RECA
REVISTA ELETRÔNICA DE COMPUTAÇÃO APLICADA
Vol. 6 Nro. 1 2025

Departamento de Computação
Uni-FACEF

Aplicações de Inteligência Artificial na Detecção de Intrusões: Um Estudo Comparativo na Base NSL-KDD

Pedro Henrique de Souza
Ciência da computação – Uni-FACEF
pehensouza16@gmail.com

Tulio Fernandes Faria
Ciência da computação – Uni-FACEF
tuliofernandes027@outlook.com.com

Prof. Dr. Daniel Facciolo Pires
Docente do Departamento de Computação – Uni-FACEF
daniel@facef.br

Resumo

Este trabalho investiga o impacto da Inteligência Artificial (IA) na cibersegurança, com ênfase em sua aplicação na detecção e prevenção de intrusões. A pesquisa apresenta uma revisão teórica das principais técnicas de IA empregadas na área, incluindo redes neurais, sistemas especialistas, agentes inteligentes e aprendizado de máquina. Em seguida, realiza-se um estudo experimental com a base de dados NSL-KDD, comparando duas abordagens: uma regra heurística tradicional, sem IA, e uma rede neural MLP, com IA. Os resultados demonstram que a abordagem sem IA apresentou alta precisão, mas baixo recall, resultando em elevada taxa de falsos negativos. Em contraste, as variantes com IA, utilizando estratégias de balanceamento de classes, mostraram ganhos substanciais em recall e F1-Score, além de melhores valores de ROC-AUC e PR-AUC. Conclui-se que a IA proporciona avanços significativos para a cibersegurança, tornando os sistemas mais eficazes, adaptáveis e capazes de responder a ataques complexos em tempo real.

Palavras-chave: inteligência artificial, cibersegurança, aprendizado de máquina, detecção de intrusões, redes Neurais.

Abstract

This work investigates the impact of Artificial Intelligence (AI) on cybersecurity, focusing on its application in intrusion detection and prevention. The study presents a theoretical review of the main AI techniques used in the field, including neural networks, expert systems, intelligent agents, and machine learning. An experimental study is then carried out using the NSL-KDD dataset, comparing two approaches: a traditional heuristic rule (without AI) and a multilayer perceptron neural network (with AI). The results show that the non-AI approach achieved high precision but low recall, resulting in a high rate of false negatives. In contrast, the AI-based approaches, using class balancing strategies, showed substantial improvements in recall and F1-Score, as well as higher ROC-AUC and PR-AUC values. It is concluded that AI provides

significant advances for cybersecurity, making systems more effective, adaptable, and capable of responding to complex attacks in real time.

Keywords: artificial intelligence, cybersecurity, machine learning, intrusion detection, neural networks.

1. Introdução

Vivemos em uma era digital marcada por uma crescente dependência de sistemas conectados à internet, o que proporciona inúmeros benefícios, mas também abre espaço para ameaças cibernéticas cada vez mais sofisticadas. Desde pequenas empresas até grandes corporações e instituições governamentais, todas estão sujeitas a ataques que podem comprometer dados sensíveis, interromper serviços e gerar prejuízos financeiros e reputacionais.

Nesse contexto, a cibersegurança tornou-se um campo estratégico e essencial, exigindo respostas mais ágeis e inteligentes do que os métodos tradicionais de defesa baseados apenas em regras fixas e análises manuais. É nesse cenário que a Inteligência Artificial (IA) tem se destacado como uma aliada poderosa, oferecendo ferramentas capazes de detectar padrões anômalos, antecipar comportamentos maliciosos e reagir automaticamente a incidentes em tempo real.

A aplicação de técnicas de IA, como Machine Learning, redes neurais e sistemas especialistas, tem transformado como as organizações lidam com ameaças digitais. A capacidade dessas tecnologias de aprender com grandes volumes de dados e adaptar-se continuamente às mudanças do ambiente proporciona uma vantagem significativa frente aos cibercriminosos, que também têm recorrido a métodos cada vez mais avançados.

O objetivo deste trabalho é comparar uma abordagem tradicional baseada em regras fixas com uma abordagem moderna baseada em Inteligência Artificial especificamente uma rede neural do tipo MLP na tarefa de detecção de intrusões, utilizando a base de dados NSL-KDD. Para isso, inicialmente são apresentados os fundamentos teóricos da Inteligência Artificial aplicados à cibersegurança, contextualizando seu papel na detecção de ameaças e anomalias em redes computacionais. Em seguida, a base NSL-KDD é descrita, destacando suas características, estrutura e relevância acadêmica. Na etapa experimental, duas estratégias são implementadas: uma abordagem sem IA, fundamentada em uma regra heurística simples, e uma abordagem com IA, por meio de uma rede neural *MLP*. Por fim, os resultados obtidos são comparados por meio de métricas de avaliação adequadas, possibilitando uma análise crítica do desempenho de cada método.

Metodologicamente, o trabalho adota um estudo experimental sobre a base de dados NSL-KDD. A análise é conduzida em duas etapas: (i) aplicação de uma regra heurística simples para identificação de tráfego anômalo, representando a abordagem tradicional sem IA, e (ii) aplicação de uma rede neural MLP com técnicas de balanceamento de classes (class_weight e SMOTE). Ambas as abordagens são avaliadas no conjunto de teste KDDTest+, com base em métricas clássicas (Acurácia, Precisão, Recall e F1-Score) e métricas robustas para dados desbalanceados (ROC-AUC e PR-AUC).

2. Inteligência Artificial e Cibersegurança

O termo Inteligência Artificial refere-se a um campo da computação que visa desenvolver algoritmos para a solução de problemas e a automação de tarefas diversas, utilizando diferentes técnicas e modelos conforme o contexto a ser abordado. A ideia de IA começou a ser debatida por volta dos anos 1950, quando Alan Turing destacou conceitos sobre como os computadores poderiam aprender. No entanto, foi em 1956, na Conferência de Dartmouth, que John McCarthy introduziu oficialmente o termo IA.

De acordo com Stryker e Kavlakoglu (2024), a IA pode ser dividida em duas categorias principais:

- IA Fraca: Projetada para executar tarefas específicas, como ocorre com os diversos sistemas de IA presentes atualmente, incluindo assistentes virtuais (ex: Siri, Alexa).
- IA Forte: Possui dois conceitos principais: Inteligência Artificial Geral e Superinteligência. A IA Geral teria capacidades semelhantes às dos seres humanos, podendo raciocinar, aprender e planejar de maneira autônoma. Já a Superinteligência superaria a capacidade do cérebro humano, ultrapassando o raciocínio e a criatividade humana. No entanto, ainda não existem exemplos reais desse tipo de IA, sendo frequentemente explorada em filmes de ficção científica.

A cibersegurança consiste na proteção e defesa de sistemas, redes e programas no ciberespaço contra possíveis ataques maliciosos, nomeadamente por meio da aplicação de normas, regulamentos, encriptação, para se conseguir evitar possíveis danos, sejam estes ao nível de hardware ou de software (OLIVEIRA, 2021)

Pode ser também considerada como um conjunto de ferramentas, políticas, conceitos de segurança, guias, abordagens de gestão de risco, melhores práticas, tecnologias que podem ser utilizadas para proteger o ciberespaço, organizações e utilizadores. Computadores, infraestruturas, serviços de sistemas de telecomunicações também devem ser protegidos, uma vez que guardam informação relevante no ciberespaço (OLIVEIRA, 2021)

Com o avanço contínuo da tecnologia, a cibersegurança tornou-se essencial em diversas áreas, garantindo a proteção de computadores, redes e sistemas contra ameaças cada vez mais sofisticadas. Ela é um ramo da Tecnologia da Informação que visa criar ambientes seguros por meio de práticas, tecnologias e processos específicos. Essas práticas são fundamentais para prevenir ataques cibernéticos, que podem ocorrer de diversas formas. Conforme o portal CERT.br, no primeiro mês de 2025, os principais tipos de ataques registrados foram: Scan, DoS *(Denial of Service*), Fraudes, Ataques Web e Invasões. Esses incidentes demonstram a crescente importância da cibersegurança para proteger dados sensíveis e garantir a integridade dos sistemas digitais.

Para lidar com esses desafios, há uma necessidade crescente de inovar e aprimorar constantemente a segurança digital. Nesse cenário, a Inteligência Artificial tem sido uma grande aliada da cibersegurança. Com o uso de IA, é possível aprimorar os processos de detecção, monitoramento e prevenção de ameaças. Isso permite respostas mais rápidas e eficientes. Assim, a integração entre IA e cibersegurança fortalece os mecanismos de defesa digital, tornando os sistemas mais inteligentes e adaptáveis.

3. Aplicações de Inteligência Artificial na Cibersegurança

O uso da Inteligência Artificial (IA) na defesa cibernética tem avançado significativamente. Ferramentas como sistemas de detecção e prevenção de intrusões (IDS/IPS) aprimorados com IA são capazes de processar dados em tempo real, aprendendo com incidentes para prever e mitigar ameaças futuras. Segundo Hartmann e Steup (2020), a IA e o aprendizado de máquina (ML) são integrados em produtos de monitoramento de rede e segurança de endpoint, com IDS inteligentes e defesa cibernética automatizada como exemplos de ativos que dependem dessas tecnologias.

3.1 Rede Neurais

Uma rede neural artificial é um modelo de aprendizado de máquina que utiliza um processamento não-linear de entradas para compor o resultado final, ou seja, a função de saída é uma correlação não-proporcional das entradas. Isso significa dizer que a saída é definida a partir de um conjunto de etapas de processamento das entradas, onde cada etapa têm um determinado impacto na saída. Essas etapas de processamento são apelidadas de grupos de neurônios pela similaridade metafórica com a atividade cerebral. Esses grupos de neurônios formam as chamadas camadas ocultas da rede neural e, para cada camada, a saída da camada anterior servirá como entrada para a atual. Sendo assim, a última camada será a responsável por gerar a classificação final (GOMES, 2019).

No contexto da cibersegurança, as *GANs* são utilizadas para criar cenários simulados de ataque, nos quais uma rede geradora tenta produzir ataques, enquanto uma rede discriminadora trabalha para identificar essas ameaças. Essa abordagem demonstra que a competição entre redes fortalece os sistemas de defesa, tornando-os mais eficazes frente a ataques que seriam invisíveis a métodos tradicionais. As Redes Neurais Profundas são capazes de identificar padrões complexos em grandes volumes de dados, evidenciando que, quando aplicadas a análises de tráfego ou comportamentos anômalos, elas ampliam significativamente a capacidade de detecção de ameaças cibernéticas(CRUZ et al., 2024).

As *CNNs* são amplamente utilizadas para a análise de dados com estrutura visual ou espacial, como imagens e padrões gráficos. Isso sugere que a sua aplicação em cibersegurança permite a identificação visual de ataques ou irregularidades em grandes volumes de dados, o que seria inviável manualmente (CRUZ et al., 2024).

A *Recurrent Deep Neural Networks* (*RNN*) é uma classe de rede neural artificial profunda onde as conexões dos neurônios internos formam grafos direcionados, como loops internos, possibilitando a persistência das informações na rede. Esse loop interno permite que uma informação seja passada de uma camada a outra dentro da rede em várias instâncias da rede. Sendo assim, é possível para a rede operar sobre sequências de vetores dados como entrada e gerar sequências de vetores como saída, aumentando a eficiência da rede e gerando modelos de representação bem mais complexos, sem perda de acurácia, se comparado às redes neurais artificiais tradicionais (GOMES, 2019).

3.2 Sistemas Especialistas

Os sistemas especialistas (*experts systems*) ou sistemas baseados em conhecimento tentam apoderar-se de um conjunto de conhecimento especializados num determinado domínio dos humanos e posteriormente aplicar esse conhecimento, tentando encontrar respostas a determinas questões aplicadas a esse domínio (MORGADO, 2016).

Geralmente, esses sistemas funcionam por meio de um conjunto de regras que analisam informações sobre um problema específico e sugerem possíveis ações para resolvê-lo. A característica essencial dos sistemas especialistas é a capacidade de simular um raciocínio lógico e estruturado.

Conceitualmente, um sistema especialista é composto por uma base de conhecimento, na qual são armazenadas informações sobre um determinado tema, e por um mecanismo de inferência, que utiliza esse conhecimento para gerar respostas por meio de regras lógicas. Antes de sua utilização, é fundamental que a base de conhecimento seja preenchida com informações relevantes. A utilidade e a confiabilidade desses sistemas dependem diretamente da qualidade e da quantidade do conhecimento adquirido, tornando a fase de aquisição de conhecimento um aspecto crítico no desenvolvimento dessas aplicações (WIRKUTTIS; KLEIN, 2017).

Na cibersegurança, os sistemas especialistas são amplamente utilizados para apoiar a tomada de decisões estratégicas, auxiliando na definição de medidas de segurança mais adequadas para diferentes cenários. Eles fornecem informações detalhadas sobre como otimizar o uso de recursos disponíveis para enfrentar ameaças específicas. Além disso, esses sistemas também desempenham um papel relevante na detecção de intrusões, contribuindo para a proteção de infraestruturas críticas contra-ataques cibernéticos.

3.3 Agentes Inteligentes

Os agentes inteligentes são entidades computacionais projetadas para atuar autonomamente em ambientes específicos, executando ações com objetivos bem definidos. Esses agentes operam continuamente e podem interagir entre si, colaborando para alcançar soluções eficientes diante de determinadas situações. As principais características dos agentes inteligentes incluem a capacidade de tomar iniciativas, reagir a estímulos externos, comunicar-se com outros agentes e cooperar na resolução de problemas(ANSARI et al., 2022)

Um sistema pode ser classificado como um agente inteligente se for capaz de compreender seu ambiente por meio de sensores e atuar sobre ele por meio de mecanismos apropriados. Atualmente, os agentes inteligentes são amplamente utilizados em diversas aplicações, como plataformas de comparação de preços e mecanismos de busca inteligentes que organizam os resultados com base na relevância das informações.

Na área da cibersegurança, os agentes inteligentes apresentam mobilidade, adaptabilidade e uma forte capacidade de colaboração, tornando-os altamente eficazes na defesa do ciberespaço contra ameaças. Eles são frequentemente empregados na mitigação de ataques de negação de serviço distribuído (*DDoS*) e na detecção de intrusões, proporcionando uma resposta rápida e eficiente a eventos suspeitos.

3.4 Aprendizado de máquina.

O aprendizado de máquina (*Machine Learning* - ML) é uma tecnologia poderosa baseada em algoritmos que aprendem e evoluem a partir de dados. Sua principal vantagem é a capacidade de identificar padrões e aprimorar seu desempenho automaticamente.

Na cibersegurança, o *Machine Learning* se destaca por sua eficiência em analisar grandes volumes de dados em tempo real. Ao contrário dos métodos tradicionais baseados em regras fixas, ele permite a detecção de ameaças emergentes e reduz significativamente o tempo de resposta a incidentes(BUGHUNT, 2023; E-SAFER, 2023; KASPERSKY, 2020)

Algumas das principais aplicações de *Machine Learning* na cibersegurança incluem (BUGHUNT, 2023; KASPERSKY, 2020):

- **Detecção de Ameaças:** utilizando sua capacidade de análise em grandes volumes de dados, o *Machine Learning* identifica padrões de comportamento malicioso com maior precisão.
- **Identificação de Vulnerabilidades:** análise contínua de sistemas para encontrar falhas de segurança. Ele verifica continuamente a existência de falhas em sistemas e aplicativos, detectando anomalias e possíveis riscos.
- **Autenticação e Monitoramento:** verificação de comportamentos atípicos de usuários para detectar possíveis acessos não autorizados. Com a capacidade de identificar padrões, o *Machine Learning* analisa o comportamento dos usuários, reconhecendo atividades suspeitas com base em seus históricos de acesso.

A combinação de *Machine Learning* com outras abordagens de IA fortalece ainda mais a segurança digital, tornando os sistemas mais resilientes e preparados para enfrentar ameaças cibernéticas complexas.

Com o crescimento das ameaças cibernéticas, a utilização de tecnologias como Inteligência Artificial e *Machine Learning* é cada vez mais essencial para garantir a segurança de dados e a continuidade.

4. IA e Cibersegurança no Mercado

Com o aumento das ameaças cibernéticas e a crescente sofisticação dos ataques, empresas em todo o mundo estão incorporando a IA como uma solução vital para reforçar suas defesas digitais. Um exemplo de como a IA pode ser decisiva nesse contexto ocorreu em 2016, quando a Darktrace ajudou uma grande instituição financeira a identificar um ataque de *ransomware*. O sistema de IA da empresa detectou comportamentos anômalos nas comunicações de um servidor e conseguiu impedir que o malware criptografasse dados críticos, evitando danos em larga escala. Esse exemplo ilustra a eficácia da IA na detecção precoce e na prevenção de ataques cibernéticos, sendo uma das principais razões pelas quais tantas empresas estão adotando essas tecnologias em suas estratégias defensivas (DARKTRACE, 2017).

A seguir, algumas das empresas que lideram a integração da IA com a cibersegurança e estão transformando a forma como as organizações protegem suas redes e dados:

- **Darktrace:** um dos primeiros casos emblemáticos ocorreu em 2016, a Darktrace é uma das principais empresas a utilizar IA para cibersegurança. Fundada em 2013, ela aplica a IA para monitorar e proteger redes corporativas contra ataques cibernéticos. Seu sistema, Enterprise Immune System, é inspirado no funcionamento do sistema imunológico humano, detectando atividades anômalas e potenciais ameaças com base no aprendizado de máquina. A Darktrace utiliza IA de forma autônoma para identificar padrões de comportamento normais e detectar variações que possam indicar ataques, como intrusões ou malware, em tempo real (DARKTRACE, 2025).

- **Cylance**: a Cylance, agora integrada à BlackBerry, é reconhecida por sua abordagem inovadora na prevenção de ameaças cibernéticas por meio do uso de inteligência artificial e aprendizado de máquina. Sua principal solução, o CylancePROTECT, utiliza modelos matemáticos avançados para analisar e prever a probabilidade de um código ser malicioso, sem depender de assinaturas tradicionais de ameaças. Essa tecnologia permite detectar e bloquear malwares, incluindo ameaças de dia zero, antes que possam causar danos, oferecendo uma proteção proativa e eficaz aos *endpoints* (BLACKBERRY, 2024).

- **IBM Security:** a IBM tem uma forte presença na área de cibersegurança e utiliza IA para fortalecer suas soluções. A plataforma IBM QRadar combina IA e aprendizado de máquina para analisar grandes volumes de dados e detectar ameaças. A empresa também aplica o Watson for Cyber Security, que usa IA para automatizar a análise de dados de segurança e gerar informações relevantes sobre ataques cibernéticos. Essas soluções permitem uma resposta mais rápida e eficaz a incidentes, ajudando as organizações a enfrentarem as crescentes ameaças digitais (IBM, 2024).

- **Vectra AI:** a Vectra AI utiliza IA para detectar ameaças cibernéticas em tempo real, com foco em ataques persistentes e avançados. Sua plataforma Cognito usa aprendizado de máquina para analisar o tráfego de rede e identificar comportamentos suspeitos. A Vectra é conhecida por sua capacidade de detectar ataques internos e ameaças avançadas que muitas vezes escapam das soluções tradicionais de segurança. A IA da Vectra permite uma defesa contínua e automática contra as ameaças mais sofisticadas, oferecendo uma proteção eficaz para infraestruturas críticas (VECTRA AI, 2025).

- **SentinelOne**: a SentinelOne é uma empresa especializada em segurança de *endpoints*, utilizando IA para detectar e neutralizar ataques em tempo real. Sua plataforma de segurança analisa comportamentos e utiliza aprendizado de máquina para identificar malwares, *ransomware* e ataques de dia zero. A tecnologia de IA da SentinelOne viabiliza uma **detecção ágil e resposta imediata** a ameaças cibernéticas, muitas vezes sem a necessidade de intervenção humana, oferecendo uma solução proativa para proteger os sistemas contra ameaças cibernéticas emergentes (SENTINELONE, 2025).

Essas empresas, entre outras, estão liderando o caminho na integração de IA com a cibersegurança, transformando a forma como as organizações se protegem contra as ameaças digitais. A IA permite detectar e neutralizar ataques com maior

precisão e rapidez, garantindo uma defesa mais robusta e eficiente. À medida que os ciberataques se tornam cada vez mais sofisticados, a adoção de IA nas soluções de segurança digital se torna cada vez mais crucial para proteger dados e garantir a continuidade das operações empresariais.

A Inteligência Artificial tem se mostrado uma grande aliada na proteção digital, oferecendo soluções práticas e automáticas para identificar, prevenir e reagir a ameaças. Com o uso de técnicas como aprendizado de máquina, redes neurais, agentes inteligentes e sistemas especialistas, os sistemas de segurança se tornam mais inteligentes e preparados para lidar com ataques cada vez mais complexos. Por isso, a presença da IA na cibersegurança já não é mais uma opção, é uma necessidade real para acompanhar a evolução dos riscos no mundo digital.

5. Métricas de Avaliação

A avaliação de modelos de detecção de intrusões requer o uso de métricas adequadas, especialmente em cenários de dados desbalanceados, como a base NSL-KDD. Para isso, é fundamental compreender alguns elementos básicos:

- **TP (True Positive)**: número de instâncias positivas corretamente classificadas.
- **TN (True Negative)**: número de instâncias negativas corretamente classificadas.
- **FP (False Positive)**: número de instâncias negativas incorretamente classificadas como positivas.
- **FN (False Negative)**: número de instâncias positivas incorretamente classificadas como negativas.

As principais métricas utilizadas neste trabalho foram **Acurácia, Precisão, Recall, F1-Score, ROC-AUC e PR-AUC**, descritas a seguir:

- **Acurácia:** de acordo com Swaminathan e Tantri (2024), a acurácia representa a proporção de previsões corretas em relação ao total de exemplos avaliados. Embora seja amplamente utilizada, essa métrica pode ser insuficiente em bases desbalanceadas, pois um modelo pode obter alta acurácia mesmo falhando na identificação de instâncias da classe minoritária — cenário comum em tarefas de detecção de intrusões. A fórmula é:

$$(TP+TN) / (TP+TN+FP+FN)$$

- **Precisão (Precision):** de acordo com Swaminathan e Tantri (2024), a precisão é definida como a proporção de exemplos classificados como positivos que realmente pertencem à classe positiva. Essa métrica é relevante quando se deseja minimizar falsos positivos. A fórmula é:

$$TP / (TP+FP)$$

- **Recall (Sensibilidade ou Taxa de Verdadeiros Positivos):** Swaminathan e Tantri (2024) definem o recall como a proporção de instâncias positivas

corretamente identificadas pelo modelo, sendo essencial para medir a capacidade de detectar eventos positivos. Em contextos de cibersegurança, baixos valores de recall indicam falhas na identificação de ataques (falsos negativos). A fórmula é:

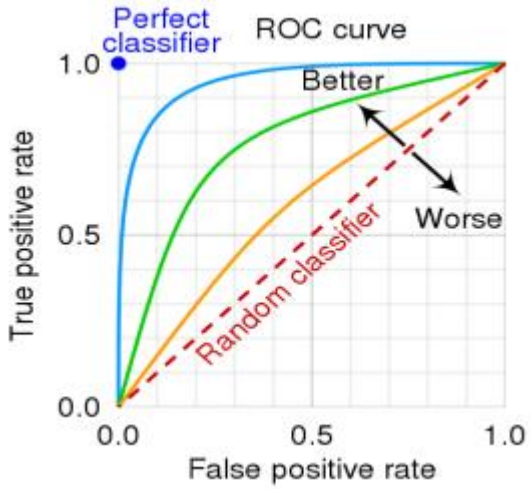
$$TP/ (TP + FN)$$

- **F1-Score:** Segundo Swaminathan e Tantri (2024), o F1-Score é a média harmônica entre precisão e recall, sendo útil especialmente em problemas com desbalanceamento entre classes. Ele penaliza modelos que priorizam apenas uma dessas métricas. A fórmula é:

$$(2\times Precisão\times Recall) / (Precisão + Recall)$$

- **ROC-AUC *(Receiver Operating Characteristic – Area Under the Curve*):** a curva ROC representa a relação entre a taxa de verdadeiros positivos (TPR) e a taxa de falsos positivos (FPR) em diferentes limiares de decisão. A métrica AUC é comumente usada para a avaliação de modelos para a **classificação binária** e mede a área total abaixo da curva ROC, de (0,0) a (1,1), sendo igual a **1** para um modelo completamente correto, que é, portanto, um classificador perfeito. Um modelo totalmente incorreto tem uma AUC de **0**. O gráfico da figura 5 demonstra a curva ROC-AUC (SATHYANARAYANAN; TANTRI, 2024).

Figura 5 - Representação do gráfico ROC-AUC

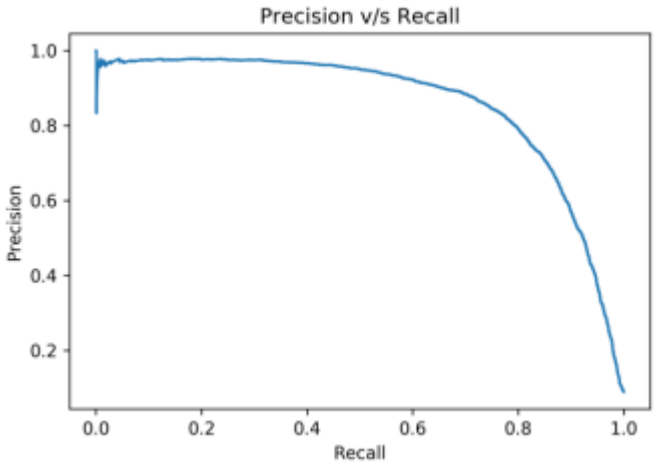


Fonte: SATHYANARAYANAN; TANTRI 2024.

- **PR-AUC (Precision-Recall – Area Under the Curve):** a curva PR relaciona precisão e recall ao longo de diferentes limiares. Segundo SATHYANARAYANANe TANTRI (2024), essa curva pode ser uma alternativa à curva ROC para dados desbalanceados. Um classificador que produz uma curva próxima ao canto superior direito é considerado bom. A figura 6 mostra um exemplo de curva precisão-recall.O PR-AUC quantifica essa área, sendo

especialmente relevante para aplicações de segurança, onde os ataques costumam ser minoria.

Figura 6 - *Representação do grafico PR-AUC*



Fonte: SATHYANARAYANAN, S.; TANTRI, B. R, 2024.

Dessa forma, a combinação dessas métricas fornece uma avaliação abrangente: enquanto a Acurácia e a Precisão capturam aspectos gerais e conservadores do desempenho, o Recall e o F1-Score permitem mensurar a capacidade do modelo em não deixar ataques passarem despercebidos. Já as métricas ROC-AUC e PR-AUC complementam a análise em cenários desbalanceados, oferecendo uma visão mais robusta da separação.

6. Desenvolvimento

Esta seção descreve o desenvolvimento experimental conduzido sobre a base de dados NSL-KDD, aplicando uma regra heurística simples (sem IA) e um modelo de rede neural MLP (com IA), permitindo a comparação de desempenho em métricas clássicas e específicas para cenários desbalanceados. O pipeline adotado compreende: pré-processamento com One-Hot Encoding (handle_unknown='ignore') para variáveis categóricas e StandardScaler para variáveis numéricas; divisão dos dados em treino, validação e teste cego (KDDTest+ mantido intocado); e tratamento do desbalanceamento por duas estratégias complementares class_weight='balanced' e SMOTE (reamostragem do treino). O treinamento da MLP utiliza EarlyStopping e ModelCheckpoint para evitar sobrereajuste e preservar o melhor modelo.

A abordagem sem IA representa um método tradicional baseado em regra fixa, enquanto a abordagem com IA aprende padrões discriminativos a partir dos dados. Ambas são avaliadas no mesmo conjunto de teste, reportando Acurácia, Precisão, Recall, F1-Score, além de ROC-AUC e PR-AUC (mais informativa em dados desbalanceados), com matrizes de confusão e, quando pertinente, curvas ROC/PR e curva de aprendizado para análise de ganho com mais dados.

6.1 Base de Dados NSL-KDD

Para a realização dos testes experimentais deste trabalho, foi utilizada a base de dados NSL-KDD, amplamente reconhecida na área de cibersegurança e frequentemente empregada como benchmark para avaliação de sistemas de detecção de intrusões. Essa base foi desenvolvida como uma evolução da KDD Cup 1999, buscando superar limitações relevantes como a presença excessiva de registros duplicados e o forte desequilíbrio entre classes, fatores que podiam levar modelos a memorizar padrões e dificultar uma avaliação justa de algoritmos de aprendizado de máquina (DIVEKAR et al., 2018).

A escolha pela base NSL-KDD se justifica por diferentes fatores. Primeiramente, trata-se de um conjunto de dados amplamente adotado em pesquisas acadêmicas, o que possibilita a comparação direta dos resultados obtidos neste trabalho com estudos previamente publicados. Além disso, a base apresenta uma estrutura organizada e bem documentada, facilitando as etapas de pré-processamento, análise e modelagem. Por fim, a NSL-KDD contém rótulos claramente definidos para cada instância, característica fundamental para a condução de experimentos supervisionados em detecção de intrusões, permitindo avaliar com precisão o desempenho de diferentes abordagens.

Embora existam conjuntos de dados mais recentes e realistas como UNSW-NB15 e CICIDS2017, que incorporam tráfego moderno e ataques contemporâneos, optou-se pela NSL-KDD por sua ampla aceitação na literatura, documentação consolidada e fácil disponibilidade. Esses fatores preservam a reprodutibilidade científica e permitem comparação direta com estudos prévios (DIVEKAR et al., 2018), além de alinharem-se ao objetivo deste trabalho sem introduzir variáveis adicionais de coleta e infraestrutura. Assim, a NSL-KDD permanece uma base sólida para estudos comparativos entre abordagens tradicionais e baseadas em IA.

A base NSL-KDD é composta por **41 atributos** que descrevem características do tráfego de rede, incluindo variáveis de rede como duração da conexão, tipo de protocolo (*protocol_type*), serviço solicitado (*service*), estado da conexão TCP (*flag*), quantidade de dados enviados e recebidos (*src_bytes* e *dst_bytes*), contadores de eventos no host e na rede, além de taxas que refletem padrões de tráfego e possíveis erros de conexão. Cada registro também recebe um rótulo indicando se a conexão é normal ou corresponde a um ataque.

Os ataques presentes no dataset são agrupados em quatro categorias principais. No conjunto de treinamento (KDDTrain+), a distribuição é composta por **53,46%** de tráfego normal, enquanto os ataques se dividem em **DoS (36,46%)**, **Probe (9,25%)**, **R2L (0,79%)** e **U2R (0,04%)**. No conjunto de teste (KDDTest+), a proporção de ataques raros aumenta: **43,08%** do tráfego é normal, seguido por **DoS (33,09%)**, **Probe (10,74%)**, **R2L (12,80%)** e **U2R (0,30%)**. Essa distribuição evidencia o desbalanceamento da base, principalmente nas classes R2L e U2R, que possuem baixa representatividade.

Para fins deste estudo, os rótulos originais foram **binarizados**, de forma que conexões normais foram codificadas como **0** e todas as categorias de ataque foram agrupadas e codificadas como **1**. Essa transformação permitiu a formulação do problema como uma tarefa de **classificação binária**, alinhando o objetivo do trabalho: distinguir tráfego legítimo de tráfego malicioso.

A etapa de preparação dos dados incluiu a identificação das colunas, a transformação de atributos categóricos em variáveis numéricas e a conversão dos rótulos de ataque em uma classificação binária (“normal” ou “malicioso”), a fim de facilitar a aplicação das abordagens de detecção propostas neste trabalho.

6.2 Etapas de Pré-processamento

Antes da aplicação das abordagens de detecção de intrusões, foi necessário realizar o pré-processamento dos dados. Essa etapa garante que a base esteja no formato adequado para a aplicação tanto de regras manuais quanto de técnicas de *Inteligência Artificial*.

6.2 Pré-processamento dos Dados

Após a importação dos arquivos KDDTrain+ e KDDTest+, realizou-se uma verificação de valores nulos ou inconsistentes por meio da função isnull().sum(). Não foram encontrados valores ausentes, indicando que o conjunto está completo e pronto para uso.

Em seguida, os atributos categóricos (protocol_type, service e flag) foram codificados por meio de ***One-Hot Encoding***, configurado com handle_unknown='ignore' para evitar erros no caso de categorias presentes no teste e ausentes no treino — situação especialmente comum na variável *service*.

Os atributos numéricos foram padronizados com **StandardScaler**, que ajusta os valores para média zero e desvio padrão igual a um, favorecendo a estabilidade e desempenho no treinamento da rede neural. Ambas as etapas — codificação categórica e padronização — foram integradas a um **ColumnTransformer**, compondo um pipeline único e automatizado de pré-processamento, assegurando consistência nas transformações aplicadas aos dados.

Para evitar vazamento de dados, o ajuste do pipeline (fit) foi realizado exclusivamente no conjunto de treino, sendo posteriormente aplicado (transform) aos conjuntos de validação e teste. O conjunto de teste oficial (KDDTest+) permaneceu completamente cego até a etapa final, preservando a imparcialidade da avaliação.

Após o pré-processamento, os dados foram divididos em X, contendo os 41 atributos transformados (numéricos padronizados e variáveis resultantes do One-Hot), e Y, correspondente ao rótulo original. Por fim, os rótulos foram **binarizados**, sendo definido normal = 0 e attack = 1, padronização adequada ao objetivo deste estudo, que consiste em distinguir tráfego benigno de tráfego malicioso na base NSL-KDD.

6.3 Abordagem Tradicional: Regra Heurística

Nesta etapa foi aplicada uma abordagem tradicional de detecção de intrusões baseada em **regras fixas**, sem o uso de algoritmos de aprendizado de máquina. Esse tipo de método é característico de sistemas clássicos de segurança cibernética, conhecidos como **sistemas especialistas baseados em assinaturas e limiares**, nos quais condições específicas do tráfego de rede são utilizadas para identificar comportamentos anômalos. O objetivo é simular a lógica de decisão de ferramentas que dependem de padrões previamente definidos para sinalizar possíveis ataques.

A regra heurística adotada neste trabalho utiliza **duas variáveis do protocolo TCP** que, segundo estudos de *Choudhary e Kesswani (2020)*, tendem a aumentar em situações de ataque:

- **serror_rate**: representa a proporção de conexões que não foram concluídas corretamente, refletindo falhas durante o processo de handshake TCP. Valores elevados dessa métrica estão frequentemente associados a tentativas de conexão maliciosas ou tráfego anômalo.

- **dst_host_serror_rate:** mede a proporção de falhas de conexão concentradas em um mesmo host de destino. Quando esse valor é alto, indica que um servidor específico está sendo sobrecarregado por múltiplas tentativas de conexão malsucedidas, comportamento típico de ataques

A escolha dessas duas variáveis decorreu da **análise exploratória da base NSL-KDD**, na qual observou-se que ataques tendem a elevar simultaneamente ambas as taxas. Assim, definiu-se uma **regra simples e interpretável** que permite distinguir o tráfego normal do potencialmente malicioso com base em limiares fixos. O critério adotado foi o seguinte:

> Se **dst_host_serror_rate > 0,5** ou **serror_rate > 0,5**, classificar como *attack*; caso contrário, classificar como *normal*.

O valor de corte **0,5** foi escolhido com base na distribuição empírica dos dados e em testes preliminares, representando um ponto de equilíbrio entre evitar falsos alarmes e ainda capturar um número razoável de ataques.

A Figura 7 apresenta visualmente a lógica dessa regra, ilustrando como a decisão é tomada com base nas duas métricas de taxa de erro.

Figura 7 - Regra Heurística

```
# Regra: attack se dst_host_serror_rate > 0.5 ou serror_rate > 0.5.
y_pred_rule = [
    'attack' if (row['dst_host_serror_rate'] > 0.5 or row['serror_rate'] > 0.5)
    else 'normal'
    for _, row in X_test_baseline.iterrows()
]
```

Fonte: Autores.

6.3.2 Preparação para avaliação

- **Partição usada:** a regra não precisa de treino, então foi aplicada diretamente no conjunto de teste (KDDTest+).

- **Seleção de atributos:** para a regra, só interessam serror_rate e dst_host_serror_rate. Os demais campos (incluindo as categóricas) não entram nesta abordagem.

- **Binarização do rótulo:** os rótulos do teste foram mapeados para binário: normal = 0, attack = 1 (todas as demais classes).

- **Geração das predições:** para cada linha do teste, aplica-se a regra e grava-se attack ou normal.

6.3.3 Resultados no conjunto de teste (KDDTest+)

Os resultados obtidos com a abordagem sem IA são apresentados a seguir, Figura 8 ilustra os resultados obtidos na Acurácia, Precisão, Recall, F1.

- **Acurácia:** 0,5322 → o modelo acertou pouco mais da metade das previsões.
- **Precisão:** 0,9831 → quase todos os exemplos classificados como ataque realmente eram ataques (poucos falsos positivos).
- **Recall:** 0,1814 → apenas 18% dos ataques foram detectados (muitos falsos negativos).
- **F1:** 0,3064 → baixo equilíbrio entre precisão e recall, mostrando fragilidade do método.

Figura 8 - Resultados obtidos na abordagem sem IA

```
Baseline sem IA:
Acurácia : 0.5322924059616749
Precisão : 0.9831152384972562
Recall   : 0.18148523338268527
F1       : 0.3064070517037232
```

Fonte: Autores.

Além disso, também foi gerado a matriz de confusão que mostra acertos/erros por classe. A Figura 9, apresenta essa matriz.

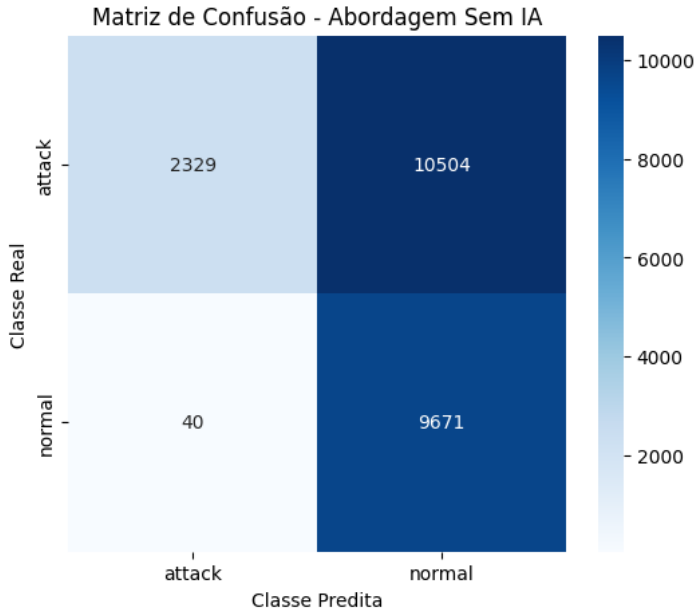
Essa distribuição mostra que a abordagem sem IA apresenta forte tendência a classificar o tráfego como normal, resultando em poucos falsos alarmes (baixo FP), mas falhando em detectar a maioria dos ataques (alto FN). Isso explica a **alta Precisão (0,98)** e o **baixo Recall (0,18)**.

6.4 Abordagem com IA: Rede Neural MLP

A rede neural do tipo *Multilayer Perceptron* (MLP) foi adotada como abordagem baseada em Inteligência Artificial para a detecção de intrusões. A MLP pertence à classe de redes neurais *feed-forward* e é composta por camadas de neurônios interconectados capazes de aprender representações não lineares (HAYKIN, 2009; BISHOP, 2006).

O treinamento foi realizado utilizando o conjunto de dados de treino da NSL-KDD, com validação estratificada de 20% para evitar *overfitting* e garantir que as classes fossem representadas proporcionalmente. A avaliação final foi conduzida em um conjunto de teste mantido totalmente separado, garantindo imparcialidade.

Figura 9 - Matriz de confusão sobre a regra sem IA



Fonte: Autores.

A arquitetura escolhida para o modelo foi a seguinte:

Dense(64, ReLU) → Dropout(0,3) → Dense(32, ReLU) → Dropout(0,3) → Dense(1, Sigmoid).

Essa configuração inclui duas camadas ocultas com 64 e 32 neurônios, respectivamente. A escolha desses valores considera um equilíbrio entre capacidade de aprendizado e custo computacional: uma quantidade maior de neurônios pode aumentar a capacidade de modelagem, mas também eleva o risco de *overfitting* e influencia no tempo de treinamento. A ativação ReLU foi utilizada nas camadas intermediárias devido à sua eficiência no treinamento de redes profundas e por reduzir problemas de gradiente.

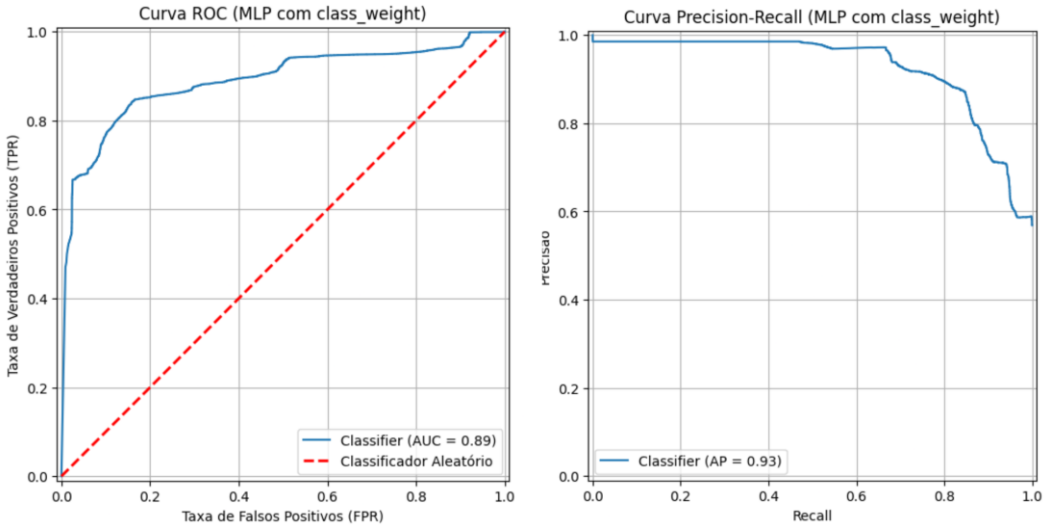
O uso do *Dropout* com taxa de 0,3 em ambas as camadas ocultas teve como objetivo reduzir o *overfitting*, desativando aleatoriamente neurônios durante o treino e forçando a rede a generalizar melhor. A camada de saída utiliza uma única unidade com ativação Sigmoid, apropriada para tarefas de classificação binária (normal vs ataque), produzindo uma probabilidade associada à classe positiva.

O modelo foi otimizado com o algoritmo Adam, amplamente utilizado por sua estabilidade e adaptação dinâmica da taxa de aprendizado, e treinado com a função de perda *binary cross-entropy*, padrão para problemas de classificação binária. Foram empregados os mecanismos de *EarlyStopping* e *ModelCheckpoint* para interromper o treinamento automaticamente ao detectar ausência de melhoria e salvar o melhor modelo encontrado, respectivamente, evitando sobreajuste e garantindo eficiência no processo de treinamento.

6.4.1 MLP com class_weight='balanced'

Ao atribuir peso maior à classe minoritária (ataques), ela força o modelo a "prestar mais atenção" nos exemplos raros, sem alterar o conjunto de dados. No teste cego, a MLP com class_weight atingiu alto equilíbrio entre Precisão e Recall, refletindo no F1-score elevado. As curvas ROC-AUC e PR-AUC também foram altas, indicando bom poder de ranqueamento, representada na Figura 10.

Figura 10 - Representação das Curvas ROC-AUC e PR-AUC



Fonte: Autores.

Comparada à regra heurística, a MLP demonstrou desempenho substancialmente superior na detecção de intrusões, recuperando grande parte dos ataques anteriormente não identificados pela abordagem sem IA. O modelo reduziu significativamente os falsos negativos, ao mesmo tempo em que manteve um número reduzido de falsos positivos, indicando uma boa capacidade de generalização. Essa configuração apresentou o melhor equilíbrio entre precisão e sensibilidade dentre os métodos avaliados, refletido no seu maior F1-Score.

Na Figura 11, observa-se que a rede neural atingiu acurácia de aproximadamente **0,797**, evidenciando que cerca de 80% das classificações foram corretas. A precisão foi de **0,934**, o que demonstra que, quando o modelo classificou uma conexão como ataque, ele estava correto em 93% das vezes, reduzindo alarmes falsos. O valor de **recall igual a 0,698** indica que o modelo foi capaz de identificar aproximadamente 70% dos ataques reais desempenho significativamente superior ao da abordagem sem IA, que alcançou apenas 0,18. O F1-Score obtido (**0,796** confirma um bom equilíbrio entre precisão e sensibilidade. Além disso, os resultados de **ROC-AUC (0,889)** e **PR-AUC (0,926)** reforçam a capacidade do modelo em separar, de forma consistente, tráfego normal de tráfego malicioso mesmo em um cenário de desbalanceamento de classes.

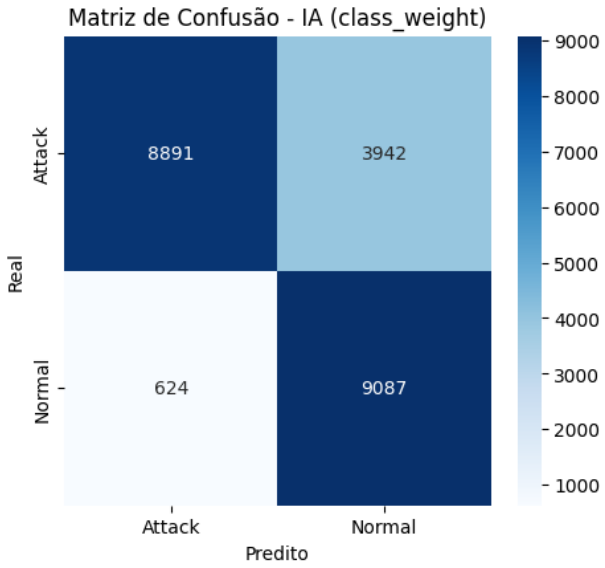
Figura 11 - Resultado da aplicação MLP com cla

```
MLP (class_weight):
Acurácia : 0.7974627395315826
Precisão : 0.9344193378875459
Recall   : 0.6928231902127328
F1       : 0.7956864148917129
ROC AUC  : 0.8899948959753361
PR  AUC  : 0.926129640865681
```

Fonte: Autores.

A matriz de confusão apresentada na Figura 12 evidencia o bom desempenho da MLP com balanceamento por `class_weight`. Observa-se que o modelo classificou corretamente 8.087 tráfegos normais (True Negatives) e 8.891 ataques (True Positives). Em contrapartida, 624 conexões legítimas foram marcadas incorretamente como ataque (False Positives) e 3.942 ataques não foram detectados (False Negatives). Esses resultados indicam uma significativa redução de falsos negativos em comparação com a abordagem sem IA, mantendo um baixo nível de falsos alarmes, o que destaca a eficácia do modelo na detecção de intrusões.

Figura 12 - Matriz de Confusão da MLP com class_weight='balanced



Fonte: Autores.

Esse modelo apresentou bom equilíbrio entre precisão e *recall*, demonstrando capacidade eficaz de identificação de ataques. Embora ainda existam falsos negativos (3.942 ataques não detectados), o *recall* obtido é significativamente superior ao da abordagem sem IA, evidenciando maior eficiência na detecção de tráfego malicioso. A precisão permaneceu elevada (0,93), indicando que o modelo raramente gera falsos alarmes e mantém alta confiabilidade nas classificações realizadas. O *F1-Score* próximo de 0,80 reforça a robustez e o equilíbrio do desempenho alcançado. Adicionalmente, as áreas sob as curvas ROC-AUC (≈ 0,89) e PR-AUC (≈ 0,93) confirmam a capacidade do modelo em discriminar de forma consistente o tráfego normal do tráfego anômalo, mesmo diante do desbalanceamento da base de dados.

Em um contexto prático, o uso de IA com a MLP (parâmetro *class_weight*)

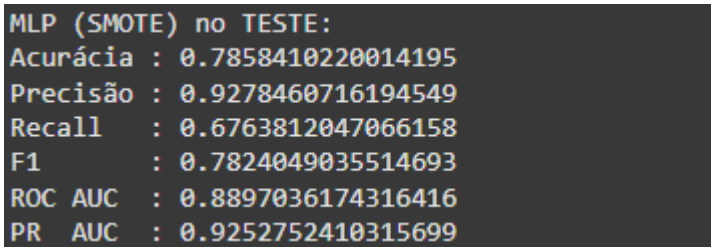
mostrou-se consideravelmente mais eficiente do que a abordagem sem IA, reduzindo de maneira expressiva os falsos negativos e oferecendo uma solução mais confiável para a cibersegurança.

6.4.2 MLP com SMOTE (reamostragem da minoria)

O método SMOTE (*Synthetic Minority Over-sampling Technique*) foi aplicado para gerar amostras sintéticas da classe minoritária (ataques) exclusivamente no conjunto de treino, mantendo os conjuntos de validação e teste inalterados. Essa abordagem reduz o desbalanceamento da base e melhora o aprendizado da fronteira de decisão pelo modelo.

No teste cego, a *MLP* com *SMOTE* apresentou desempenho muito próximo ao class_weight, o que pode ser visto na Figura 13. Em geral, obtivemos AUC-ROC e PR-AUC ligeiramente maiores com SMOTE, enquanto o *class_weight* manteve o melhor F1-score. O *SMOTE* aumenta a cobertura de ataques (*Recall*) com custo pequeno em Precisão. Em termos de áreas sob as curvas (ROC-AUC/PR-AUC), *SMOTE* ficou levemente à frente, indicando melhor ranqueamento de probabilidade e Aprendizado.

Figura 13 - Resultados da abordagem com Smote



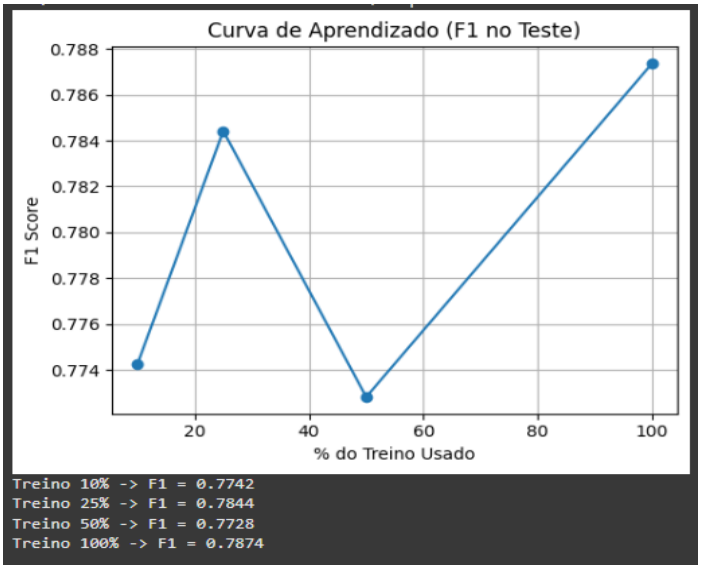
Fonte: Autores.

Além disso, a Figura 14 apresenta a **Curva de Aprendizado (Learning Curve)** do modelo MLP, mostrando o desempenho medido pelo **F1-Score** no conjunto de teste em função da quantidade de dados de treino utilizada (10%, 25%, 50% e 100%).

Observa-se que o modelo já atinge desempenho relativamente estável mesmo com pequenas frações dos dados (10%–25%), alcançando F1 próximo de **0,78**. Ao utilizar 50% do treino, houve uma pequena queda (F1 = 0,7728), possivelmente devido à variação estocástica do processo de treinamento. Com 100% do treino, o modelo atingiu seu melhor resultado (F1 = 0,7874), embora o ganho em relação a 25% ou 10% tenha sido marginal.

A curva evidência que o modelo **satura rapidamente**: cerca de 50% dos dados já capturam a maior parte da sua capacidade preditiva, e treinar com todo o conjunto apenas gera ganhos incrementais. Isso indica que a MLP, na configuração utilizada, não depende fortemente de todo o dataset para alcançar bom desempenho, embora mais exemplos ainda contribuam para melhorar.

Figura 14 - Curva de Aprendizado (Learning Curve) do modelo MLP com SMOTE



Fonte: Autores.

6.5 Análise Comparativa

Os resultados obtidos no conjunto KDDTest+, apresentados na **Tabela 1**, evidenciam que a regra heurística, embora apresente alta precisão (0,98), sofre com baixo recall (0,18), o que implica elevada taxa de falsos negativos e consequente subdetecção de ataques, cenário indesejável em sistemas de detecção de intrusões.

Em contraste, as abordagens baseadas em MLP (com *class_weight* e SMOTE) alcançaram desempenho substancialmente superior, com F1-Score em torno de 0,78–0,79 e áreas sob as curvas (ROC-AUC ≈0,89–0,90; PR-AUC ≈0,92–0,93). A variante com *class_weight* apresentou melhor equilíbrio entre precisão e recall no limiar padrão (0,5), enquanto o SMOTE obteve ligeira vantagem em ROC-AUC e PR-AUC, indicando maior robustez no ranqueamento probabilístico.

Tabela 1 - Comparação entre os resultados das abordagens

Comparação entre abordagens

	Acurácia	Precisão	Recall	F1	ROC-AUC	PR-AUC
Sem IA (Regra)	0.5323	0.9831	0.1815	0.3064	NaN	NaN
IA (class_weight)	0.7975	0.9344	0.6928	0.7957	0.890	0.9261
IA (SMOTE)	0.7932	0.9606	0.6638	0.7851	0.909	0.9362

Fonte: Autores.

A escolha entre elas deve considerar a política de risco da organização:

- **Se a prioridade for detectar o maior número possível de ataques (reduzir FN)** → recomenda-se SMOTE ou ajuste do limiar de decisão.

- **Se a prioridade for reduzir falsos alarmes (FP)** → recomenda-se o uso de *class_weight* com limiar padrão.

Assim, ambas as variantes de IA superam claramente a abordagem sem IA, tornando a detecção mais eficaz e robusta frente a ataques cibernéticos.

7 Conclusão

Este trabalho comparou, de forma sistemática, duas estratégias de detecção de intrusões na base NSL-KDD: uma abordagem sem Inteligência Artificial (IA), baseada em regra heurística com limiares fixos, e uma abordagem com IA, utilizando uma rede neural MLP. O pipeline proposto incluiu pré-processamento adequado (One-Hot Encoding com *handle_unknown='ignore'* e padronização), validação sobre o conjunto de treinamento e avaliação posterior em um conjunto de teste não visto anteriormente, garantindo uma análise imparcial e reprodutível.

A abordagem sem IA apresentou comportamento excessivamente conservador: a regra somente sinalizou ataque quando os indicadores *serror_rate* e *dst_host_serror_rate* atingiram valores elevados. Esse comportamento reduziu falsos positivos, porém não capturou a maioria dos ataques, resultando em *recall* e *F1-score* baixos, apesar da alta precisão. Em termos práticos, isso implica que muitos ataques não seriam detectados em um ambiente real, o que limita o uso exclusivo dessa estratégia em sistemas modernos de detecção de intrusões.

Por outro lado, as abordagens com IA demonstraram ganhos significativos. A MLP treinada com *class_weight='balanced'* apresentou melhor equilíbrio entre precisão e *recall*, atingindo F1 ≈ 0,79. Já a MLP com SMOTE obteve valores superiores de ROC-AUC (≈0,90) e PR-AUC (≈0,93), sugerindo maior capacidade de ordenação probabilística das amostras — característica importante quando limiares de decisão podem ser ajustados dinamicamente em contextos operacionais.

Como contribuições, este trabalho implementa um pipeline metodologicamente sólido, quantifica os ganhos da IA frente a uma baseline sem aprendizado de máquina e evidencia a relevância de métricas adequadas para dados desbalanceados, como PR-AUC. Além disso, foi analisado o impacto do limiar de decisão no equilíbrio entre falsos positivos e falsos negativos, aspecto crítico para sistemas de detecção aplicados em ambientes reais.

Em síntese, os resultados confirmam a hipótese central: a IA supera significativamente a abordagem baseada em regras fixas na detecção de intrusões, oferecendo maior capacidade de identificação de ataques sem comprometer a precisão de maneira relevante. Tanto o *class_weight* quanto o SMOTE apresentaram desempenho competitivo, demonstrando que ambas as estratégias são adequadas e viáveis para cenários práticos.

A principal limitação deste estudo é a utilização exclusiva da base NSL-KDD, que, apesar de amplamente utilizada na literatura, representa um ambiente simulado. Como trabalhos futuros, sugerem-se: (i) aplicar o método em bases mais recentes, como UNSW-NB15 e CICIDS-2017; (ii) investigar arquiteturas mais avançadas, como LSTM e CNN; e (iii) explorar sistemas híbridos que integrem heurísticas e aprendizado

de máquina, alinhados às necessidades operacionais de SOCs e IDS/IPS contemporâneos.

Por fim, ressalta-se a relevância prática dos resultados. A comparação entre as abordagens demonstra que técnicas de IA podem fortalecer Centros de Operações de Segurança (SOCs) e sistemas IDS/IPS, ampliando a detecção de ataques sofisticados e aumentando a resiliência de infraestruturas críticas. Assim, os resultados obtidos não apenas evidenciam o avanço proporcionado pela IA no contexto acadêmico, mas também reforçam seu potencial como ferramenta estratégica na cibersegurança moderna.

Referências

ANSARI, M. F. et al. The impact and limitations of artificial intelligence in cybersecurity: a literature review. IJARCCE, 2022. Disponível em: https://doi.org/10.17148/ijarcce.2022.11912. Acesso em: 8 ago. 2025.

BISHOP, C. M. Pattern Recognition and Machine Learning. New York: Springer, 2006. Disponível em: https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/. Acesso em: 20 fev. 2025.

BLACKBERRY. How Unified Endpoint Security uses advanced technology. BlackBerry Docs, [2024]. Disponível em: https://docs.blackberry.com/en/unified-endpoint-security/blackberry-ues/overview/What-is-Unified-Endpoint-Security/How-ues-uses-advanced-technology. Acesso em: 31 jul. 2025.

BUGHUNT. Machine Learning: o que é e como pode ajudar na cibersegurança? 28 fev. 2023. Disponível em: https://blog.bughunt.com.br/machine-learning. Acesso em: 21 set. 2025.

CERT.br. Estatísticas de Incidentes Notificados ao CERT.br. São Paulo: Núcleo de Informação e Coordenação do Ponto BR, 2025. Disponível em: https://stats.cert.br/incidentes/. Acesso em: 11 jul. 2025.

CRUZ, J. V. S. et al. Inteligência artificial e cibersegurança: análise de ameaças emergentes e estratégias defensivas. REVISTA DELOS, 2024. Disponível em: https://doi.org/10.55905/rdelosv17.n61-193. Acesso em: 29 ago. 2025.

CHOUDHARY, S.; KESSWANI, N. Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 Datasets using Deep Learning in IoT. Procedia Computer Science, 2020. Disponível em: https://doi.org/10.1016/j.procs.2020.03.367. Acesso em: 11 jul. 2025.

DARKTRACE. Defending against ransomware: A live threat scenario. Darktrace Blog, 2017. Disponível em: https://darktrace.com/es/blog/defending-against-ransomware-a-live-threat-scenario. Acesso em: 15 maio 2025.

DARKTRACE. Cyber AI: Augment your security team and stop novel threats. 2025. Disponível em: https://www.darktrace.com/cyber-ai. Acesso em: 8 jul. 2025.

DIVEKAR, A. et al. Benchmarking datasets for Anomaly-based Network Intrusion Detection: KDD CUP 99 alternatives. In: IEEE 3rd International Conference on Computing, Communication and Automation (ICCCA), 2018, p. 1-6. Disponível em:https://arxiv.org/abs/1811.05372. Acesso em: 15 maio 2025.

E-SAFER. A cibersegurança através de Machine Learning. 2023. Disponível em: https://e-safer.com.br/ciberseguranca-com-machine-learning. Acesso em: 21 maio 2025.

GOMES, E. A. S. Aplicabilidade de Algoritmos de Aprendizado de Máquina para Detecção de Intrusão e Análise de Anomalias de Rede. 2019. Trabalho de Conclusão de Curso (Graduação em Engenharia de Sistemas) - Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: https://repositorio.ufmg.br/server/api/core/bitstreams/89b6e643-313b-4e28-be5b-e99d3c6eb8a1/content. Acesso em: 24 set. 2025.

HARTMANN, K.; STEUP, C. Hacking the AI - the Next Generation of Hijacked Systems. In: 12th International Conference on Cyber Conflict (CyCon), 2020. Disponível em: https://ccdcoe.org/uploads/2020/05/CyCon_2020_18_Hartmann_Steup.pdf. Acesso em: 21 maio 2025.

HAYKIN, S. Neural Networks and Learning Machines. 3. ed. Upper Saddle River: Pearson, 2009. Disponível em: https://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf. Acesso em: 20 fev. 2025.

IBM. IBM Security QRadar SIEM. IBM Brasil, 2024. Disponível em: https://www.ibm.com/br-pt/qradar. Acesso em: 6 ago. 2025.

KASPERSKY. IA e aprendizagem por máquina na segurança virtual — Como moldarão o futuro. 2020. Disponível em: https://www.kaspersky.com.br/resource-center/definitions/ai-cybersecurity. Acesso em: 13 ago. 2025.

MORGADO, R. C. O recurso e a contribuição potencial da inteligência artificial para a cibersegurança em ambientes digitais. 2016. Dissertação (Mestrado em Engenharia Eletrotécnica e de Computadores) - Universidade de Lisboa, Lisboa, 2016. Disponível em: https://www.researchgate.net/publication/316241417_O_recurso_e_a_contribuicao_potencial_da_inteligencia_artificial_para_a_ciberseguranca_em_ambientes_digitais. Acesso em: 9 abr. 2025.

NEGIDA, A. Simple Definition and Calculation of Accuracy, Sensitivity and Specificity, 2015. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC4614595. Acesso em: 10 abr. 2025.

OLIVEIRA, V. Cibersegurança e Inteligência Artificial. 2021. Trabalho de Conclusão de Curso (Licenciatura em Engenharia Informática) - Universidade Nova de Lisboa, Lisboa, 2021. Disponível em: https://run.unl.pt/bitstream/10362/117660/1/TGI0405.pdf. Acesso em: 11 ago. 2025.

SATHYANARAYANAN, S.; TANTRI, B. R. Confusion matrix-based performance evaluation metrics. African Journal of Biomedical Research, p. 4023-4031, 30 nov. 2024b. Disponível em: https://doi.org/10.53555/ajbr.v27i4s.4345. Acesso em: 13 set. 2025.

SENTINELONE. Plataforma de Cibersegurança Empresarial com IA. 2025. Disponível em: https://www.sentinelone.com. Acesso em: 27 jun. 2025.
STRYKER, C.; KAVLAKOGLU, E. O que é inteligência artificial (IA)? | IBM. 9 ago. 2024. Disponível em: https://www.ibm.com/br-pt/topics/artificial-intelligence. Acesso em: 30 jul. 2025.

VECTRA AI. Vectra AI: Real-time Threat Detection and response. Vectra, 2025. Disponível em: https://www.vectra.ai. Acesso em: 26 maio 2025.

WIRKUTTIS, N.; KLEIN, H. Artificial intelligence in cybersecurity. Cyber, Intelligence, and Security, 2017. Disponível em: https://www.inss.org.il/wp-content/uploads/2017/03/Artificial-Intelligence-in-Cybersecurity.pdf. Acesso em: 19 jul. 2025.