Abordagens de Aprendizado de Máquina para Segurança em Redes IoT

Leylane Teixeira Bastos
Departamento de Engenharia Elétrica
Universidade Federal do Piauí (UFPI)
Teresina, Brasil
leylane@ufpi.edu.br

Juan de Sousa Rodrigues
Departamento de Engenharia Elétrica
Universidade Federal do Piauí (UFPI)
Teresina, Brasil
eng.juan.rodrigues@gmail.com

José Maria P. de Menezes Júnior
Departamento de Engenharia Elétrica
Universidade Federal do Piauí (UFPI)
Teresina, Brasil
josemenezesjr@ufpi.edu.br

***Resumo*—O avanço da Internet das Coisas (IoT) ampliou a conectividade entre dispositivos e trouxe ganhos em automação e eficiência, mas também aumentou os riscos de segurança. Este trabalho investiga o uso de modelos de aprendizado de máquina para detectar tráfego malicioso em redes IoT, com foco na mitigação do desbalanceamento de dados — desafio comum, já que ataques representam uma fração reduzida dos eventos. Foram avaliados cinco algoritmos supervisionados: Random Forest, Support Vector Machine (SVM), LightGBM, XGBoost e Explainable Boosting Machine (EBM). Aplicou-se oversampling como técnica de balanceamento, e o desempenho dos modelos foi analisado por meio de métricas como acurácia, precisão, recall, F1-score e tempo de treinamento. Os resultados mostraram que o Random Forest obteve o melhor desempenho em recall (98,92%) e apresentou F1-score de 96,04%. Já o LightGBM e o XGBoost também atingiram F1-scores superiores a 95%, com tempos de treinamento inferiores a 25 minutos. O EBM destacou-se pela interpretabilidade e desempenho consistente. Conclui-se que a combinação entre aprendizado de máquina e técnicas de balanceamento contribui para soluções mais robustas em segurança IoT, desde que se equilibrem sensibilidade, custo computacional e transparência.**

***Palavras-chave*—Internet das Coisas, Segurança, Detecção de Ataques, Balanceamento de Dados, Aprendizado de Máquina, Desempenho de Modelos, Oversampling.**

I. Introdução

A Internet das Coisas (IoT, *Internet of Things*) representa uma das transformações tecnológicas mais impactantes do século XXI, conectando uma ampla variedade de dispositivos, desde termostatos inteligentes e câmeras de segurança até equipamentos médicos e veículos autônomos. Essa expansão tem impulsionado avanços significativos em automação, eficiência e conveniência, promovendo a integração desses dispositivos em diversas esferas da vida cotidiana e industrial. De acordo com estimativas da consultoria *IoT Analytics*, o número de conexões IoT deve ultrapassar 29 milhões até 2027 (Sinha, 2023), evidenciando o crescimento exponencial dessa tecnologia.

No entanto, esse avanço também intensifica desafios críticos, especialmente no que se refere à segurança cibernética. Os dispositivos IoT, por sua própria natureza, frequentemente coletam, processam e transmitem informações sensíveis, tornando-se alvos atrativos para ataques. A ampliação desse ecossistema resulta em uma superfície de ataque cada vez maior, expondo vulnerabilidades que podem ser exploradas por agentes mal-intencionados. Durante uma das edições da conferência DEF CON, uma das maiores convenções hacker do mundo, foram identificadas 47 novas vulnerabilidades em 23 tipos de dispositivos IoT (Constantin, 2016), evidenciando a magnitude do problema. Ataques como *Distributed Denial of Service* (DDoS), infiltrações maliciosas e explorações de vulnerabilidades representam ameaças reais e crescentes à segurança dessas redes.

O uso de técnicas de aprendizado de máquina (AM) tem se mostrado promissor na detecção de anomalias e ameaças em redes IoT, dado o grande volume de dados gerado continuamente por dispositivos conectados. Esses modelos aprendem padrões a partir de dados históricos e são capazes de identificar desvios que possam indicar comportamentos maliciosos.

Porém, um desafio importante nas técnicas de detecção de ameaças em ambientes de Internet das Coisas (IoT) é o desbalanceamento dos dados. Em cenários típicos de IoT, o tráfego gerado por operações normais ocorre com muito mais frequência do que eventos de ataque, tornando escassos os exemplos de comportamentos maliciosos. Esse desbalanceamento ocorre por diversos motivos, como a baixa incidência de ataques reais, os custos elevados de coleta desses dados e as dificuldades na rotulagem precisa dos eventos anômalos.

Como consequência, modelos de aprendizado de máquina treinados com esses dados tendem a ser enviesados em favor da classe majoritária, dificultando a identificação precisa de padrões associados a ameaças. Para enfrentar esse problema, técnicas como *oversampling*, *undersampling* e geração de dados sintéticos vêm sendo amplamente utilizadas, buscando melhorar a sensibilidade dos modelos sem comprometer sua precisão.

A eficácia dessas abordagens depende diretamente da escolha de algoritmos e métricas adequadas para avaliação. Em bases desbalanceadas, a acurácia pode se tornar enganosa, mascarando falhas na detecção da classe minoritária. Por isso, métricas como *recall*, *precisão* e F1-score são preferidas, pois permitem uma análise mais criteriosa do desempenho, especialmente no contexto de segurança, onde a falha em identificar ataques pode ter consequências críticas (Krishnaveni et al., 2011). Além disso, é essencial considerar o custo computacional associado ao uso de técnicas de balanceamento, que podem aumentar significativamente o tempo de treinamento

dos modelos.

Neste trabalho, investiga-se a aplicação de técnicas de aprendizado de máquina para a classificação de tráfego malicioso em redes IoT, com ênfase na mitigação dos efeitos do desbalanceamento. Avaliam-se cinco algoritmos — Random Forest, SVM, LightGBM, XGBoost e EBM — considerando diferentes métricas e tempos de execução, com e sem oversampling. A análise busca compreender como essas técnicas impactam o desempenho dos modelos, fornecendo subsídios para a construção de soluções mais eficazes, eficientes e transparentes na proteção de ambientes IoT.

II. Internet das Coisas

A Internet das Coisas (IoT, do inglês *Internet of Things*) refere-se à integração de dispositivos físicos à internet, permitindo que objetos cotidianos adquiram capacidades de coleta, processamento e troca de dados. Sensores, atuadores e sistemas embarcados conectam-se em rede para monitorar e interagir com o ambiente, promovendo automação, eficiência operacional e tomada de decisão inteligente.

A arquitetura da IoT é composta por camadas interdependentes, nas quais sensores capturam informações do meio, que são então transmitidas para processamento e resposta por atuadores ou sistemas remotos. Essa estrutura amplia as possibilidades de aplicação em diversos setores, como indústria, saúde, transporte e domicílios inteligentes.

Entretanto, a interconexão de dispositivos heterogêneos, frequentemente realizada por redes sem fio, impõe desafios significativos à segurança cibernética. A garantia da integridade, confidencialidade e disponibilidade dos dados torna-se fundamental para assegurar a confiabilidade dos sistemas. Dessa forma, a IoT representa uma transformação digital que exige soluções robustas para sua implementação segura e eficaz.

A. Segurança de dados

A Internet das Coisas (IoT) é composta por uma ampla variedade de dispositivos interconectados, que vão desde eletrodomésticos e sensores industriais até dispositivos médicos. Essa diversidade pode ser agrupada em três categorias principais (ZANELLA et al., 2014):

- **Dispositivos Finais (Edge Devices)**: englobam dispositivos de uso pessoal, como *smartphones* e *wearables*, bem como dispositivos voltados para aplicações específicas, como sensores de temperatura residenciais ou rastreadores de atividades físicas.
- **Rede de Sensores e Atuadores**: refere-se ao conjunto de sensores, que captam dados do ambiente, e atuadores, que executam ações com base nesses dados. Essa camada é essencial para a obtenção de informações em tempo real e para a realização de respostas automáticas.
- **Infraestrutura de Rede e Servidores**: abrange servidores, bancos de dados e sistemas de gerenciamento responsáveis pelo processamento, armazenamento e organização dos dados gerados pelos dispositivos conectados.

A comunicação entre esses dispositivos é viabilizada por protocolos como MQTT, CoAP, HTTP e DDS (Gamal, 2014), cada um adequado a diferentes necessidades de desempenho, largura de banda e confiabilidade. No entanto, esses protocolos também apresentam vulnerabilidades críticas, como a ausência de criptografia adequada, falhas nos mecanismos de autenticação e autorização, suscetibilidade a ataques *DoS/DDoS*, *buffer overflows* e atualizações de firmware inseguras (Abdulghani et al., 2020). A inexistência de padrões universais, aliada a falhas na gestão de identidades e interfaces de gerenciamento, contribui para o aumento dos riscos.

Além disso, o uso predominante de comunicação sem fio amplia a superfície de ataque, exigindo protocolos mais robustos. Para mitigar essas ameaças, é fundamental incorporar práticas de segurança desde a fase de projeto, incluindo criptografia forte, autenticação robusta, monitoramento contínuo e atualizações periódicas. A proteção da privacidade e a conformidade com regulamentações também se apresentam como desafios relevantes, tornando a segurança na IoT uma preocupação constante e estratégica.

B. Tipos de ameaça

As redes IoT estão sujeitas a uma variedade de ataques que comprometem a disponibilidade, a integridade e a confidencialidade dos sistemas e dados. Entre os principais tipos de ameaça estão os ataques de negação de serviço (DoS/DDoS), que visam interromper a disponibilidade dos serviços ao sobrecarregar dispositivos com tráfego excessivo. Em ataques DoS, um único dispositivo é utilizado para gerar essa sobrecarga, enquanto em ataques DDoS múltiplos dispositivos distribuídos atuam simultaneamente. Técnicas comuns incluem *SYN Flood*, *UDP Fragmentation* e *ICMP Fragmentation*, explorando diferentes vulnerabilidades dos protocolos de rede (Vanitha; Uma; Mahidhar, 2017).

Os ataques de reconhecimento também são recorrentes e geralmente representam a fase inicial de uma invasão. Nessa etapa, o atacante coleta informações sobre a rede e os dispositivos conectados por meio de técnicas como *Ping Sweep*, *OS Scan* e *Port Scan*, possibilitando o mapeamento de ativos, identificação de sistemas operacionais e localização de possíveis brechas exploráveis (Leevy et al., 2021).

A exploração de vulnerabilidades em serviços web é outra ameaça crítica. Ataques como *SQL Injection*, *Command Injection*, *Cross-Site Scripting (XSS)* e *Upload Malicioso* visam falhas em entradas de dados, permitindo a execução de comandos, acesso não autorizado a informações e propagação de *malware* (Luo et al., 2021).

A falsificação de comunicação, ou *spoofing*, ocorre quando um atacante se passa por outro dispositivo na rede para interceptar ou manipular o tráfego. No *ARP Spoofing*, por exemplo, o invasor associa seu endereço MAC a um IP legítimo; no *DNS Spoofing*, registros DNS são alterados para redirecionar usuários a sites maliciosos (NETO SAJJAD DADKHAH, 2023).

Outra ameaça comum são os ataques por força bruta, que consistem em tentativas repetidas e automatizadas de adivinhar

senhas ou chaves de acesso. O ataque por dicionário é um exemplo típico, utilizando listas predefinidas de palavras para obter acesso. Sistemas sem mecanismos de proteção robustos, como limitação de tentativas ou autenticação multifator, são especialmente vulneráveis.

Um caso emblemático é o malware *Mirai*, conhecido por realizar ataques DDoS em larga escala a partir de *botnets* compostas por dispositivos IoT comprometidos. O Mirai explora credenciais fracas ou padrões de fábrica para infectar dispositivos, que então passam a ser controlados remotamente para gerar grandes volumes de tráfego malicioso. Suas variantes, como *GREIP*, *GREETH* e *UDP Plain*, empregam diferentes técnicas para sobrecarregar redes, explorando falhas específicas nos protocolos de comunicação. Esses ataques ilustram a gravidade das ameaças em ambientes IoT e justificam a adoção de abordagens baseadas em aprendizado de máquina para sua detecção e mitigação.

A Tabela I apresenta um resumo dos principais tipos de ataques discutidos nesta seção, juntamente com suas consequências mais relevantes. Essa sistematização permite visualizar, de forma compacta, os impactos potenciais de cada ameaça no contexto de redes IoT, servindo como base para a definição de estratégias de detecção e mitigação.

Tabela I
PRINCIPAIS TIPOS DE ATAQUES EM IOT E SUAS CONSEQUÊNCIAS

Tipo de Ataque	**Consequência Principal**
DoS/DDoS	Interrupção de serviços, sobrecarga e paralisação de dispositivos.
Reconhecimento	Exposição da topologia da rede e identificação de vulnerabilidades.
Exploração de Serviços Web	Acesso indevido, execução de comandos e vazamento de dados.
Spoofing (ARP/DNS)	Interceptação, redirecionamento e manipulação de tráfego.
Força Bruta	Quebra de senhas, acesso não autorizado e invasão de sistemas.
Botnets (Mirai)	Controle remoto e lançamento de ataques DDoS massivos.

III. APRENDIZADO DE MÁQUINA

A crescente complexidade das ameaças cibernéticas em ambientes de Internet das Coisas (IoT) tem evidenciado as limitações dos métodos tradicionais de segurança. Nesse contexto, o aprendizado de máquina emerge como uma abordagem estratégica, oferecendo meios para identificar padrões anômalos e responder a comportamentos suspeitos em tempo quase real.

Diversos algoritmos têm se destacado na literatura por sua eficácia em tarefas de detecção de ameaças, entre eles o *XGBoost*, *Random Forest*, *Support Vector Machine* (SVM), *LightGBM* e a *Explainable Boosting Machine* (EBM). Cada um desses métodos apresenta características particulares quanto à acurácia, interpretabilidade, custo computacional e adequação a diferentes contextos de segurança.

Contudo, a adoção de modelos preditivos em ambientes IoT envolve desafios relevantes. A transparência nas decisões, a mitigação de vieses algorítmicos e a proteção da privacidade dos dados são aspectos centrais. Soma-se a isso o desbalanceamento comum nos conjuntos de dados de segurança, em que ataques representam uma fração muito pequena dos registros. Essa condição pode comprometer o desempenho dos modelos, exigindo o uso de técnicas de reamostragem e ajustes específicos para garantir maior robustez e confiabilidade às soluções baseadas em ML.

Além disso, a natureza dinâmica dos ambientes IoT demanda modelos capazes de adaptação contínua. Estratégias de aprendizado incremental ou online têm sido exploradas para permitir que os sistemas se atualizem automaticamente à medida que novos dados são coletados, mantendo sua eficácia ao longo do tempo.

A. Machine Learning para Detecção de Ameaças

A aplicação de algoritmos de *Machine Learning* (ML) na detecção de ameaças em ambientes IoT tem se consolidado como uma abordagem eficiente e adaptável. Diferentemente dos métodos tradicionais, os modelos baseados em ML oferecem capacidades avançadas de análise comportamental, detecção de anomalias e previsão de ataques, sendo especialmente úteis em contextos dinâmicos e com alto volume de dados.

Uma das principais vantagens do ML é sua habilidade de adaptação contínua (Angra; Ahuja, 2017). Modelos treinados com dados históricos podem ser atualizados em tempo real à medida que novas informações são coletadas, permitindo respostas ágeis a padrões emergentes. Essa característica é essencial para lidar com a natureza mutável das ameaças em redes IoT.

A detecção de anomalias é um dos pontos fortes do ML na segurança cibernética. Algoritmos especializados conseguem identificar desvios sutis no comportamento de dispositivos ou usuários, sinalizando possíveis intrusões ou falhas operacionais. Técnicas como o *Isolation Forest* têm se mostrado eficazes para isolar padrões atípicos no tráfego de rede (Liu; Ting; Zhou, 2008), contribuindo para a identificação precoce de dispositivos comprometidos.

Além disso, o ML permite a previsão de ameaças com base em padrões históricos. Modelos preditivos podem antecipar comportamentos maliciosos e viabilizar ações preventivas antes da ocorrência efetiva de ataques. Essa abordagem proativa fortalece a resiliência dos sistemas IoT diante de agentes sofisticados.

Outro diferencial importante é a capacidade de análise sobre grandes volumes e múltiplas fontes de dados. Ao explorar correlações complexas entre variáveis, os algoritmos extraem insights relevantes para a compreensão do comportamento do sistema como um todo. No entanto, desafios como o controle de falsos positivos e a curadoria dos dados de entrada ainda exigem atenção.

Por fim, destaca-se o uso de modelos de aprendizado profundo, como redes neurais e arquiteturas baseadas em LSTM, em sistemas de detecção de ameaças avançadas. Esses modelos são capazes de capturar padrões altamente complexos e sutis,

proporcionando um nível adicional de proteção em ambientes críticos.

Ao complementar abordagens tradicionais de segurança, o *Machine Learning* contribui para o desenvolvimento de soluções mais inteligentes, capazes de se adaptar e evoluir continuamente frente a novos desafios. A seção seguinte discute os algoritmos mais aplicados à detecção de ameaças, com ênfase em suas características técnicas e desempenho em cenários de IoT.

B. Algoritmos de Aprendizado de Máquina Utilizados

Foram selecionados os algoritmos *XGBoost*, *Random Forest*, *SVM*, *LightGBM* e *EBM*, devido à sua reconhecida eficácia em tarefas de classificação, robustez contra sobreajuste e histórico de aplicação bem-sucedida em problemas de detecção de anomalias, incluindo segurança em redes IoT. A seguir, apresenta-se uma breve descrição de cada técnica.

XGBoost (Extreme Gradient Boosting) é uma biblioteca otimizada de *gradient boosting*, baseada em árvores de decisão construídas sequencialmente (Chen, 2016). Destaca-se pela alta performance, suporte à paralelização e mecanismos de regularização (*L1* e *L2*), que reduzem o *overfitting*. Além disso, lida eficientemente com dados ausentes, suporta validação cruzada e é amplamente adotado em competições de ciência de dados.

Random Forest é um método de *ensemble* que combina diversas árvores de decisão treinadas com subconjuntos aleatórios dos dados. As decisões são tomadas por maioria (classificação) ou média (regressão). É conhecido por sua boa capacidade de generalização, desempenho estável em dados desbalanceados e resistência ao *overfitting*.

SVM (*Support Vector Machine)* busca encontrar um hiperplano ótimo que maximize a margem entre classes (Cortes; Vapnik, 1995). É particularmente eficaz em espaços de alta dimensionalidade e pode utilizar funções *kernel* para lidar com dados não linearmente separáveis. Por isso, é adequado para cenários como a classificação de tráfego em redes IoT.

LightGBM é uma biblioteca de *gradient boosting* desenvolvida com foco em eficiência e escalabilidade (Ke et al., 2017). Utiliza técnicas como GOSS e EFB para acelerar o treinamento e reduzir o uso de memória, além de empregar crescimento baseado em folhas, o que proporciona convergência mais rápida. Suporta variáveis categóricas de forma nativa e é eficaz em grandes volumes de dados.

EBM (*Explainable Boosting Machine)* é um modelo baseado em *boosting* com ênfase na interpretabilidade (Interpretml, 2023). Adota uma estrutura aditiva e monotônica que facilita a explicação do impacto de cada variável. Além disso, é robusto contra *overfitting* e apropriado para cenários em que transparência na tomada de decisão é essencial.

Esses algoritmos foram avaliados quanto à sua eficácia na detecção de tráfego malicioso em redes IoT, conforme descrito nas seções seguintes. Cada modelo apresenta características distintas em termos de desempenho, interpretabilidade e custo computacional, o que influencia diretamente sua adequação a diferentes cenários de segurança em ambientes IoT. A análise a seguir discute essas diferenças, considerando tanto os resultados obtidos quanto os compromissos envolvidos em sua aplicação prática.

C. Desbalanceamento de Dados em Datasets de IoT: Desafios e Técnicas de Mitigação

Dados desbalanceados ocorrem quando uma das classes, geralmente a de interesse, como a classe minoritária, possui significativamente menos amostras que as demais. Essa condição é comum em contextos como diagnósticos médicos, detecção de fraudes e filtragem de spam, e também se manifesta fortemente em ambientes de segurança em redes IoT.

O desbalanceamento pode comprometer a capacidade dos modelos de aprendizado de máquina, levando ao chamado Paradoxo da Acurácia (Azank, 2020), em que modelos com alta acurácia geral falham na identificação correta da classe minoritária. Por isso, torna-se essencial empregar métricas mais sensíveis, como *precision*, *recall*, F1-score e matriz de confusão, para uma avaliação mais precisa do desempenho.

Em redes IoT, esse problema é particularmente recorrente, dado que o tráfego malicioso representa apenas uma fração do volume total de dados, o que tende a enviesar os modelos em favor da classe majoritária (Pei et al.,2023) e comprometer a eficácia na detecção de ameaças.

Diversas técnicas têm sido aplicadas para mitigar esse desequilíbrio (Lahera, 2019):

- ***Oversampling***: replica instâncias da classe minoritária para aumentar sua representação no treinamento, sem alterar a proporção total de dados.
- ***Undersampling***: reduz o número de exemplos da classe majoritária, equilibrando o conjunto de dados, mas com possível perda de informação.
- ***Variational Autoencoders* (VAE)**: utilizam redes neurais para gerar amostras sintéticas da classe minoritária, promovendo maior diversidade nos dados.
- **SMOTE (*Synthetic Minority Over-sampling Technique*)**: cria exemplos artificiais por interpolação entre instâncias reais da classe minoritária, sendo uma das abordagens mais difundidas.

A escolha da técnica mais adequada depende do cenário. Em casos de escassez severa de dados maliciosos, como no dataset utilizado neste trabalho, com classificação binária entre tráfego malicioso e benigno, o uso combinado dessas estratégias pode aumentar significativamente a robustez do modelo e sua capacidade de generalização.

Em síntese, a aplicação estratégica de algoritmos de *Machine Learning* na segurança de redes IoT permite enfrentar desafios como o desbalanceamento de dados e implementar soluções adaptativas e proativas. Ao aprender continuamente com fluxos de dados dinâmicos, esses modelos oferecem uma camada adicional de proteção, aprimorando a detecção de comportamentos anômalos e fortalecendo a resiliência dos sistemas.

Na próxima seção, será apresentada a metodologia adotada neste trabalho, detalhando a aplicação prática dos modelos de aprendizado de máquina discutidos anteriormente, com o

objetivo de avaliar sua eficácia na classificação de ameaças em ambientes IoT.

IV. METODOLOGIA

Este estudo baseia-se no treinamento e avaliação de cinco modelos de aprendizado de máquina para a classificação de tráfego benigno e malicioso em redes de Internet das Coisas (IoT). Inicialmente, os modelos são treinados com um conjunto de dados fornecido pelo Instituto Canadense de Cibersegurança. Em seguida, aplica-se a técnica de *oversampling* para balanceamento das classes, com o objetivo de aprimorar a capacidade de detecção das ameaças. O desempenho dos modelos é analisado com base nas métricas discutidas na fundamentação teórica. Todo o código-fonte utilizado nos experimentos está disponível publicamente no repositório GitHub de Juan Rodrigues (2023).

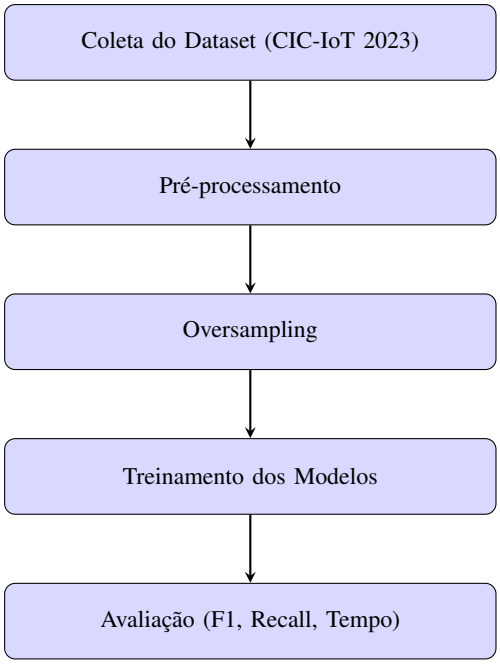

Figura 1. Fluxograma da metodologia adotada

A. Banco de Dados

O conjunto de dados utilizado neste trabalho é o CIC-IoT 2023, disponibilizado pelo Canadian Institute for Cybersecurity (CIC). Esse dataset foi gerado em ambientes controlados, simulando cenários de redes IoT com dispositivos reais atuando como fontes de tráfego benigno e agentes de ataque (NETO SAJJAD DADKHAH, 2023). O conjunto de dados contém aproximadamente 46,6 milhões de registros, dos quais cerca de 45,5 milhões correspondem a tráfego malicioso e apenas 1,1 milhão a tráfego legítimo (Tabela II).

Cada registro é descrito por 47 atributos, incluindo estatísticas de pacotes, características dos protocolos, taxas de transmissão e outras métricas extraídas do tráfego de rede.

Tabela II
CATEGORIA DE ATAQUES

Categoria	Registros
DDoS	33.984.560
DoS	8.090.738
Mirai	2.634.124
Benign	1.098.195
Spoofing	486.504
Recon	354.565
Web	24.829
BruteForce	13.064

Essa diversidade de atributos permite construir modelos capazes de capturar padrões maliciosos em diferentes camadas do protocolo.

Apesar da predominância artificial de ataques no dataset, essa distribuição foi intencionalmente induzida no processo de coleta para garantir representatividade dos diversos tipos de ameaças. Em redes reais, a maioria do tráfego é benigno, o que reforça a necessidade de empregar estratégias de balanceamento de classes para simular tais condições. Para fins de validação, o conjunto de dados foi particionado em 70% para treinamento e 30% para teste, mantendo a distribuição original das classes.

B. Balanceamento de Dados

Devido ao desbalanceamento entre as classes no conjunto de dados, com predominância de registros de tráfego malicioso em relação ao tráfego benigno, foi aplicada a técnica de *oversampling* por replicação direta das instâncias da classe minoritária. Essa estratégia visa atenuar o viés do processo de aprendizado e aumentar a sensibilidade dos modelos na identificação de comportamentos maliciosos. O procedimento foi restrito ao conjunto de treinamento, a fim de preservar a imparcialidade da avaliação no conjunto de teste.

Optou-se pelo *oversampling* simples em vez de técnicas sintéticas, como o SMOTE ou métodos baseados em VAE (descritos na seção de Revisão Teórica), considerando a necessidade de preservar a integridade das amostras reais e a simplicidade do método em contextos de alta dimensionalidade.

Após o balanceamento, o conjunto de treinamento passou a apresentar uma proporção equitativa entre as classes, favorecendo uma aprendizagem mais justa e possibilitando que os algoritmos considerassem de forma mais equilibrada os padrões de ambos os tipos de tráfego. Essa etapa mostrou-se fundamental para a melhoria da sensibilidade (*recall*) e da métrica F1-Score dos modelos, como será apresentado nos resultados.

C. Modelagem

A etapa de modelagem consistiu na aplicação de cinco algoritmos de aprendizado de máquina para a tarefa de classificação binária do tráfego de rede: *XGBoost*, *Random Forest*, *Support Vector Machine* (SVM), *LightGBM* e *Explainable Boosting Machine* (EBM). Esses modelos foram selecionados com base em seu desempenho consolidado em

tarefas de classificação e por suas diferentes características em termos de interpretabilidade, custo computacional e robustez, conforme discutido na fundamentação teórica.

D. Avaliação dos Modelos

Para a avaliação do desempenho dos modelos, foram utilizadas as métricas de *precision*, *recall* e F1-score, por apresentarem maior sensibilidade à distribuição desigual entre classes, uma característica marcante em conjuntos de dados com prevalência de tráfego malicioso, como no presente estudo. A adoção dessas métricas busca mitigar os efeitos do chamado *Paradoxo da Acurácia*, fenômeno em que modelos com alta acurácia global podem falhar na detecção da classe minoritária, resultando em taxas elevadas de falsos negativos. No contexto de segurança em redes IoT, essa limitação é especialmente crítica, pois compromete a identificação eficaz de ameaças.

A acurácia, embora amplamente utilizada em tarefas de classificação, tende a ser enganosa em cenários com classes desbalanceadas, pois pode mascarar um viés do modelo em favor da classe majoritária. Em contrapartida, a métrica de *precision* avalia a confiabilidade das detecções — isto é, a proporção de casos classificados como ataque que realmente correspondem a ameaças.

Já o *recall* mede a capacidade do modelo em identificar corretamente todas as instâncias da classe minoritária, sendo, portanto, essencial em sistemas de segurança. Nestes casos, uma baixa taxa de *recall* pode significar a não detecção de ataques reais, o que compromete diretamente a eficácia do sistema e aumenta os riscos à integridade da rede.

O F1-score, por sua vez, corresponde à média harmônica entre *precision* e *recall*, oferecendo uma medida única que equilibra os impactos de falsos positivos e falsos negativos. Por combinar essas duas perspectivas, o F1-score é especialmente útil em cenários desbalanceados, como o analisado neste trabalho, pois permite avaliar o desempenho geral do modelo sem favorecer desproporcionalmente nenhuma das classes.

A aplicação de técnicas de balanceamento, como o oversampling, tende a aumentar o *recall*, reduzindo os falsos negativos, mas pode levar a uma queda na *precision*, devido ao aumento de falsos positivos. O F1-score reflete esse trade-off de forma mais equilibrada, sendo particularmente valioso para comparar o desempenho entre os diferentes modelos sob cenários realistas de desbalanceamento.

V. RESULTADOS

A avaliação dos modelos de aprendizado de máquina foi realizada considerando os cenários com e sem aplicação de oversampling. As métricas analisadas foram acurácia, precisão, recall, F1-score e tempo de treinamento.

A Tabela III resume o desempenho dos cinco modelos de aprendizado de máquina avaliados, considerando os cenários com e sem aplicação de oversampling. De forma geral, observa-se que o uso dessa técnica contribuiu para o aumento do *recall* em modelos como o Random Forest, indicando uma redução na taxa de falsos negativos, especialmente relevantes neste estudo, em que o tráfego benigno representa a classe minoritária e tende a ser sub-representado nos processos de aprendizado.

No entanto, conforme esperado em bases fortemente desbalanceadas, esse ganho em *recall* foi acompanhado por uma queda na *precision*, indicando um aumento na taxa de falsos positivos, ou seja, instâncias benignas classificadas erroneamente como maliciosas. Esse trade-off reforça a importância da utilização do F1-score como métrica de equilíbrio, especialmente em contextos em que é necessário ponderar tanto a capacidade de detecção de comportamentos legítimos quanto a confiabilidade das decisões tomadas pelo modelo.

O modelo SVM apresentou comportamento inverso aos demais: após o balanceamento, houve queda no *recall* e aumento na *precision*, tornando-se mais seletivo, com menos alarmes falsos, mas menos sensível à detecção de tráfego malicioso, o que é indesejável em sistemas de segurança. Além disso, seu tempo de treinamento foi significativamente superior ao dos outros modelos, especialmente com oversampling, o que limita sua aplicação em cenários com restrições operacionais.

LightGBM e XGBoost demonstraram bom desempenho geral, com boa estabilidade entre os cenários. Ambos mantiveram F1-scores elevados, com tempos de execução reduzidos e variações moderadas entre *precision* e *recall*, o que os torna modelos atrativos para aplicações em tempo real. Já o modelo EBM, embora apresente menor *precision* no cenário sem balanceamento, destacou-se por sua interpretabilidade e consistência após o oversampling, sendo adequado para contextos onde a explicabilidade é fator crítico.

Esses resultados reforçam a importância de escolher métricas adequadas para avaliação em bases desbalanceadas. Em ambientes IoT, onde a prioridade muitas vezes recai sobre a detecção de ameaças (isto é, *recall*), o uso de técnicas de balanceamento e a análise cuidadosa do F1-score tornam-se estratégias indispensáveis para construir sistemas mais confiáveis e efetivos.

A Figura 2 apresenta a comparação do F1-score entre os modelos avaliados, nos cenários com e sem oversampling. Observa-se que todos os algoritmos mantiveram desempenho elevado, com variações pontuais. O Random Forest apresentou o F1-score mais consistente entre os dois cenários, demonstrando robustez frente ao desbalanceamento dos dados. O XGBoost e o LightGBM também mantiveram desempenho estável, com leve oscilação entre recall e precisão.

O SVM, por outro lado, apresentou queda significativa no F1-score após o oversampling, reflexo de uma redução expressiva no recall, mesmo com o aumento da precisão. Já o EBM teve desempenho mais equilibrado, com pequena variação entre os cenários e a vantagem adicional da interpretabilidade do modelo. Esse gráfico resume, de forma visual, a eficácia relativa dos algoritmos sob diferentes condições de balanceamento, sendo uma ferramenta útil para comparar desempenho geral em tarefas de detecção binária em ambientes IoT.

A Figura 3 ilustra o trade-off entre *precision* e *recall* para cada modelo avaliado, comparando os resultados antes e depois da aplicação do oversampling. Cada par de pontos representa o deslocamento nos indicadores causado pela

Tabela III
DESEMPENHO DOS MODELOS COM E SEM OVERSAMPLING

Modelos	Oversampling	Acurácia (%)	Precisão (%)	Recall (%)	F1-score (%)	Tempo de Treinamento
Random Forest	Não	99,68	96,45	96,67	96,56	56 min
Random Forest	Sim	99,61	93,50	98,92	96,04	6h 48min
SVM	Não	99,25	94,66	90,11	92,25	8h 32min
SVM	Sim	98,78	99,23	82,98	89,38	42h 14min
XGBoost	Não	99,63	96,74	95,33	96,02	10min 37s
XGBoost	Sim	99,60	98,03	93,81	95,82	21min 40s
LightGBM	Não	99,57	96,48	94,49	95,46	8min 36s
LightGBM	Sim	99,57	98,42	93,15	95,62	18min 05s
EBM	Não	99,51	80,96	93,91	94,86	4h 05min
EBM	Sim	99,43	98,51	90,98	94,41	15h 18min

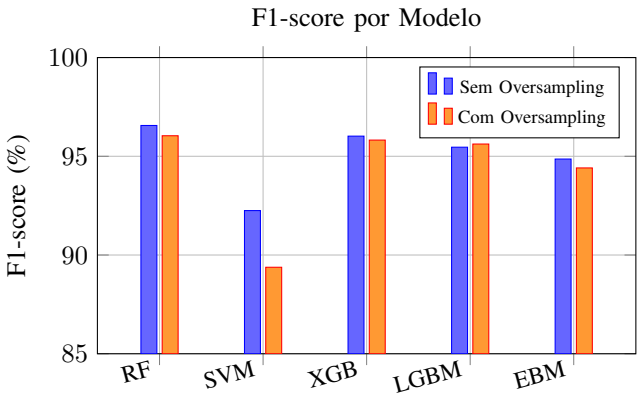

Figura 2. F1-score com e sem oversampling para cada modelo avaliado

técnica de balanceamento. De forma geral, observa-se que o oversampling elevou o *recall* em alguns modelos, como o Random Forest, indicando uma maior capacidade de detecção de tráfego malicioso, o que se traduz em redução da taxa de falsos negativos. Esse ganho, no entanto, veio acompanhado de uma queda na *precision*, refletindo o aumento de falsos positivos.

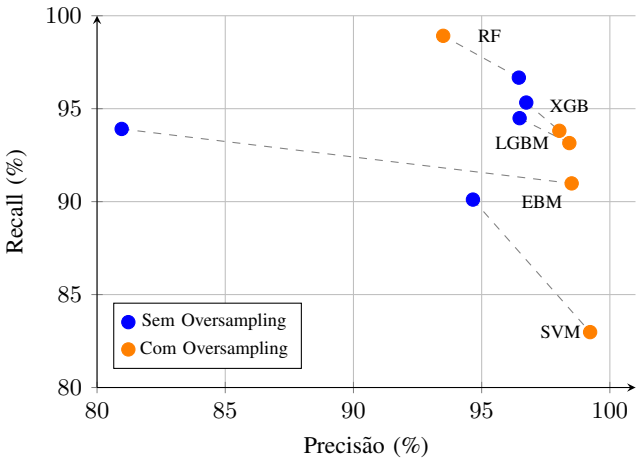

Figura 3. Trade-off entre precisão e recall com e sem oversampling para os modelos avaliados. As linhas tracejadas conectam os pares de pontos de cada modelo.

No caso do SVM, o oversampling gerou efeito inverso: aumentou expressivamente a *precision*, mas reduziu o *recall*. O modelo tornou-se mais conservador, com menos alarmes falsos, mas também menos capaz de detectar ataques reais, um risco relevante em contextos de segurança.

Os modelos LightGBM e XGBoost demonstraram deslocamentos mais discretos entre os dois cenários, mantendo um bom equilíbrio entre precisão e sensibilidade. O EBM, por sua vez, apresentou um aumento notável na *precision* após o balanceamento, com uma leve redução no *recall*, sustentando um perfil de modelo estável e interpretável.

Esses resultados reforçam a importância de avaliar múltiplas métricas em conjunto, especialmente em contextos de alta criticidade como redes IoT. Nesses casos, perdas em *recall* podem comprometer a eficácia do sistema, tornando necessário priorizar soluções que garantam a sensibilidade na detecção de ameaças, ainda que isso implique em um aumento controlado de falsos positivos.

A Figura 4 apresenta o trade-off entre F1-score e tempo de treinamento dos modelos avaliados, utilizando escala logarítmica para melhor representar as diferenças entre os algoritmos. Cada modelo aparece com dois pontos conectados, indicando os resultados nos cenários com e sem oversampling.

Observa-se que o *Random Forest* manteve um F1-score elevado, mas com aumento expressivo no tempo de treinamento após o balanceamento. O *SVM* foi o mais impactado, com tempo significativamente superior aos demais e redução no desempenho. Por outro lado, os modelos *LightGBM* e *XGBoost* demonstraram bom equilíbrio entre desempenho e custo computacional, mantendo F1-scores altos com tempos de execução reduzidos. Já o *EBM*, embora mais custoso em termos de tempo, apresentou desempenho competitivo e a vantagem da interpretabilidade.

VI. CONCLUSÃO

A análise comparativa dos cinco modelos avaliados demonstrou que o oversampling impacta de forma distinta cada algoritmo, afetando tanto o desempenho quanto o custo computacional. No caso do Random Forest, a aplicação do balanceamento elevou significativamente o *recall*, indicando uma melhora na identificação do tráfego benigno, classe minoritária e mais difícil de detectar, embora com uma leve queda na *precisão*, sinalizando aumento nos falsos positivos.

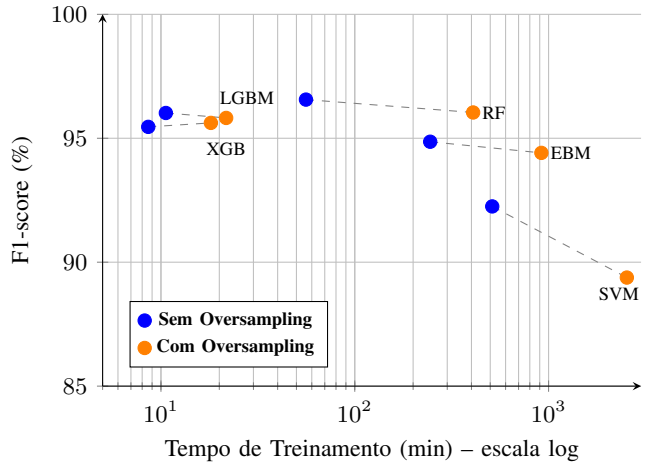

Figura 4. Trade-off entre F1-score e tempo de treinamento (escala logarítmica). Cada modelo aparece com dois pontos conectados, representando os cenários com e sem oversampling.

Em contrapartida, nos demais modelos, como SVM, XGBoost e EBM, observou-se o comportamento inverso: a *precisão* aumentou, mas com redução no *recall*, sugerindo que se tornaram mais conservadores, privilegiando a assertividade nas classificações ao custo de uma menor sensibilidade para identificar o tráfego benigno.

O oversampling aumentou consideravelmente o tempo de treinamento em todos os modelos, sendo o SVM o mais afetado. Esse custo computacional elevado é relevante em bases desbalanceadas como a usada neste estudo, em que a classe minoritária (tráfego benigno) representa uma pequena fração. Assim, embora o oversampling melhore a representatividade dessa classe, seu uso deve equilibrar ganhos em desempenho e tempo de processamento.

Entre os algoritmos, o Random Forest destacou-se pela boa capacidade de detecção (elevado *recall*), sendo ideal para cenários que priorizam sensibilidade a ameaças. O LightGBM combinou bom desempenho com baixo custo computacional, enquanto o EBM se sobressaiu pela interpretabilidade, sendo adequado a contextos que exigem decisões transparentes, como auditorias e ambientes regulatórios.

Conclui-se que o uso de aprendizado de máquina, aliado a técnicas adequadas de balanceamento, pode fortalecer significativamente a segurança em redes IoT. No entanto, é fundamental equilibrar precisão, sensibilidade, interpretabilidade e eficiência computacional. Como perspectivas futuras, recomenda-se a investigação de técnicas mais sofisticadas de balanceamento, como SMOTE e *Variational Autoencoders*, além da aplicação de estratégias de paralelização e otimização para redução do tempo de treinamento.

Adicionalmente, os resultados indicam recomendações práticas para cada modelo, considerando diferentes contextos. O Random Forest é indicado para ambientes hospitalares ou infraestruturas críticas, onde a sensibilidade na detecção de ameaças é prioritária. O LightGBM, com bom desempenho e baixo custo computacional, é adequado para ambientes residenciais ou dispositivos com limitações de processamento. Já o EBM destaca-se em sistemas industriais ou corporativos que exigem interpretabilidade e transparência, como em setores regulados.

Referências

[1] R. M. ABDULGHANI, M. M. ALREHILI, A. A. ALMUHANNA, and O. ALHAZMI, "Vulnerabilities and security issues in IoT protocols," pp. 7-12, 2020.

[2] S. ANGRA and S. AHUJA, "Machine learning and its applications: A review," pp. 57–60, 2017.

[3] F. AZANK, "Paradoxo da acurácia. Um conceito fundamental para sua jornada em Ciência de Dados," 2020. [Online]. Available: https://medium.com/p/56baa75334f2

[4] T. CHEN and C. E. GUESTRIN, "XGBoost: A scalable tree boosting system," *ACM*, 2016. [Online]. Available: https://arxiv.org/pdf/1603.02754

[5] L. CONSTANTIN, "Hackers found 47 new vulnerabilities in 23 IoT devices at DEF CON," 2016. [Online]. Available: https://www.csoonline.com/article/557921/hackers-found-47-new-vulnerabilities-in-23-iot-devices-at-def-con.html

[6] C. CORTES and V. VAPNIK, "Random forests," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[7] E. C. PINTO NETO, S. DADKHAH, R. FERREIRA, A. ZOHOURIAN, R. LU, and A. A. GHORBANI, "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," *Sensors*, vol. 23, no. 13, p. 5941, 2023.

[8] M. MANSOUR, A. GAMAL, A. I. AHMED, L. A. SAID, A. ELBAZ, N. HERENCSAR, and A. SOLTAN, "Internet of Things: A comprehensive overview on protocols, architectures, technologies, simulation tools, and future directions," *Energies*, vol. 16, no. 8, p. 3465, 2014.

[9] INTERPRETML, "Explainable Boosting Machine," 2023. [Online]. Available: https://interpret.ml/docs/ebm.html

[10] G. KE, Q. MENG, T. FINLEY, T. WANG, W. CHEN, W. MA, Q. YE, and T. LIU, "LightGBM: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[11] C. V. KRISHNAVENI and T. S. RANI, "On the classification of imbalanced datasets," *International Journal of Computer Science and Technology*, vol. 2, no. 1, pp. 2229–4333, 2011.

[12] G. LAHERA, "Unbalanced datasets: What to do about them," 2019. [Online]. Available: https://medium.com/strands-tech-corner/unbalanced-datasets-what-to-do-144e0552d9cd

[13] J. L. LEEVY, J. HANCOCK, T. M. KHOSHGOFTAAR, and N. SELIYA, "IoT reconnaissance attack classification with random undersampling and ensemble feature selection," in *2021 IEEE 7th International Conference on Collaboration and Internet Computing (CIC)*, pp. 41–49, 2021.

[14] F. T. LIU, K. M. TING, and Z. H. ZHOU, "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*, pp. 413–422, 2008.

[15] C. LUO, M. TANG, Y. CHEN, and M. QIU, "A novel web attack detection system for Internet of Things via ensemble classification," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5810–5818, 2021.

[16] W. PEI, B. XUE, M. ZHANG, L. SHANG, X. YAO, and Q. ZHANG, "A survey on unbalanced classification: How can evolutionary computation help?," *IEEE Transactions on Evolutionary Computation*, pp. 1–1, 2023.

[17] J. RODRIGUES, "DatasetIoT," 2023. [Online]. Available: https://github.com/juanr-0/DatasetIoT

[18] S. SINHA, "State of IoT 2023: Number of connected IoT devices growing 16% to 16.7 billion globally," 2023. [Online]. Available: https://iot-analytics.com/number-connected-iot-devices/

[19] K. VANITHA, S. V. UMA, and S. MAHIDHAR, "Distributed denial of service: Attack techniques and mitigation," *IEEE*, vol. 2, no. 1, pp. 226–231, 2017.

[20] A. ZANELLA, N. BUI, A. CASTELLANI, L. VANGELISTA, and M. ZORZI, "Internet of Things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.