios relacionados à robustez, explicabilidade, escalabilidade e governança organizacional limitam sua adoção plena.

Assim, a eficácia desses modelos não pode ser analisada apenas sob a perspectiva estatística, mas deve considerar variáveis técnicas e institucionais que condicionam sua implementação em redes corporativas.

A partir desse arcabouço teórico, o próximo tópico apresentará a metodologia adotada neste estudo, explicitando o delineamento da pesquisa bibliográfica qualitativa e os critérios utilizados para análise da literatura selecionada.

# 3 METODOLOGIA

O presente estudo caracteriza-se como uma pesquisa de natureza aplicada, com abordagem qualitativa e procedimento técnico bibliográfico. A pesquisa aplicada justifica-se pelo objetivo de produzir conhecimento com potencial de aplicação prática em ambientes corporativos de segurança da informação. A abordagem qualitativa foi adotada por permitir análise interpretativa e crítica da literatura científica, visando compreender a eficácia dos algoritmos de Machine Learning na detecção de tráfego malicioso sob múltiplas dimensões — técnica e organizacional.

Quanto aos procedimentos, trata-se de pesquisa bibliográfica sistematizada, fundamentada em obras e artigos científicos publicados nos últimos vinte anos, em bases indexadas de reconhecida relevância acadêmica, tais como *IEEE Xplore, ACM Digital Library, ScienceDirect e SpringerLink.*

# 4 RESULTADOS E DISCUSSÕES OU ANÁLISE DOS DADOS

## 4.1 Predominância de Modelos Supervisionados

A análise da literatura evidencia predominância de modelos supervisionados na detecção de tráfego malicioso. Estudos como os de KHAN et al. (2019) e VINAYAKUMAR et al. (2019) indicam que algoritmos como Random Forest, Support Vector Machines e Redes Neurais Profundas apresentam taxas de acurácia superiores a 90% em diversos conjuntos de dados experimentais.

A métrica F1-score mostrou-se particularmente relevante para avaliação equilibrada entre precisão e recall, especialmente em contextos nos quais há desbalanceamento entre tráfego benigno e malicioso.

Entretanto, observa-se que resultados elevados frequentemente estão associados a bases de dados controladas, como CIC-IDS e UNSW-NB15, que não necessariamente refletem a complexidade do tráfego real corporativo, conforme alertado por SOMMER e PAXSON (2010).

## 4.2 Limitações Técnicas Identificadas

Apesar do alto desempenho estatístico, foram identificadas limitações recorrentes como a dependência de dados rotulados de alta qualidade, sensibilidade a mudanças no padrão de tráfego, elevada taxa de falsos positivos em ambientes dinâmicos e a vulnerabilidade a ataques adversariais.

BIGGIO e ROLI (2018) demonstram que modelos podem ser manipulados por meio de pequenas alterações nos dados de entrada, comprometendo sua confiabilidade. Tal vulnerabilidade representa risco relevante para ambientes corporativos, nos quais atacantes adaptativos podem explorar brechas nos modelos de detecção.

Além disso, modelos baseados em Deep Learning demandam elevado poder computacional, o que pode implicar custos significativos de infraestrutura.

## 4.3 Desafios Organizacionais e Operacionais

Do ponto de vista organizacional, a literatura indica que a eficácia prática depende de fatores como a governança de dados, a integração com sistemas legados à capacitação técnica da equipe e a política de atualização contínua dos modelos.

A simples adoção de algoritmos de ML não garante melhoria automática da segurança. A eficácia está condicionada à maturidade digital da organização e à capacidade de interpretar e agir sobre os alertas gerados.

Observa-se tendência à adoção de modelos híbridos, combinando detecção por assinatura e detecção baseada em aprendizado de máquina, buscando equilíbrio entre previsibilidade e adaptabilidade.

### 4.4 Síntese Analítica

A análise dos estudos revisados permite afirmar que os algoritmos de Machine Learning apresentam eficácia estatística elevada em ambientes experimentais, a aplicabilidade prática em redes corporativas depende de variáveis técnicas e organizacionais, modelos híbridos demonstram maior potencial de viabilidade operacional e a robustez contra ataques adversariais ainda representa um desafio significativo.

Esses achados dialogam diretamente com a questão de pesquisa proposta, indicando que a eficácia não pode ser avaliada exclusivamente por métricas quantitativas, mas deve considerar o contexto de implementação.

## 5 CONCLUSÃO/CONSIDERAÇÕES FINAIS

A presente pesquisa analisa a eficácia de algoritmos de Machine Learning na detecção de tráfego malicioso em redes corporativas, considerando dimensões técnicas e organizacionais. Com base na revisão bibliográfica qualitativa realizada, conclui-se que os modelos de aprendizado supervisionado apresentam desempenho estatístico elevado na identificação de padrões maliciosos, especialmente quando avaliados por métricas como acurácia, precisão, recall e F1-score.

Entretanto, a eficácia desses algoritmos não se manifesta de forma absoluta em ambientes corporativos reais. A literatura demonstra que fatores como qualidade dos dados, desbalanceamento de classes, variabilidade do tráfego, suscetibilidade a ataques adversariais e custos computacionais influenciam diretamente o desempenho operacional dos modelos.

Verifica-se que a implementação isolada de técnicas de Machine Learning não garante aumento automático da segurança organizacional. A eficácia depende da integração com políticas de governança de dados, atualização contínua dos modelos, monitoramento especializado e compatibilidade com sistemas legados. Nesse sentido, modelos híbridos, que combinam detecção por assinatura e aprendizado de máquina, apresentam maior potencial de viabilidade prática.

Os objetivos propostos foram atingidos. A análise dos principais algoritmos utilizados na literatura recente permitiu identificar tendências metodológicas predominantes. A discussão das métricas de avaliação evidenciou que a taxa de falsos positivos constitui variável crítica para ambientes corporativos. A investigação das limitações técnicas e organizacionais confirmou que a aplicabilidade desses modelos transcende o desempenho estatístico.

A hipótese de que algoritmos de Machine Learning apresentam desempenho superior aos métodos tradicionais em ambientes experimentais é confirmada. Contudo, a hipótese de que sua eficácia prática depende de variáveis organizacionais também se confirma, indicando que o fator humano e estrutural permanece determinante.

Como limitação do estudo, destaca-se a natureza exclusivamente bibliográfica da pesquisa, que não contempla experimentação empírica própria. Sugere-se, para pesquisas futuras, a realização de estudos de caso em redes corporativas reais, bem como análises quantitativas comparativas com implementação prática de modelos supervisionados e não supervisionados.

Conclui-se que a eficácia de algoritmos de Machine Learning na detecção de tráfego malicioso em redes corporativas existe, porém é condicionada a fatores técnicos, estruturais e estratégicos. A adoção dessas tecnologias deve ser acompanhada de planejamento institucional, investimento em infraestrutura e políticas robustas de governança da informação.

## REFERÊNCIAS


BIGGIO, Battista; ROLI, Fabio. Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition, v. 84, p. 317-331, 2018.

BRUNDAGE, Miles et al. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. Oxford: Future of Humanity Institute, 2018.

BUCZAK, Anna L.; GUVEN, Erhan. A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, v. 18, n. 2, p. 1153-1176, 2016.

CHANDOLA, Varun; BANERJEE, Arindam; KUMAR, Vipin. Anomaly detection: A survey. ACM Computing Surveys, v. 41, n. 3, p. 1-58, 2009.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep learning. Cambridge: MIT Press, 2016.

KHAN, Muhammad Afzal et al. A survey of machine learning-based network intrusion detection systems. IEEE Access, v. 7, p. 70926-70950, 2019.
SIENA, Osmar. **Metodologia da pesquisa científica**: elementos para elaboração e apresentação de trabalhos acadêmicos. Porto Velho: [s.n.], 2007. Disponível em: http://www.mestradoadm.unir.br/site_antigo/doc/manualdetrabalhoacademicoatual.pdf. Acesso em: 10 de janeiro de 2013.

SOMMER, Robin; PAXSON, Vern. Outside the closed world: On using machine learning for network intrusion detection. In: IEEE Symposium on Security and Privacy. Berkeley: IEEE, 2010. p. 305-316.

STALLINGS, William. Network security essentials: Applications and standards. 6. ed. Boston: Pearson, 2017.

VERKERKEN, Dries et al. A survey on distributed machine learning for intrusion detection systems. IEEE Communications Surveys & Tutorials, v. 24, n. 1, p. 501-534, 2022.

VINAYAKUMAR, R. et al. Deep learning approach for intelligent intrusion detection system. IEEE Access, v. 7, p. 41525-41550, 2019.

ZHANG, Jian et al. Network anomaly detection: A survey and comparative analysis of stochastic and deterministic methods. Computer Networks, v. 51, n. 12, p. 3440-3462, 2007.