Uma Avaliação das Características de Redes Wi-Fi(802.11) para a Detecção de Ataques de Personificação baseada em Inteligência Artificial Explicável – XAI

Elton Henrique Lunardi Gimenes

Universidade Federal de Uberlândia
Faculdade de Computação
Bacharelado em Sistemas de Informação

Monte Carmelo - MG
2024

Elton Henrique Lunardi Gimenes

Uma Avaliação das Características de Redes Wi-Fi(802.11) para a Detecção de Ataques de Personificação baseada em Inteligência Artificial Explicável – XAI

Trabalho de Conclusão de Curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Sistemas de Informação.

Área de concentração: Segurança da Informação

Orientador: Prof. Dr. Adriano Mendonça Rocha
Coorientador: Prof. Dr. Silvio Ereno Quincozes

Monte Carmelo - MG
2024

Este trabalho é dedicado ao amor da minha vida.

"Há três espécies de cérebros: uns entendem por si próprios; os outros discernem o que os primeiros entendem; e os terceiros não entendem nem por si próprios nem pelos outros; os primeiros são excelentíssimos; os segundos excelentes; e os terceiros totalmente inúteis."

(Maquiavel)

Resumo

A eficiência na detecção de ataques de personificação em redes Wi-Fi no padrão IEEE 802.11 é crucial para a segurança de redes. Este trabalho explora o uso da Inteligência artificial explicável - (XAI) para avaliar a eficácia de diferentes conjuntos de características na detecção desses ataques, utilizando a ferramenta SHapley Additive exPlanations (SHAP) para investigar a contribuição individual de cada característica. Com base nos experimentos que envolvem 15 conjuntos de características e a aplicação do classificador eXtreme Gradient Boosting (XGBoost), identificamos que características como *frame.len*, *wlan.fc.subtype* e *wlan.duration* são vitais tanto para o tráfego normal quanto para ataques específicos como *Cafe Latte* e *Evil Twin*. A análise SHAP revela que certas características têm um papel variável, sendo associadas tanto positiva quanto negativamente com a presença de ataques, o que promove maior transparência e confiança nas decisões dos sistemas de detecção de intrusões e ajuda na identificação e prevenção de atividades maliciosas na rede.

Palavras-chave: Sistema de Detecção de Intrusões, Inteligência Artificial Explicável, SHapley Additive exPlanations (SHAP), Ataque de Personificação, Segurança em Redes Wi-Fi.

Lista de siglas

IA Inteligência Artificial - *Artificial Inteligence*

AP Acess Point - *Ponto de Acesso*

ARP Address Resolution Protocol - *Protocolo de Resolução de Endereços*

BIAS Bluetooth Impersonation AttacksS

IDS Sistema de Detecção de Instrusões - *Intrusion Detection Systems*

IP Internet Protocol address - *Endereço de Protocolo da Internet*

ML Machine Learning

SHAP SHapley Additive exPlanations

SSID Service Set Identifier - *Identificador de Conjunto de Serviços*

WEP Wired Equivalent Privacy

WPA2 Wi-Fi Protected Access 2

WPA3 Wi-Fi Protected Access 3

WLAN Wireless Local Area Network

VPN Virtual Private Network - *Rede privada virtual*

XGBoost eXtreme Gradient Boosting

XAI Inteligência artificial explicável - *Explainable Artificial Intelligence*

Sumário

Capítulo 1

Introdução

A segurança em redes Wi-Fi (802.11) é um campo de constante evolução e de crítica importância, especialmente diante da crescente dependência de indivíduos e organizações em conexões sem fio. Os ataques de personificação, particularmente, representam uma ameaça significativa nesse ambiente. Esses ataques, onde um adversário imita uma entidade legítima na rede, comprometem a autenticidade das comunicações, permitindo atividades maliciosas como interceptação de dados, disseminação de *malware* ou negação de serviço. A detecção eficaz desses ataques é crucial para manter a confiança e a integridade das redes Wi-Fi (802.11).

No contexto atual, as técnicas de Inteligência Artificial - (IA) têm desempenhado um papel fundamental na melhoria da precisão e eficiência dos sistemas de detecção de intrusões. Entretanto, a natureza opaca de muitos modelos de IA cria um desafio em termos de confiabilidade e compreensibilidade das decisões tomadas por esses sistemas. Aqui, o Sistema de Detecção de Instrusões - (IDS) oferece uma solução promissora ao proporcionar modelos de IA não apenas eficientes, mas também interpretáveis e transparentes.

Focando especificamente nos ataques de personificação em redes Wi-Fi (802.11), este trabalho emprega técnicas de XAI para avaliar e interpretar conjuntos de características utilizados na detecção desses ataques. Utilizando o algoritmo XGBoost e a ferramenta SHAP, investiga-se a influência de diferentes características na precisão do modelo de detecção, visando não apenas aprimorar a capacidade de detecção, mas também a explicabilidade das decisões, o que é essencial para a adoção e confiança em sistemas de detecção de intrusões baseados em IA.

1.1 Motivação

A prevalência de redes Wi-Fi (802.11) em uma multiplicidade de ambientes torna-as alvos atraentes para agentes mal-intencionados. Em particular, ataques de personificação representam uma ameaça crescente, permitindo aos atacantes obter acesso não autorizado a recursos de rede e dados confidenciais. Estudos e relatórios recentes, vistos na Seção 2.4,

sublinham a gravidade e a prevalência crescente desses ataques, reforçando a necessidade urgente de abordagens de detecção mais eficazes e transparentes.

Métodos tradicionais de detecção de intrusões estão se tornando insuficientes, dada a complexidade e sofisticação crescentes dos ataques de personificação. A integração da IA e do Machine Learning (ML) nas estratégias de segurança de redes Wi-Fi (802.11) oferece um caminho promissor para melhorar a detecção de intrusões. No entanto, a adoção dessas tecnologias avançadas traz consigo um desafio significativo: a falta de transparência. Modelos de IA, especialmente aqueles baseados em aprendizado profundo, atuam frequentemente como *caixas-pretas*, onde as decisões são tomadas de forma que os usuários finais e os especialistas em segurança não conseguem entender facilmente. Esta opacidade pode ser problemática, especialmente em um contexto de segurança cibernética, onde compreender o porquê por trás de uma detecção ou classificação é crucial para a tomada de decisões informadas, a adequação de respostas e a mitigação de vulnerabilidades. Além disso, a falta de explicabilidade pode dificultar a identificação de erros nos modelos, resultando em falsos positivos ou falsos negativos que comprometem a eficácia do sistema de detecção de intrusões.

O XAI emerge como uma abordagem inovadora para abordar esses desafios, capacitando os sistemas de detecção com a habilidade de explicar suas decisões. Por meio de técnicas explicativas, é possível desvendar a lógica por trás das previsões dos modelos, oferecendo *insights* que não apenas aumentam a confiança dos usuários, mas também facilitam a otimização dos modelos e a correção de eventuais falhas. Assim, o XAI não só melhora a transparência e a responsabilidade dos sistemas de IA aplicados à segurança cibernética, mas também contribui para a evolução desses sistemas em direção a uma maior precisão e confiabilidade.

Este estudo é motivado pela necessidade imperativa de aprimorar a segurança em redes Wi-Fi (802.11), adotando o XAI para fornecer não apenas uma detecção eficaz, mas também um entendimento claro dos processos de tomada de decisão dos modelos de IA. A relevância da investigação é ancorada por relatórios recentes que destacam a sofisticação crescente dos ataques de personificação:

- *The420.in* reporta um aumento significativo nos ataques de *whaling*, demonstrando a evolução e a sofisticação dos ataques de personificação e sua relevância para a segurança em redes corporativas (*"CEO Impersonation Strikes: Understanding the Growing Threat of Whaling Attacks"*, 2023). O *whaling* é um tipo de ataque de *phishing* direcionado que visa indivíduos de alto nível em organizações, como CEOs, procurando enganá-los para obter ganhos financeiros ou informações confidenciais.

- *DarkReading* destaca ataques recentes à cadeia de suprimentos, onde agressores se passaram pelo Dependabot do GitHub, ilustrando a complexidade e o potencial impacto desses ataques na integridade dos sistemas (*"Supply Chain Attackers Escalate*

With GitHub Dependabot Impersonation", 2023). Esse tipo de ataque representa uma forma sofisticada de personificação, onde o adversário se passa por um serviço ou indivíduo confiável para inserir código malicioso em projetos de *software*.

Além desses ataques contemporâneos e sofisticados, este estudo também abordará ataques de personificação mais tradicionais, mas não menos perigosos, como *evil twin*, *cafe-latte*, e *hirte*:

- O ataque *evil twin* envolve a criação de um ponto de acesso Wi-Fi (802.11) malicioso que imita um ponto legítimo, enganando os usuários para que se conectem a ele, possibilitando ao atacante interceptar informações sensíveis transmitidas pela vítima.
- No ataque *cafe-latte*, o adversário cria um ponto de acesso Wi-Fi (802.11) para interceptar as mensagens de autenticação Wired Equivalent Privacy (WEP) de uma vítima, permitindo-lhe decifrar rapidamente a chave WEP e ganhar acesso à rede.
- O ataque *hirte* é uma variação do *cafe-latte*, onde o atacante também visa obter a chave WEP, mas utiliza uma abordagem diferente para induzir a vítima a enviar uma grande quantidade de tráfego, facilitando a quebra da chave.

Esses tipos de ataques, embora mais conhecidos, continuam evoluindo e representam uma ameaça significativa em ambientes de redes Wi-Fi (802.11). A inclusão desses ataques no escopo da pesquisa reitera a abrangência e a atualidade do problema, sublinhando a importância de desenvolver métodos de detecção que sejam não apenas eficazes, mas também transparentes e compreensíveis, reforçando a segurança e a confiança nas decisões tomadas pelos sistemas de IA.

1.2 Justificativa

Embora a detecção de ataques de personificação em redes Wi-Fi (802.11) tenha sido extensivamente abordada em trabalhos como os de Aminanto e Kim (2016) e Aminanto e Kim (2017), a dimensão da explicabilidade nas técnicas de detecção ainda é um campo pouco explorado. Enquanto Aminanto e Kim (2016) utilizaram técnicas de aprendizado profundo para otimizar a detecção de ataques de personificação, focando principalmente na eficiência da seleção de características através de redes neurais e *autoencoders*, eles não forneceram detalhes suficientes sobre a interpretabilidade das características selecionadas, limitando a compreensão dos processos decisórios do modelo.

Aminanto e Kim (2017), por sua vez, investigaram métodos de ponderação de características em técnicas de aprendizado de máquina para detectar ataques de personificação de maneira precisa. Contudo, sua pesquisa focou na eficácia da seleção sem abordar a

transparência do processo decisório, deixando uma lacuna na compreensão de como as características influenciam diretamente os resultados do modelo.

Diante dessas limitações, este trabalho busca avançar no campo da segurança cibernética, integrando técnicas de XAI para elucidar de forma clara a contribuição e relevância das características na detecção de ataques de personificação. A aplicação de XAI visa não apenas aprimorar a precisão da detecção de ataques, mas também proporcionar *insights* profundos sobre os mecanismos de decisão dos modelos, aumentando assim a transparência e confiança nos sistemas de detecção de intrusões.

Neste contexto, os trabalhos de Tu et al. (2021) e Antonioli, Tippenhauer e Rasmussen (2020) demonstram como técnicas emergentes, como aprendizado por reforço e análise de ataques em tecnologias Bluetooth, podem ser integradas para fornecer uma defesa robusta e compreensiva contra a personificação em múltiplas camadas de uma rede. Estas investigações são particularmente pertinentes para este estudo, pois destacam a necessidade de uma abordagem holística que não só combate, mas também explica e adapta-se às estratégias de ataques em evolução.

Portanto, a integração de XAI em sistemas de detecção de intrusões representa um passo significativo para a melhoria da segurança em redes Wi-Fi (802.11), garantindo não só a detecção eficaz, mas também a compreensão completa dos processos que levam à identificação de ameaças, essencial para o desenvolvimento de estratégias de defesa mais eficientes e adaptáveis.

1.3 Objetivos

Nesta seção, são estabelecidos os objetivos que orientam o escopo desta pesquisa. O objetivo geral descreve a ambição central do estudo, enquanto os objetivos específicos fragmentam essa meta maior em tarefas focadas, detalhando as abordagens metodológicas e as expectativas de contribuições para o campo de estudo.

1.3.1 Objetivo Geral

O objetivo geral deste trabalho é desenvolver e avaliar um modelo de detecção de intrusões utilizando o algoritmo XGBoost, integrado com a técnica de interpretabilidade SHAP, para melhorar a precisão e a explicabilidade na detecção de ataques de personificação em redes Wi-Fi (802.11).

1.3.2 Objetivos Específicos

Para alcançar o objetivo geral mencionado, este trabalho de conclusão de curso foca nos seguintes objetivos específicos:

- ❏ Reproduzir experimentos da literatura que abordam a detecção de ataques de personificação em redes Wi-Fi (802.11), utilizando um conjunto de dados que inclui tipos específicos de ataques como *evil twin*, *cafe latte* e *hirte*. Este passo visa validar as metodologias existentes e estabelecer uma base comparativa para a avaliação do modelo proposto.

- ❏ Implementar o uso do classificador XGBoost para avaliar sua performance em diferentes conjuntos de características extraídos da literatura. A escolha do XGBoost se deve à sua capacidade comprovada de fornecer resultados robustos e eficientes em diversos problemas de classificação.

- ❏ Aplicar a biblioteca SHAP para interpretar a influência das características na decisão do modelo XGBoost. O uso desta técnica de XAI visa desvendar como cada característica contribui para a classificação do modelo, proporcionando *insights* valiosos para a compreensão e aperfeiçoamento do sistema de detecção de intrusões.

1.4 Hipóteses

Considerando a complexidade e diversidade dos ataques de personificação em redes Wi-Fi (802.11) e a necessidade de sistemas de detecção que sejam transparentes e compreensíveis, propõem-se as seguintes hipóteses para nortear a investigação e validação dos métodos sugeridos neste estudo:

1.4.1 Hipótese Principal

A integração de técnicas de XAI com modelos de aprendizado de máquina, como o XGBoost, eleva de maneira significativa a precisão e a explicabilidade na detecção de ataques de personificação em redes Wi-Fi (802.11).

Essa hipótese baseia-se na premissa de que a aplicação de técnicas explicativas pode elucidar como características específicas impactam nas previsões do modelo, proporcionando um entendimento mais aprofundado que facilita a otimização e a confiabilidade no sistema de detecção de intrusões.

1.4.2 Hipóteses Secundárias

Características identificadas por técnicas de XAI demonstram maior eficácia na detecção de ataques de personificação do que aquelas selecionadas por métodos tradicionais de aprendizado de máquina.

Modelos de detecção que incorporam explicabilidade (XAI) tendem a ser mais valorizados por usuários finais e especialistas em segurança devido à maior transparência e confiança nas decisões do modelo.

A eficiência do modelo de detecção de intrusões varia significativamente entre diferentes tipos de ataques de personificação, como *evil twin*, *cafe latte*, e *hirte*, em função das características únicas de cada ataque.

A validação destas hipóteses não apenas confirmará a eficácia das técnicas utilizadas, mas também contribuirá para o aperfeiçoamento contínuo das práticas de segurança em redes Wi-Fi (802.11).

1.4.3 Justificativa das Hipóteses

A formulação destas hipóteses é justificada pela observação de lacunas na literatura existente, especialmente quanto à explicabilidade dos modelos de detecção de ataques de personificação. Estudos anteriores, embora tenham alcançado eficácia na detecção, geralmente não esclarecem a transparência nas decisões dos modelos de IA, o que pode limitar sua utilidade prática e aceitação pelos profissionais de segurança ((AMINANTO; KIM, 2016), (AMINANTO; KIM, 2017)). Assim, este estudo pretende explorar como a explicabilidade afeta tanto a performance técnica quanto a percepção dos sistemas de detecção, fornecendo *insights* importantes para o desenvolvimento de futuros sistemas de segurança cibernética.

1.5 Disposição da Monografia

Este documento está organizado da seguinte forma: O Capítulo 2 aborda uma revisão da literatura, discutindo trabalhos relacionados ao tema de detecção de ataques de personificação em redes Wi-Fi (802.11) e o uso de técnicas de Inteligência Artificial, com ênfase no XAI. O Capítulo 3 detalha a metodologia empregada neste estudo, incluindo a descrição do conjunto de dados, o processo de seleção de características, o treinamento e a avaliação do modelo utilizando XGBoost e SHAP. O Capítulo 5 apresenta uma analise do resultados obtidos, verificando o desempenho do modelo em termos de acurácia, precisão, *recall* e *F1-score*, além de explorar a contribuição das diferentes características na predição do modelo por meio de análises SHAP. Por fim, o Capítulo 6 conclui o trabalho, resumindo os principais achados, discutindo as implicações práticas e sugerindo direções para pesquisas futuras no campo da detecção de intrusões e da Inteligência Artificial Explicável.

Capítulo 2

Fundamentação Teórica

Antes de explorar as especificidades dos ataques de personificação e seu impacto nas redes Wi-Fi (802.11), é fundamental estabelecer uma compreensão sólida da natureza e do escopo desses ataques. Esta seção visa detalhar os diferentes tipos de ataques de personificação, elucidando suas metodologias, objetivos e implicações para a segurança das redes. A análise desses ataques proporcionará a base necessária para entender as estratégias de detecção e defesa aplicáveis, estabelecendo o contexto para a discussão subsequente sobre sistemas de detecção de intrusões e a aplicação de técnicas avançadas de Inteligência Artificial para o aprimoramento da segurança em redes Wi-Fi (802.11).

2.1 Ataques de Personificação

Os ataques de personificação nas redes Wi-Fi (802.11) ocorrem quando um agente mal-intencionado finge ser outro dispositivo ou usuário legítimo para ganhar acesso não autorizado a informações ou redes. Isso pode incluir ataques do tipo *man-in-the-middle*, no qual o atacante intercepta e potencialmente altera a comunicação entre dois dispositivos, ou ataques de falsificação de IP, no qual o atacante usa endereços IP de outros dispositivos como se fossem próprios.

A Figura 1 exemplifica um ataque de personificação, no qual o nó atacante (N3) finge ser um nó legítimo (N1) para enganar outros nós restantes da rede. O atacante estabelece um caminho de comunicação com N5 e, simultaneamente, intercepta os dados destinados a esse nó. Este tipo de ataque explora a confiança entre os nós da rede e a falta de autenticação adequada no estabelecimento de conexões. O objetivo do atacante é manipular os nós N1 e N2 para que acreditem que estão comunicando diretamente com N5, enquanto, na verdade, toda a comunicação está sendo redirecionada através de N3, permitindo ao atacante acesso aos dados ou a capacidade de inserir dados maliciosos.

Os ataques de personificação emergem como ameaças prevalentes nas redes Wi-Fi (802.11), onde atacantes simulam entidades legítimas, como usuários ou estações base, buscando acesso indevido a redes ou sistemas de computadores (AMINANTO et al.,

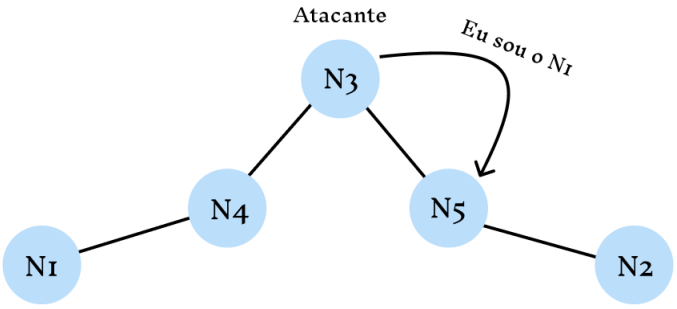

Figura 1 – Demonstração de um ataque de personificação

2018; QUINCOZES; KAZIENKO; COPETTI, 2018). Tais ataques exploram várias técnicas, incluindo clonagem de dispositivos, estabelecimento de pontos de acesso falsos, *spoofing* de endereços IP e ataques de *replay* (AMINANTO et al., 2018; BARBEAU; HALL; KRANAKIS, 2006; QUINCOZES; KAZIENKO; COPETTI, 2018).

Classificação dos Ataques

Baseados nos propósitos dos atacantes, esses ataques podem ser classificados em diversas categorias, como a interceptação de comunicações ou a obtenção de chaves criptográficas usadas na comunicação entre clientes conectados (AMINANTO; KIM, 2017; KOLIAS et al., 2016; QUINCOZES; KAZIENKO; COPETTI, 2018).

2.1.1 Ataque Evil Twin

O ataque *Evil Twin* é uma técnica sofisticada de personificação, onde um Acess Point - (AP) falso é criado para se parecer exatamente com um AP legítimo em termos de Service Set Identifier - (SSID) e segurança, com o objetivo de enganar as vítimas para que se conectem a ele em vez do AP genuíno. Essa tática explora a confiança que os usuários depositam em suas redes conhecidas e a automação de conexões a redes conhecidas em dispositivos. Uma vez conectadas ao AP malicioso, todas as transmissões de dados das vítimas podem ser interceptadas pelo atacante.

Este tipo de ataque é particularmente eficaz em locais públicos com alta densidade de redes Wi-Fi (802.11), como aeroportos, cafés e hotéis, onde os usuários frequentemente buscam conexões de internet gratuitas. Os atacantes podem facilmente configurar um AP com um nome de rede semelhante ou idêntico aos oferecidos na área, aumentando a probabilidade de conexões inadvertidas ao AP falso.

A capacidade do AP malicioso de fornecer um sinal mais forte é crucial, pois os dispositivos geralmente se conectam ao AP com o sinal mais forte disponível. Isso não só aumenta a chance de capturar mais vítimas, como também melhora a estabilidade e a qualidade da conexão, facilitando o monitoramento prolongado e a coleta de dados.

Durante um ataque *Evil Twin*, o atacante pode realizar uma variedade de atividades maliciosas, incluindo, mas não se limitando a, *sniffing* de tráfego, roubo de identidade, disseminação de *malware* e ataques de *phishing*. Isso é facilitado pelo posicionamento do atacante como um intermediário (*Man-In-The-Middle*), onde ele pode interceptar, alterar e retransmitir comunicações entre a vítima e seus destinos pretendidos sem levantar suspeitas.

No cenário 1, considera-se que o atacante encontra-se dentro do alcance da vítima, mas não necessariamente dentro da área de cobertura de um ponto de acesso AP legítimo. Já no cenário 2, o atacante está situado dentro das áreas de cobertura tanto da vítima quanto do AP legítimo ao qual a vítima pretende se conectar. A Figura 2 demonstra esses cenários.

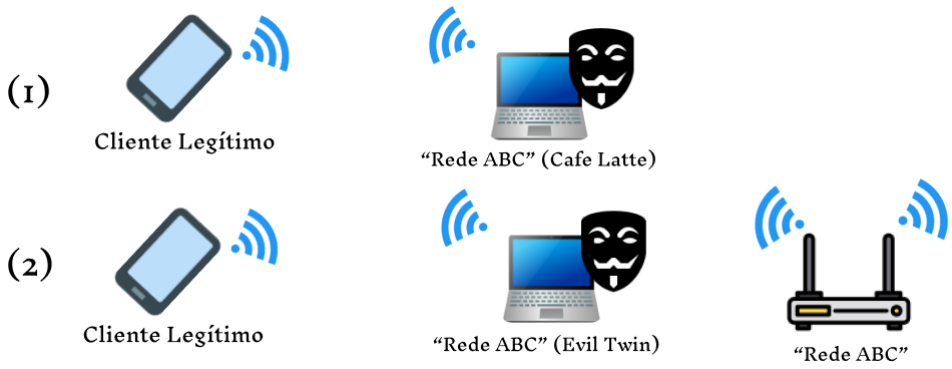

Figura 2 – Cenários de ataque de personificação explorados na avaliação.

No cenário 1, uma vez que a vítima se conecta a esta rede, terá os dados necessários coletados para que outra ferramente decifre a chave WEP A consequência deste ataque é a violação da confidencialidade das comunicações da vítima, permitindo que o atacante espione todas as informações transmitidas.

No cenário 2, o atacante, utilizando o mesmo conjunto de ferramentas do cenário 1, também está dentro do alcance do AP legítimo. Neste caso, ele cria uma rede falsa que emite um sinal mais forte do que o AP legítimo, persuadindo a vítima a conectar-se a ela. Esta rede maliciosa possui o mesmo SSID do AP legítimo, facilitando o engano. Uma vez conectada à rede do atacante, todas as informações transmitidas pela vítima, incluindo dados sensíveis como senhas e informações bancárias, são interceptadas antes de serem retransmitidas ao AP legítimo, resultando em uma exposição significativa de dados confidenciais sem o conhecimento da vítima.

Para a detecção e mitigação de ataques *Evil Twin*, é recomendado o uso de ferramentas e técnicas de segurança avançadas, como a implementação de autenticação de dois fatores para redes, o uso de Virtual Private Network - (VPN)s para criptografar dados de comunicação, e a verificação cuidadosa da autenticidade das redes Wi-Fi (802.11) antes da conexão (ZOVI; MACAULAY, 2005; QUINCOZES; KAZIENKO; COPETTI, 2018).

2.1.2 Ataque Cafe Latte

O ataque *Cafe Latte* é uma técnica de personificação que se concentra em explorar clientes individuais conectados a redes Wi-Fi (802.11) seguras pelo protocolo WEP, sem a necessidade de o atacante estar fisicamente próximo aos pontos de acesso. Esse tipo de ataque explora uma vulnerabilidade no protocolo WEP para interceptar e decifrar chaves de criptografia usando requisições Address Resolution Protocol - (ARP) forjadas.

Diferentemente de outros métodos que requerem uma presença física próxima à rede alvo, o ataque *Cafe Latte* pode ser executado a partir de uma distância significativa, aproveitando a função dos dispositivos de se reconectar automaticamente a redes conhecidas. Os atacantes utilizam essa característica para enviar pacotes ARP maliciosos que são projetados para forçar o dispositivo a revelar informações que podem ser usadas para deduzir a chave WEP.

A eficácia do ataque *Cafe Latte* reside na sua capacidade de compilar rapidamente os dados necessários para quebrar a chave WEP, frequentemente em questão de minutos. O método envolve o envio intensivo de pacotes ARP para o dispositivo alvo, cada um contendo diferentes variações de dados criptografados, permitindo que o atacante execute uma análise estatística para decifrar a chave WEP.

Este tipo de ataque destaca as deficiências significativas do protocolo WEP e sublinha a necessidade de protocolos mais seguros como Wi-Fi Protected Access 2 (WPA2) ou Wi-Fi Protected Access 3 (WPA3) em ambientes modernos. As implicações para a segurança das redes Wi-Fi (802.11) são consideráveis, sugerindo a importância de medidas preventivas como a atualização para protocolos de segurança mais robustos e a vigilância constante de atividades suspeitas que podem indicar a presença de ataques como o *Cafe Latte* (AMINANTO et al., 2018; AHMAD; RAMACHANDRAN, 2007; QUINCOZES; KAZIENKO; COPETTI, 2018).

2.1.3 Ataque Hirte

O ataque *Hirte* amplia a metodologia do ataque *Cafe Latte*, empregando técnicas de fragmentação para quebrar chaves criptográficas em comunicações WEP. Essa abordagem se baseia na divisão da requisição em múltiplos fragmentos e na manipulação do comprimento do primeiro fragmento para alterar o Internet Protocol address - (IP) de origem durante o processo de remontagem pelo cliente. Esta manipulação facilita significativamente a obtenção da chave WEP, tornando o ataque particularmente eficaz.

A técnica envolve o envio de pacotes que são intencionalmente fragmentados de maneira que, quando reagrupados pelo algoritmo de *reassemble* do dispositivo alvo, resultam em uma mensagem que parece ser de uma fonte confiável. Esta estratégia explora vulnerabilidades no protocolo WEP que não foi projetado para lidar de forma segura com pacotes fragmentados que podem ser manipulados durante a transmissão.

O ataque *Hirte* é uma demonstração da necessidade de protocolos mais robustos e seguros, como WPA2 ou WPA3, que oferecem mecanismos mais eficazes para proteger contra alterações no tráfego de rede.

2.2 Sistema de Detecção de Intrusões

Como ilustrado na Figura 3, um Sistema de Detecção de Intrusão em Rede (IDS) é uma parte integrante da arquitetura de segurança de uma rede corporativa. O IDS é colocado de forma estratégica dentro da rede para monitorar o tráfego que passa através dos dispositivos de rede, como *switches* e roteadores, bem como o tráfego que entra e sai da rede através do *firewall*. Ele atua como uma sentinela, analisando os pacotes de dados que circulam no ambiente corporativo para identificar padrões suspeitos ou anomalias que possam indicar uma tentativa de intrusão ou atividade maliciosa. Nessa configuração, o IDS tem uma visão ampla do tráfego de rede, permitindo-lhe observar comunicações entre sistemas internos, como servidores de *e-mail* e *web*, e também o tráfego entre esses sistemas e o mundo externo, potencialmente interceptando ataques antes que eles atinjam os ativos críticos da rede.

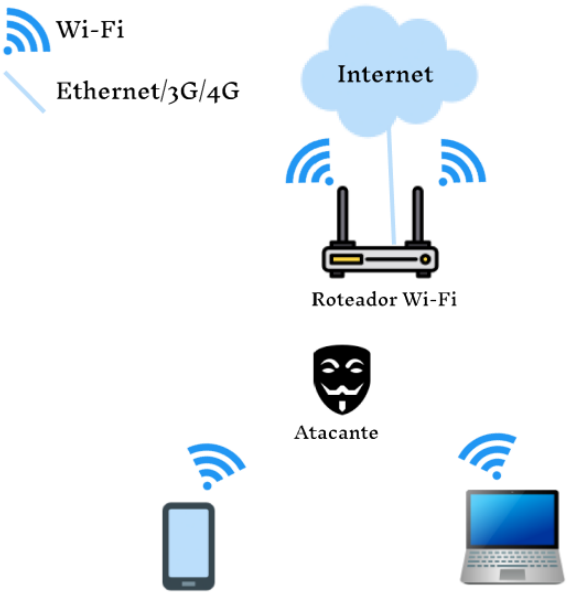

Figura 3 – Modelo de funcionamento de um IDS

Algoritmos de aprendizado de máquina desempenham um papel vital na automação do processo de seleção de características e na melhoria dos IDS. Eles podem aprender a partir de dados históricos para identificar padrões e comportamentos, sendo capazes de classificar e prever atividades maliciosas. Técnicas como árvores de decisão, redes neurais, e máquinas de vetor de suporte são comumente empregadas para analisar e decidir quais características são mais indicativas de intrusões ou ataques.

A Figura 4 ilustra o processo típico de aprendizado de máquina, começando com a entrada de dados, neste caso, uma imagem que requer classificação. A fase de extração de características é onde o algoritmo identifica as informações relevantes da entrada, essenciais para o processo de tomada de decisão. Segue-se a classificação, um passo crítico onde o modelo de aprendizado de máquina, treinado em dados previamente rotulados, utiliza as características extraídas para determinar a categoria adequada para a entrada. O resultado é uma decisão de classificação que forma a saída do sistema, evidenciando a capacidade dos algoritmos de aprendizado de máquina para automatizar e aprimorar a detecção de intrusões.

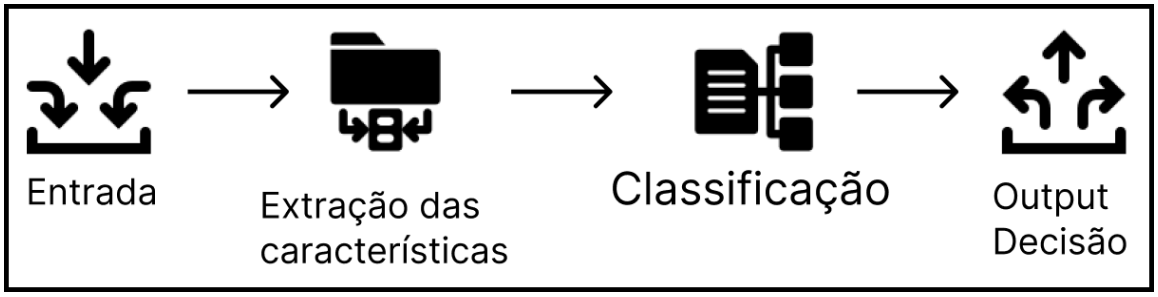

Figura 4 – Diagrama esquemático do processo de aprendizado de máquina.

2.2.1 Classificação de Dados na Detecção de Intrusões

A classificação de dados é uma técnica crucial no campo da ciência de dados e aprendizado de máquina, usada para categorizar diferentes tipos de dados em grupos específicos. Essa metodologia é particularmente útil na detecção de intrusões, uma vez que permite a identificação automática de atividades anormais ou maliciosas em uma rede. Ao analisar padrões de tráfego e comportamento dos usuários, sistemas de detecção de intrusões empregam algoritmos de classificação para diferenciar entre atividades legítimas e potenciais ameaças, incluindo ataques de personificação. Estes últimos, devido à sua natureza enganosa, representam um desafio único, pois imitam a identidade de dispositivos legítimos para obter acesso não autorizado, tornando sua detecção uma prioridade neste trabalho de conclusão de curso (QUINCOZES; KAZIENKO; COPETTI, 2018).

Dentro do espectro de ferramentas e algoritmos disponíveis para a classificação, o XGBoost (Extreme Gradient Boosting) destaca-se pela sua alta eficiência e eficácia na modelagem de problemas complexos, incluindo a detecção de intrusões. O XGBoost é um exemplo primário de um classificador baseado em árvores de decisão, que faz parte de uma categoria maior conhecida como métodos de *ensemble*. Os métodos de *ensemble* visam melhorar o desempenho dos modelos de aprendizado de máquina ao combinar as previsões de múltiplos modelos de treinamento, proporcionando uma solução mais robusta e precisa (GOUVEIA; CORREIA, 2020).

O XGBoost, em particular, opera construindo sequencialmente árvores de decisão, onde cada nova árvore corrige os erros cometidos pelas árvores anteriores em um processo conhecido como *boosting*. Este método é caracterizado pelo seu uso de gradientes no processo de otimização, permitindo uma ajuste fino e eficiente dos parâmetros do modelo. Através desta técnica, o XGBoost é capaz de lidar com uma ampla gama de dados e tipos de problemas de classificação, tornando-o uma ferramenta poderosa para a detecção de intrusões e ataques de personificação (CHEN; GUESTRIN, 2016).

2.2.2 Seleção de Características

A seleção de características é um processo fundamental em aprendizado de máquina e análise de dados, especialmente em IDS, no qual identificar as características mais relevantes pode significar a diferença entre detectar um ataque e deixá-lo passar despercebido. O objetivo é escolher as características dos dados que contribuem mais para a precisão do modelo e identificar quais dessas características são mais influentes e pontuais na análise de um possível ataque, melhorando a eficiência e a eficácia do sistema ao reduzir a dimensionalidade e eliminar o ruído dos dados.

Existem diversas ferramentas para a seleção de características, no entanto, neste trabalho nenhuma delas será aplicada. Essa decisão se justifica pelo propósito do trabalho, o qual visa estudar conjuntos já propostos por outros trabalhos na literatura à luz da explicabilidade fornecida pelas ferramentas a serem exploradas neste trabalho de conclusão de curso.

2.3 Inteligência Artificial Explicável – XAI

Explicabilidade em inteligência artificial refere-se à capacidade de entender e interpretar os resultados gerados por modelos de IA. Ela se torna especialmente importante em aplicações onde as decisões tomadas por algoritmos podem ter impactos significativos, como na área da saúde, finanças e segurança — especialmente na cibersegurança. A explicabilidade é essencial para construir confiança nos sistemas de IA, permitindo que os usuários compreendam como as decisões são feitas e potencialmente contestem ou verifiquem essas decisões.

A aplicação da XAI para detecção de intrusões mostrou eficácia, conforme será apontado por esse estudo, que utiliza-se o XGBoost, um algoritmo de aprendizado de máquina que se destaca por sua precisão e eficiência. Métodos SHAP são utilizados para interpretar e justificar as saídas de classificadores como o XGBoost, facilitando a compreensão dos modelos e aumentando a confiança dos usuários em suas previsões (BROECK et al., 2022).

A biblioteca SHAP, especificamente, tem sido uma ferramenta valiosa para XAI, permitindo a análise da importância dos recursos individuais na previsão de modelos de aprendizado de máquina.

2.4 Trabalhos Relacionados

Esta seção revisa trabalhos relevantes que exploram diversas estratégias de detecção e prevenção de ataques de personificação, além de discutir a aplicação de técnicas de IA e ML para aprimorar a segurança em redes sem fio. Em particular, aborda-se estudos que se concentram tanto na eficácia das técnicas de detecção quanto na explicabilidade dos modelos utilizados, identificando lacunas e oportunidades para futuras investigações.

2.4.1 Detecção de Ataques de Personificação

Vários estudos na literatura abordam a problemática da detecção de ataques de personificação em redes Wi-Fi. Por exemplo, o artigo de Aminanto e Kim (2016) explora o uso de *deep learning* para identificar ataques de personificação, destacando a eficácia dessa abordagem em melhorar a precisão da detecção. Embora esse trabalho demonstre resultados promissores, ele não enfoca a explicabilidade do modelo, que é um aspecto central no nosso estudo.

Similarmente, Aminanto e Kim (2017) investigam técnicas de seleção de características para otimizar a detecção de intrusões. Este trabalho é particularmente relevante para o estudo, pois também concentra-se na seleção de características para melhorar a performance do modelo, embora o trabalho vá além ao integrar a explicabilidade na análise.

Além desses trabalhos, recentes pesquisas exploram a segurança em diferentes contextos tecnológicos, ilustrando a diversidade e a complexidade dos desafios de segurança atuais. Zhu e Cao () realizam uma análise crítica de um esquema de autenticação e acordo de chave para a Internet dos Veículos, revelando vulnerabilidades contra ataques de personificação, o que destaca a importância contínua da inovação em métodos de detecção e prevenção de ataques.

2.4.2 Exploração de Técnicas de Segurança em Ambientes de Computação em Névoa

No artigo de Tu et al. (2018), os autores abordam o desafio significativo de garantir a segurança em ambientes de computação em névoa, que, devido à sua arquitetura descentralizada e distribuída, apresentam vulnerabilidades únicas a ataques de personificação. Neste contexto, os ataques de personificação são particularmente problemáticos porque podem levar à usurpação de identidades de dispositivos conectados, resultando em acessos não autorizados a dados críticos e serviços essenciais.

Para enfrentar esses desafios, Tu et al. (2018). propõem um método sofisticado que integra técnicas de criptografia avançada e autenticação robusta. A abordagem sugere a utilização de certificados de segurança atualizados dinamicamente e algoritmos de criptografia baseados em chave pública para validar a identidade dos dispositivos na névoa, um

método que se mostra eficaz não só na detecção mas também na prevenção de tentativas de personificação.

Este método é especialmente relevante para redes Wi-Fi, onde a natureza igualmente distribuída dos dispositivos e a variabilidade dos pontos de acesso podem criar lacunas de segurança semelhantes. A implementação de uma estratégia de autenticação contínua e adaptativa, como proposto por Tu et al. (2018), pode fornecer uma camada adicional de segurança em redes Wi-Fi, mitigando o risco de ataques de personificação e garantindo que apenas dispositivos e usuários autenticados possam acessar a rede.

A pesquisa de Tu et al. (2018) é fundamental para o entendimento de como as técnicas de segurança podem ser adaptadas de ambientes de computação em névoa para redes Wi-Fi, sugerindo que as soluções desenvolvidas para um ambiente podem ser efetivamente transpostas para o outro, considerando as semelhanças em suas vulnerabilidades e desafios de segurança.

2.4.3 Prevenção de Ataques em Redes *Ad Hoc* Móveis

O artigo de Tamilselvan e Sankaranarayanan (2007) aborda a problemática de segurança em redes móveis *ad hoc*, um ambiente de rede que possui características similares às redes Wi-Fi em termos de sua estrutura descentralizada e dependência da comunicação direta entre dispositivos. As redes *ad hoc* são particularmente suscetíveis a ataques de personificação, onde atacantes podem facilmente se passar por outros nós devido à falta de uma infraestrutura centralizada de autenticação.

Neste estudo, Tamilselvan e Sankaranarayanan (2007) propõem métodos robustos de autenticação que são essenciais para mitigar riscos de personificação. Eles desenvolvem e testam algoritmos que podem ser implementados diretamente nos dispositivos, permitindo que eles autentiquem de forma independente a identidade de outros nós na rede. Esta abordagem é baseada na utilização de chaves criptográficas compartilhadas e protocolos de autenticação que garantem que os nós sejam quem dizem ser antes de qualquer troca de dados.

A relevância deste trabalho para redes Wi-Fi reside na aplicabilidade desses métodos de autenticação em um ambiente similarmente distribuído e descentralizado. As técnicas desenvolvidas para redes *ad hoc* podem ser adaptadas para proteger redes Wi-Fi, onde a autenticação e a segurança são continuamente desafiadas pela presença de múltiplos acessos e a necessidade de uma gestão de segurança eficaz em um ambiente dinâmico e frequentemente não seguro.

Este estudo também discute a importância de protocolos de segurança adaptativos que possam responder dinamicamente a ameaças emergentes, algo igualmente crítico para a segurança em redes Wi-Fi. A implementação de tais protocolos pode melhorar significativamente a resiliência de redes Wi-Fi contra ataques de personificação, fornecendo uma

camada de segurança que é tanto proativa quanto reativa em face de ameaças variadas e sofisticadas.

2.4.4 Personificação em Redes Sociais Online

O trabalho de Goga, Venkatadri e Gummadi (2015) aborda uma faceta intrigante da personificação, focando especificamente nos ataques dentro de redes sociais. Este estudo é crítico por iluminar como identidades falsas são criadas e utilizadas para enganar usuários e sistemas de segurança em plataformas altamente interativas e sociais. A investigação detalha métodos pelos quais atacantes executam esses ataques e a eficácia das técnicas de aprendizado de máquina na identificação e prevenção de tais ameaças.

A pesquisa demonstra o uso de algoritmos avançados para analisar padrões de comportamento que diferenciam usuários legítimos de contas falsas ou duplicadas, que podem ser aplicados transversalmente para aumentar a segurança em outros domínios digitais, inclusive em redes Wi-Fi. Em ambientes de rede Wi-Fi, a personificação pode ocorrer quando atacantes tentam acessar a rede assumindo a identidade de um usuário legítimo, através de técnicas como *spoofing* de MAC ou IP. Os *insights* do estudo são especialmente relevantes para desenvolver sistemas de detecção mais robustos que utilizam aprendizado de máquina para monitorar e analisar padrões de acesso e comportamento de rede, procurando anomalias que possam indicar tentativas de intrusão.

Além disso, o estudo destaca a importância da implementação de medidas proativas de segurança que possam adaptar-se dinamicamente ao comportamento do usuário e a padrões de tráfego inusuais, sugerindo uma aplicabilidade direta dessas técnicas em redes Wi-Fi onde a personificação pode levar a acessos não autorizados ou a ataques mais sérios.

Este trabalho de Goga, Venkatadri e Gummadi (2015) oferece uma visão profunda sobre a dinâmica de ataques de personificação e propõe soluções que, embora inicialmente focadas em redes sociais online, fornecem um *framework* valioso para fortalecer a segurança em redes Wi-Fi e outros ambientes digitais vulneráveis a identidades falsificadas e acessos fraudulentos.

2.4.5 Uso de Aprendizado por Reforço para Detectar Ataques

O estudo de Tu et al. (2021) representa um avanço significativo na aplicação de tecnologias emergentes para a segurança de redes, especificamente através do uso de aprendizado por reforço para detectar ataques de personificação em comunicações dispositivo a dispositivo. Esta pesquisa explora a capacidade do aprendizado por reforço em adaptar-se e responder dinamicamente a ameaças emergentes, o que é crucial em ambientes de rede dinâmicos e em constante mudança, como as redes Wi-Fi.

A metodologia proposta utiliza agentes de aprendizado por reforço que são treinados para identificar padrões de comportamento anômalo que tipificam ataques de personifica-

ção. Esses agentes aprendem com a interação contínua com o ambiente de rede, ajustando suas estratégias de detecção com base no sucesso ou falha de suas ações anteriores. Isso permite que o sistema de segurança evolua com o tempo, melhorando sua capacidade de detectar novos tipos de ataques que não foram previamente identificados.

A aplicabilidade desta técnica em redes Wi-Fi é particularmente pertinente, dado o alto volume e a diversidade de dispositivos que acessam essas redes. Redes Wi-Fi, frequentemente expostas a um espectro amplo de ameaças de segurança, podem beneficiar significativamente de sistemas de segurança que não apenas reagem a ameaças conhecidas, mas também aprendem e se adaptam a novas táticas de ataque.

Este estudo não apenas destaca a eficácia do aprendizado por reforço na detecção de ataques de personificação, mas também abre caminho para futuras pesquisas e desenvolvimento de sistemas de segurança que são verdadeiramente adaptativos e proativos. A capacidade de adaptar-se e aprender com o ambiente é um passo crucial na criação de defesas mais robustas contra os ataques cibernéticos sofisticados e em constante evolução que caracterizam o cenário de ameaças moderno.

2.4.6 Ataques de Personificação em Bluetooth

O estudo realizado por Antonioli, Tippenhauer e Rasmussen (2020) introduz uma análise profunda sobre os ataques de personificação em tecnologias Bluetooth, uma dimensão crucial para compreender as vulnerabilidades de segurança em redes Wi-Fi, especialmente considerando que muitos dispositivos Wi-Fi são equipados com funcionalidades Bluetooth. Este trabalho é vital para mapear como ataques de personificação podem ser executados através de diferentes camadas de conectividade e quais medidas de segurança podem ser efetivamente implementadas para mitigar tais ameaças.

O estudo detalha uma série de ataques de personificação, conhecidos como Bluetooth Impersonation AttacksS (BIAS), que exploram falhas nas implementações do protocolo *Bluetooth* para realizar ataques de *man-in-the-middle* (MITM) ou de falsificação de identidade. Estes ataques são particularmente significativos pois demonstram que as vulnerabilidades não estão limitadas a dispositivos isoladamente, mas são extensíveis a toda a rede na qual esses dispositivos operam.

A pesquisa de Antonioli, Tippenhauer e Rasmussen (2020) é relevante para a segurança em redes Wi-Fi na medida em que oferece *insights* sobre como dispositivos conectados via *Bluetooth* podem comprometer redes mais amplas, incluindo infraestruturas de Wi-Fi. Além disso, o estudo propõe métodos para detectar e prevenir esses ataques, incluindo melhorias nos procedimentos de autenticação e validação, que podem ser adaptados para reforçar a segurança em redes Wi-Fi.

Esta abordagem não só aumenta a compreensão das técnicas de ataque e de defesa em camadas de comunicação múltiplas mas também incentiva a integração de protocolos de

segurança mais robustos em dispositivos que suportam múltiplas interfaces de comunicação, fortalecendo a resiliência de redes contra ataques de personificação multifacetados.

2.4.7 Necessidade de um trabalho para a explicabilidade em redes Wi-Fi

A integração de técnicas de explicabilidade em sistemas de detecção de intrusões em redes Wi-Fi representa uma área significativamente sub explorada na literatura de segurança cibernética. Embora estudos recentes tenham começado a abordar a importância da transparência e da compreensão das decisões tomadas por sistemas automatizados, a aplicação específica de tais técnicas em contextos de detecção de ataques de personificação em redes Wi-Fi ainda é escassa.

Pesquisas anteriores, como as realizadas por Aminanto e Kim (2016) e Aminanto e Kim (2017), demonstram avanços na detecção de intrusões usando aprendizado de máquina, mas frequentemente deixam de lado a questão crítica da explicabilidade. Esses estudos focam predominantemente na eficiência e na eficácia da detecção, negligenciando como os modelos de aprendizado de máquina alcançam seus resultados. A falta de transparência nos modelos de aprendizado profundo, muitas vezes criticada em diversos domínios aplicados, é particularmente problemática em segurança cibernética, onde entender o *porquê* de uma detecção é essencial para a confiança nas decisões do sistema e para a adequada resposta a incidentes.

Este vácuo de pesquisa é evidenciado pela escassez de estudos que combinem técnicas de aprendizado de máquina com métodos explicativos como a análise SHAP para elucidar a contribuição das variáveis na detecção de comportamentos maliciosos. Apesar da disponibilidade de ferramentas de XAI que poderiam preencher essa lacuna, poucos pesquisadores aplicaram essas ferramentas para melhorar a compreensão dos modelos de detecção de ataques em redes Wi-Fi.

A necessidade de explicabilidade é ainda mais crítica quando considera-se os tipos específicos de ataques estudados, como *evil twin*, *cafe latte*, e *hirte*. Esses ataques apresentam desafios únicos devido às suas sofisticadas técnicas de engano e a capacidade de se misturarem ao tráfego legítimo, o que exige um nível mais profundo de análise para a identificação correta e resposta apropriada.

- A explicabilidade nas detecções destes ataques poderia facilitar a identificação de falsos positivos e aprimorar os protocolos de resposta a incidentes, permitindo que os administradores de rede entendam claramente as razões por trás das alertas de segurança.

- Além disso, uma melhor compreensão das características que indicam a presença de ataques poderia levar ao desenvolvimento de novas estratégias de mitigação,

personalizadas para as táticas específicas usadas pelos invasores em ambientes de rede Wi-Fi.

Portanto, há uma necessidade latente e presente de pesquisa adicional nesta área para desenvolver sistemas de detecção de intrusões que não apenas identifiquem ataques de forma eficiente, mas também forneçam explicações compreensíveis e acionáveis que possam ser utilizadas para fortalecer a segurança de redes Wi-Fi contra formas cada vez mais sofisticadas de personificação.

Capítulo 3

Materiais e Métodos

Neste capítulo, é abordada a descrição e análise dos conjuntos de dados e das características utilizadas para alimentar o modelo de detecção de intrusões. A seleção criteriosa de conjuntos de dados e a identificação das características mais relevantes são etapas fundamentais para assegurar a precisão e eficácia do modelo na identificação de ataques de personificação. A metodologia adotada para a escolha dos conjuntos de dados é explorada, detalhando-se as características selecionadas e justificando sua importância no contexto da detecção de intrusões em redes Wi-Fi (802.11). Este processo metodológico é vital para compreender como os dados são preparados e analisados, estabelecendo a base para as fases subsequentes de treinamento e avaliação do modelo proposto no estudo.

3.1 Discussão sobre as Metodologias

- O algoritmo XGBoost foi escolhido devido à sua capacidade de lidar eficientemente com diversas características e sua robustez em fornecer resultados consistentes, além de ser complementado pela SHAP para uma interpretação precisa e fundamentada das decisões do modelo.

- A escolha do conjunto C2, que mostrou os melhores resultados, baseia-se na sua composição de características que são cruciais para identificar padrões específicos de tráfego de rede associados a ataques de personificação. Estudos similares destacaram conjuntos de características eficazes, mas poucos proporcionaram a clareza e a profundidade de análise possibilitadas pelo uso de SHAP.

3.2 Conjuntos de Dados e Características

Nesta seção, será realizada uma análise detalhada do conjunto de dados utilizado, essencial para a implementação e validação do modelo de detecção de intrusões. O foco será no conjunto de dados AWID2, uma referência na área de segurança de redes Wi-

Fi (802.11), que fornece uma base rica para a análise de ataques de personificação. A seleção e a descrição das características específicas extraídas deste conjunto de dados são fundamentais para entender como o modelo processará as informações e identificará potenciais ameaças, estabelecendo uma ligação direta entre os dados brutos e as técnicas de aprendizado de máquina aplicadas.

3.2.1 Conjunto de Dados (Dataset)

Para esta seção, a metodologia empregada foi inspirada no artigo de Quincozes, Kazienko e Copetti (2018), onde se explora o uso do conjunto de dados AWID2, conforme reportado por Kolias et al. (2016), para a condução de experimentos em redes Wi-Fi (802.11) reais. O conjunto de dados AWID2 é notável por sua vasta coleção de registros, apresentando duas variações distintas: a primeira com aproximadamente 211 milhões de entradas e uma versão mais reduzida, com cerca de 2,3 milhões de registros. Estas duas variantes do conjunto de dados foram divididas em segmentos separados para facilitar o processo de treinamento e teste dos modelos de detecção de intrusões. Tanto a versão completa quanto a reduzida incluem dados referentes a 15 tipos de ataques, incluindo amostras de ataques que não serão utilizadas neste trabalho. Para o escopo deste trabalho, especificamente, as amostras de ataques de personificação foram: *Cafe-Latte*, *Evil Twin* e *Hirte*.

O foco deste trabalho recai sobre a análise de amostras relacionadas a ataques de personificação, resultando na seleção de 20.079 registros desses ataques específicos da versão reduzida do AWID2. Seguindo a recomendação de manter uma proporção equilibrada de 1:1 entre amostras normais e de ataques, conforme sugerido por Aminanto, Tanuwidjaja e Yoo (2017) e também adotado em (QUINCOZES; KAZIENKO; COPETTI, 2018), um número igual de amostras normais foi também selecionado. A distribuição destas amostras por classe de ataque, tanto para treinamento quanto para teste, é apresentada na Tabela 1.

Tabela 1 – Distribuição de amostras por classe (treinamento e teste) — baseado em (QUINCOZES; KAZIENKO; COPETTI, 2018).

Tipo	**Amostras de Treinamento**	**Amostras de Teste**
Evil Twin	2.633	611
Cafe Latte	45.889	379
Hirte	0	19.089
Normal	48.522	20.079

3.2.2 Características

Dentro o conjunto de dados AWID2 (descrito na seção 3.2.1), cada registro é composto por 155 características. A seguir são descritas as principais, selecionadas a partir de estudos comparativos de seu desempenho para a detecção de ataques de personificação (QUINCOZES; KAZIENKO; COPETTI, 2018).

7. **frame.len** - refere-se ao comprimento do quadro de dados que está sendo transmitido em uma rede. Valores maiores indicam quadros de dados maiores.

66. **wlan.fc.subtype** - indica o subtipo do campo de controle na Wireless Local Area Network (WLAN). Esse subtipo pode sinalizar qual a função do quadro de dados, como gerenciamento, controle ou dados.

74. **wlan.duration** - representa a duração do envio de um quadro na WLAN, o que pode influenciar o tempo que o canal de comunicação é ocupado e/ou utilizado.

47. **radiotap.datarate** - é a taxa de transmissão de dados do quadro, normalmente medida em Mbps. Uma taxa de dados mais alta pode indicar uma comunicação rápido ou um sinal de maior qualidade.

71. **wlan.fc.pwrmgt** - indicador do gerenciamento de energia no capo de controle de quadro na WLAN. Com ele é possível verificar se um dispositivo está em modo de economia de energia.

154. **data.len** - refere-se ao comprimento dos dados contidos no quadro, que pode ser diferente do comprimento total dos dados (frame.len).

145. **wlan.qos.tid** - esta característica é o identificador do tráfego em quadros *QoS(Quality of Service)* na WLAN. Isso ajuda na priorização do tráfego, fazendo com que os dados críticos sejam transmitidos de forma eficiente.

51. **radiotap.channel.type.ofdm** - indica se a modulação *OFDM(Orthogonal Frequency-Division Multiplexing)* é usada no canal de comunicação. O OFDM é uma técnica de modulação bastante utilizada para comunicação de alta velocidade em redes sem fio.

50. **radiotap.channel.type.cck** - refere-se ao uso da modulação *CCK(Complementary Code Keying)*. A CCK é uma técnica de modulação mais antiga, utilizada em padrões de rede sem fio como 802.11b.

3.2.3 Conjuntos de Características

De todas as 155 características disponíveis no conjunto de dados AWID2, apenas uma fração dessas características é considerada útil para a identificação eficaz de atividades

maliciosas. Com base nos conjuntos de características ressaltados na Tabela 3, procedeu-se a uma etapa de pré-seleção. Durante este processo, 83 características que não eram comuns aos conjuntos estudados foram excluídos da análise. Este procedimento foi desenvolvido com o intuito de refinar a seleção de dados para uma avaliação mais precisa e focada dos tipos de ataques de personificação, mantendo a integridade metodológica inspirada pelo trabalho seminal de Quincozes, Kazienko e Copetti (2018).

A revisão dos estudos anteriormente mencionados revela uma notável ausência de consolidação na seleção de características, mesmo ao focar em tipos específicos de ataques cibernéticos. A principal barreira identificada nestes estudos é a falta de um acordo comum sobre quais conjuntos de características são essenciais para análise. A Tabela 2 detalha as características escolhidos por diferentes pesquisadores.

Tabela 2 – Conjuntos de características avaliados na literatura. Em destaque, o melhor conjunto (C2), analisado neste trabalho.

Ref.	**Id.**	**Conjunto de Características**
(AMINANTO; TANUWIDJAJA; YOO, 2017)	C1	4, 8, 47, 68, 71
(AMINATO et al, 2017b)	**C2**	**8, 9, 47, 50, 51, 67, 71, 75, 145, 154**
(AMINANTO; TANUWIDJAJA; YOO, 2017)	C3	4, 7, 14, 31, 38, 64, 66, 67, 68, 70, 73, 75, 79, 82, 83, 90, 93, 94, 107, 112, 118, 120, 131, 134,136, 140
(AMINANTO; TANUWIDJAJA; YOO, 2017)	C4	1, 2, 3, 4, 8, 38, 61, 67, 70, 75, 76, 77, 78, 79, 80, 82, 107, 108, 109, 110, 111, 112, 119
(KALEEM D.; FERENS, 2017)	C5	4, 8, 77, 79, 82, 154
(AMINANTO; KIM, 2017)	C6	5, 38, 70, 71, 154
(AMINANTO; KIM, 2017)	C7	47, 50, 51, 67, 68, 71, 73, 82
(AMINANTO; KIM, 2017)	C8	4, 7, 38, 77, 82, 94, 107, 118
(AMINANTO; KIM, 2017)	C9	47, 64, 82, 94, 107, 108, 122, 154
(AMINANTO; KIM, 2017)	C10	11, 38, 61, 66, 68, 71, 76, 77, 107, 119, 140
(ALOTAIBI, 2016) (ALOTAIBI; ELLEITHY, 2016)	C11	8, 47, 50, 61, 66, 67, 68, 71, 73, 75, 76, 77, 78, 79, 80, 82, 110, 140, 141, 142
(KALEEM; FERENS,)	C12	4, 8, 77, 79, 82, 154
(WANG, 2016)	C13	4, 5, 6, 7, 8, 9, 38, 75, 81, 82
(AMINANTO et al., 2018)	C14	47, 64, 82, 94, 107, 108, 122, 154
(AMINANTO et al., 2016)	C15	4, 7, 8, 29, 38, 47, 62, 66, 67, 68, 70, 72, 73, 77, 79, 80, 82, 88, 93, 94, 98, 104, 107, 108, 112, 113, 122, 125, 126, 127, 140, 141, 142, 144, 148

Há uma variação significativa no número de características selecionados nos conjuntos. Alguns pesquisadores optam por conjuntos mais enxutos, com no mínimo 5 características, enquanto outros incluem até 35 características em suas análises.

Apesar das diferenças nos conjuntos de características selecionados, algumas características se destacam pela sua relevância, sendo frequentemente escolhidos em diversas pesquisas.

Os 15 conjuntos estudados, todos derivados do conjunto de dados AWID2, abarcam 188 características repetidas, dos quais 67 são únicos. Destes 154 características no total, 88 são descartados na análise consolidada. A Tabela 3 exibe as 15 características mais

Tabela 3 – Características mais selecionadas entre os conjuntos estudados. Baseado em (QUINCOZES; KAZIENKO; COPETTI, 2018).

Índice	**Nomenclatura**	**Seleções**	**%**
82	wlan.seq	11	68,75
107	wlan_mgt.fixed.timestamp	8	50
8	frame.len	8	50
38	radiotap.mactime	7	43,75
4	frame.time_epoch	7	43,75
77	wlan.da	7	43,75
79	wlan.sa	6	37,5
154	data.len	6	37,5
47	radiotap.datarate	6	37,5
68	wlan.fc.ds	6	37,5
71	wlan.fc.pwrmgt	5	31,25
75	wlan.duration	5	31,25
67	wlan.fc.subtype	5	31,25
94	wlan_mgt.fixed.capabilities.preamble	5	31,25
108	wlan_mgt.fixed.beacon	5	31,25

frequentemente selecionadas na revisão de literatura. Adicionalmente, outros 52 características foram incluídas em ao menos um conjunto proposto para a detecção de intrusões em redes Wi-Fi (802.11) pela comunidade acadêmica. Embora a maioria dos estudos explore diferentes tipos de ataques, incluindo a falsificação, alguns focam especificamente neste tipo de ataque (AMINANTO; KIM, 2017) (AMINANTO et al., 2018) (AMINANTO et al., 2016).

A característica mais comum entre os estudos é o número de sequência (*wlan.seq*), presente em todos os quadros transmitidos em uma comunicação 802.11, com exceção dos quadros de controle. Este número normalmente varia de 0 a 4.095 e aumenta com cada quadro transmitido consecutivamente. Monitorar alterações nesta características pode ajudar a identificar atividades maliciosas, e está presente em 11 dos 16 conjuntos de características analisados (ALOTAIBI, 2016).

Em relação à frequência de escolha, as características relacionados ao carimbo de tempo (*wlan_mgt.fixed.timestamp*) e ao comprimento dos quadros (*frame.len*) são também altamente valorizados para detecção de intrusões, uma vez que fornecem dados importantes sobre a ordem temporal e o volume de tráfego nas redes Wi-Fi (802.11) (ALOTAIBI, 2016).

Neste trabalho de conclusão de curso, planeja-se superar a falta de consolidação na seleção de características empregando técnicas de XAI, que proporcionam uma maior clareza e entendimento sobre as decisões tomadas pelos modelos de IA, facilitando assim a identificação e seleção das características mais impactantes para a detecção de intrusões em redes Wi-Fi (802.11).

3.3 Método para a Avaliação

A avaliação da hipótese deste estudo será realizada por meio de uma metodologia criteriosa, que inclui a definição de métricas de desempenho, seleção de características, uso de uma base de dados específica bem como todo seu pré-processamento necessário para os moldes de nossa pesquisa. As métricas de avaliação selecionadas são precisão, *recall*, *F1-Score* e acurácia, que são padrões reconhecidos para aferição da eficácia em modelos de classificação.

A base de dados utilizada para treinamento e teste dos modelos será a mesma empregada nos trabalhos de referência, obtida e disponível na (University of the Aegean, 2024). Essa base é amplamente reconhecida pela comunidade científica na área de segurança de redes e oferece uma representação fidedigna de tráfego de rede com uma variedade abrangente de ataques de personificação, essenciais para o teste de robustez dos modelos de IDS.

3.3.1 Métricas de Avaliação

Para avaliar o desempenho dos modelos de classificação, utiliza-se as seguintes métricas:

3.3.1.1 Precisão (Precision)

A precisão avalia a qualidade das previsões positivas do modelo. Importante em contextos onde os falsos positivos são críticos, é calculada por:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

Um alto valor indica que o modelo é assertivo nas previsões positivas que faz, mas deve ser considerado junto ao recall para entender a eficácia geral na classificação positiva.

3.3.1.2 Recall ou Sensibilidade

O recall mede a capacidade do modelo em capturar todas as instâncias reais positivas, essencial em situações onde não detectar positivos é problemático:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

Um alto valor de recall sugere que o modelo captura a maioria das instâncias positivas, mas pode também estar acompanhado de um número elevado de falsos positivos.

3.3.1.3 F1-Score

O F1-Score combina precisão e recall em uma única métrica que busca equilíbrio entre as duas:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

Serve como uma medida da eficiência geral do modelo quando a distribuição de classes é desigual.

3.3.1.4 Acurácia

A acurácia reflete a proporção de todas as previsões corretas feitas pelo modelo:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Enquanto uma métrica intuitiva, a acurácia pode ser enganosa em casos de grande desequilíbrio de classes e deve ser analisada em conjunto com outras métricas para um entendimento completo do desempenho do modelo.

Capítulo 4

Implementações e Resultados

Este capítulo se dedica ao detalhamento das etapas de implementação e experimentação do modelo de detecção de intrusões. O foco inicial está no processo de pré-processamento do conjunto de dados, uma fase vital para assegurar a qualidade e a relevância dos dados que alimentarão o modelo de aprendizado de máquina. A integridade e a precisão dos dados são essenciais para a eficácia do modelo em identificar ataques de personificação, tornando o pré-processamento um passo crítico na preparação para as fases subsequentes de treinamento e validação do modelo. Este processo abrange desde a conversão do formato dos dados até a seleção e codificação das características, estabelecendo uma base sólida para as etapas de implementação e experimentação que serão abordadas em seguida.

4.1 Pré-processamento do *Dataset*

A etapa de pré-processamento é crucial na modelagem para aprendizado de máquina, destacando-se no uso de modelos como o XGBoost para a detecção de ataques. Os dados, em seu estado bruto, podem apresentar problemas como valores ausentes ou formatos não adequados, influenciando diretamente a performance dos modelos. Em outros casos, valores vazios ou ausentes, podem influenciar na detecção de um ataque, dessa forma, não necessariamente estes valores são errados ou desconexos com o ambiente deste trabalho.O XGBoost, por sua vez, é eficiente ao lidar com esses dados incompletos, ignorando automaticamente os valores ausentes ou com formatos desconhecidos para a o aprendizado de máquina, o que nos permite também uma análise robusta e confiável mesmo em condições não ideais.

A primeira ação no processo de pré-processamento é a conversão do *dataset* bruto, inicialmente em formato não estruturado, para um formato CSV, que é mais estruturado e adequado para o armazenamento de dados tabulares. Este processo envolve a leitura do arquivo original e a reorganização de cada linha para que sejam segmentadas por vírgulas, resultando em um arquivo CSV novo e estruturado. Essa transformação é feita utilizando

funções básicas de leitura e escrita de arquivos, juntamente com operações de manipulação de strings para assegurar o formato correto dos registros.

Após converter os dados para o formato CSV, eles são carregados em uma estrutura de *DataFrame*, facilitando a manipulação subsequente com o auxílio de bibliotecas especializadas, como o *Pandas*. Nesse momento, também se define os nomes das colunas do *DataFrame*, baseando-se no conhecimento do domínio e na documentação técnica disponível, o que é crucial para a clareza e eficácia na manipulação dos dados durante as fases de análise e seleção de características.

O próximo passo é a filtragem e limpeza dos dados, definindo critérios específicos para selecionar apenas as informações relevantes para o estudo. No caso deste projeto, o *dataset* é filtrado para incluir apenas determinadas classes de interesse, como os ataques de personificação *Cafe Latte*, *Normal*, *Hirte* e *Evil Twin*. Isso é feito aplicando condições lógicas às colunas pertinentes. Os dados filtrados são então salvos em um novo arquivo CSV, preparado para ser utilizado nas etapas de treinamento e teste dos modelos de aprendizado de máquina.

Este processo de pré-processamento garante que os dados estejam corretamente preparados e estruturados, minimizando erros e maximizando a eficiência e a confiabilidade dos modelos de predição desenvolvidos. A manipulação e filtragem cuidadosa dos dados são fundamentais para garantir que apenas as informações pertinentes sejam consideradas, proporcionando uma base sólida para a análise subsequente e para a tomada de decisões baseadas em dados.

A divisão de treino e teste adotada neste estudo reflete a estrutura proposta no artigo estendido, garantindo a integridade e a replicabilidade da análise em conformidade com os padrões acadêmicos estabelecidos.

4.2 Código Principal de Treino e Avaliação

O código principal deste projeto encapsula a lógica essencial para o pré-processamento do *dataset*, treinamento do modelo de classificação e avaliação do seu desempenho, seguindo a estrutura proposta no artigo original que divide os dados em conjuntos de treino e teste. As etapas e métodos empregados são descritos a seguir de forma detalhada e impessoal.

Inicialmente, os *datasets* de treino e teste são carregados, acompanhados de uma cuidadosa seleção de características. Esta fase é vital para garantir que apenas as características relevantes e corretamente preparadas sejam utilizadas no treinamento e avaliação do modelo. Durante este processo, matrizes de características e rótulos associados são separados e organizados, preparando o cenário para a manipulação e análise subsequente.

Segue-se a conversão de valores categóricos para numéricos nos conjuntos de dados. Essa transformação, realizada através de técnicas de codificação, é essencial para que os

algoritmos de aprendizado de máquina possam interpretar e processar os dados eficientemente, eliminando barreiras linguísticas que poderiam impedir o fluxo de análise.

O próximo passo é o treinamento do modelo utilizando o algoritmo de *boosting* XGBoost, que é conhecido por sua eficácia em construir modelos de árvores de decisão sequenciais que melhoram continuamente as previsões. O modelo é ajustado utilizando os dados de treino para aprender a correlação entre as características e os rótulos.

Uma vez treinado, o modelo é rigorosamente avaliado utilizando o conjunto de teste. Métricas de desempenho, como precisão, *recall*, *F1-score* e acurácia, são calculadas para determinar a eficácia do modelo. Estas métricas proporcionam uma compreensão quantitativa do desempenho do modelo, facilitando a identificação de pontos fortes e áreas que podem necessitar de melhorias.

Por fim, para aprofundar a compreensão do comportamento do modelo e entender a influência de cada característica nas decisões tomadas, utiliza-se a ferramenta de explicabilidade SHAP. Esta ferramenta fornece visualizações e análises detalhadas, mostrando o impacto de cada característica nas previsões do modelo, o que permite obter *insights* valiosos sobre quais atributos são mais determinantes nas classificações.

Essas etapas são fundamentais não só para validar a abordagem proposta no artigo original, mas também para garantir que o modelo ofereça não apenas uma classificação eficiente, mas transparência e compreensão sobre suas previsões, que são cruciais para a confiança nas decisões automatizadas em contextos de segurança de rede.

4.3 Pré-processamento do Dataset feito em Python

Este pré-processamento envolve várias etapas detalhadas abaixo, que preparam o *dataset* para ser utilizado de forma eficaz nos modelos preditivos.

4.3.1 Instalação de Pacotes Necessários

Assegura-se que todas as bibliotecas necessárias estejam instaladas e prontas para uso, incluindo bibliotecas especializadas para manipulação de dados e modelagem preditiva, como **Pandas**, **NumPy**, **XGBoost** e **SHAP**, todas essas bibliotecas estão contidas no arquivo *prep_requirements.txt*

```
%pip install -r prep_requirements.txt
```

4.3.2 Conversão de Dados Brutos para Formato CSV

Os dados brutos são lidos de um arquivo e convertidos para o formato CSV, o qual é mais adequado para manipulação em ferramentas de análise de dados:

```
import csv
```

```
import pandas as pd

with open("dataset/AWID-ATK-R-Tst/1", "r") as arquivo_entrada:
    with open("dataset/AWID-ATK-R-Tst/saida_Tst.csv", "w", newline="")
    as arquivo_saida:
        writer = csv.writer(arquivo_saida)
        for linha in arquivo_entrada:
            writer.writerow(linha.strip().split(","))
```

4.3.3 Carregamento e Estruturação do *DataFrame*

Após a conversão, o arquivo CSV é carregado como um *DataFrame*. Atribuem-se nomes às colunas com base no conhecimento do domínio:

```
df = pd.read_csv("dataset/AWID-ATK-R-Tst/saida_Tst.csv", sep=",", header=None)
column_names = [...]
df.columns = column_names
```

4.3.4 Filtragem de Dados

Filtra-se os dados para incluir apenas as classes de interesse, e o resultado é salvo em um novo arquivo CSV:

```
valores_permitidos = ["evil_twin", "cafe_latte", "normal", "hirte"]
df_filtrado = df[df["class"].isin(valores_permitidos)]
df_filtrado.to_csv("dataset/AWID-ATK-R-Tst/dataset_Tst_limpo.csv", index=False)
```

Esta abordagem assegura que os dados estejam corretamente preparados e estruturados, minimizando erros e maximizando a eficiência e a confiabilidade dos modelos de predição desenvolvidos.

4.4 Código Principal de Treino e Avaliação em Python

Este código encapsula a lógica fundamental para o treinamento do modelo de classificação e avaliação de seu desempenho. O processo segue a metodologia proposta no artigo original, que organiza os dados em conjuntos de treino e teste. As operações realizadas são descritas a seguir.

4.4.1 Instalação de Pacotes Necessários

Antes de iniciar o processamento, garante-se que todas as bibliotecas necessárias estejam disponíveis, incluindo **XGBoost** para treinamento do modelo e **SHAP** para in-

terpretabilidade, **Pandas** para a leitura e manipulação do *dataset* e **matplotlib** para plotagem de outros gráficos. todos esse módulos para instalação constam no arquivo de texto *requeriments*

```
%pip install requirements.txt
```

4.4.2 Carregamento e Seleção de Características

Os *datasets* de treino e teste são carregados. Selecionam-se apenas as características relevantes para a análise, baseadas em critérios pré-definidos:

```
import pandas as pd

train_df = pd.read_csv("path/to/train.csv")
test_df = pd.read_csv("path/to/test.csv")

# Seleção de características baseada em análise prévia
features = ['feature1', 'feature2', 'feature3', ...]
train_df = train_df[features]
test_df = test_df[features]
```

4.4.3 Codificação de Valores Categóricos

Converte-se valores categóricos para numéricos, utilizando técnicas de codificação, para que possam ser processados pelos algoritmos de aprendizado:

```
from sklearn.preprocessing import LabelEncoder

label_encoders = {}
for column in train_df.columns:
    if train_df[column].dtype == 'object':
        le = LabelEncoder()
        train_df[column] = le.fit_transform(train_df[column])
        test_df[column] = le.transform(test_df[column])
        label_encoders[column] = le
```

4.4.4 Treinamento do Modelo

O modelo é treinado utilizando o algoritmo de *boosting* XGBoost, conhecido pela eficácia em construir árvores de decisão sequenciais:

```
import xgboost as xgb
```

```
# Configuração de parâmetros para o modelo
params = {'max_depth': 10, 'eta': 0.1, 'objective': 'multi:softprob', 'num_class':
dtrain = xgb.DMatrix(train_df.drop('label', axis=1), label=train_df['label'])
model = xgb.train(params, dtrain, num_boost_round=100)
```

4.4.5 Avaliação do Modelo

Após o treinamento, o modelo é avaliado utilizando várias métricas:

```
from sklearn.metrics import precision_score, f1_score,
recall_score, accuracy_score

y_pred = model.predict(dtest)
y_true = test_df['label']
precision = precision_score(y_true, y_pred, average='macro')
recall = recall_score(y_true, y_pred, average='macro')
f1 = f1_score(y_true, y_pred, average='macro')
accuracy = accuracy_score(y_true, y_pred)

print(f'Precision: {precision}, Recall: {recall}, F1-Score: {f1},
Accuracy: {accuracy}')
```

4.4.6 Análise de Importância das Características com SHAP

Utiliza-se SHAP para entender a influência de cada característica nas previsões do modelo:

```
import shap

explainer = shap.TreeExplainer(model)
shap_values = explainer.shap_values(X_train)

shap.summary_plot(shap_values, X_train, plot_type="bar")
```

Este fluxo de trabalho garante que o modelo não apenas seja eficiente, mas também que suas decisões sejam transparentes e compreensíveis, um aspecto crucial para aplicações práticas.

4.5 Resultados

Esta seção discorre sobre os resultados obtidos na investigação das técnicas de detecção de ataques de personificação em redes Wi-Fi (802.11). Primeiramente, na Seção 4.5.1, é demonstrada uma análise comparativa de diferentes conjuntos de características extraídos da literatura. Em seguida, na Seção 4.5.2, serão reportados os resultados obtidos através da ferramenta SHAP, usada para alcançar-se a explicabilidade dos modelos treinados.

4.5.1 Comparação de Conjuntos

A análise compreensiva dos diferentes conjuntos de características e a aplicação de métodos avançados de aprendizado de máquina permitiram a identificação de padrões significativos e a eficácia variada entre os conjuntos avaliados. A seguir, apresentam-se os detalhes dos resultados alcançados, discutindo-se os conjuntos de características que se destacaram nas métricas de precisão, *recall* e *F1-Score*.

4.5.1.1 Acurácia

Ao analisar os resultados da Figura 5, é possível analisar que o conjunto C2 se destacou em termos da métrica acurácia, atingindo a maior marca entre os 15 conjuntos analisados: 97.06%. Já os conjuntos C11, C7, C9 e C14, respectivamente, atingiram acurácias entre 96.25% e 95.37%. Os demais conjuntos não são exibidos no gráfico pois apresentaram métricas em níveis demasiadamente inferiores e, portanto, não serão alvo da discussão deste trabalho. Como ilustrado na Figura 5, a acurácia dos modelos variou

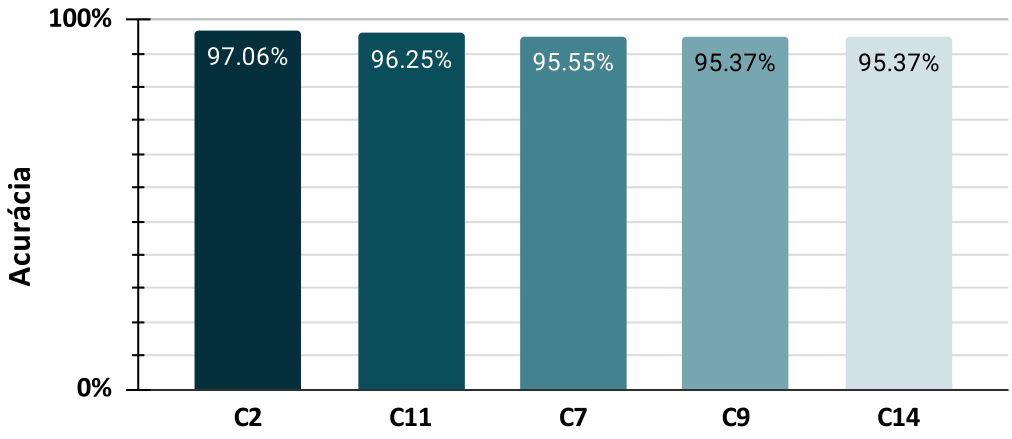

Figura 5 – Acurácia dos 5 melhores conjuntos de características.

conforme os conjuntos de características utilizados, em sua maioria alcançando uma acurácia substancialmente mais alta, o que sugere que as características neles incluídas são mais informativas para a classificação dos dados de rede.

Isso indica que, em termos de acurácia, os ataques de personificação são mais eficientemente detectados pelo conjunto de características ‘data.len’,‘frame.len’,‘wlan.fc.subtype’, ‘wlan.duration’,‘wlan.fc.pwrmgt’,‘radiotap.datarate’,‘wlan.qos.tid’ e ‘radiotap.channel.type.ofdm’, que constituem o conjunto 2. Na Seção 4.5.2, exploraremos essas características utilizando ferramentas de explicabilidade apropriadas. Essa análise tem como objetivo oferecer uma interpretação mais profunda sobre por que essas características específicas são significantes para a detecção dos ataques examinados.

4.5.1.2 Recall

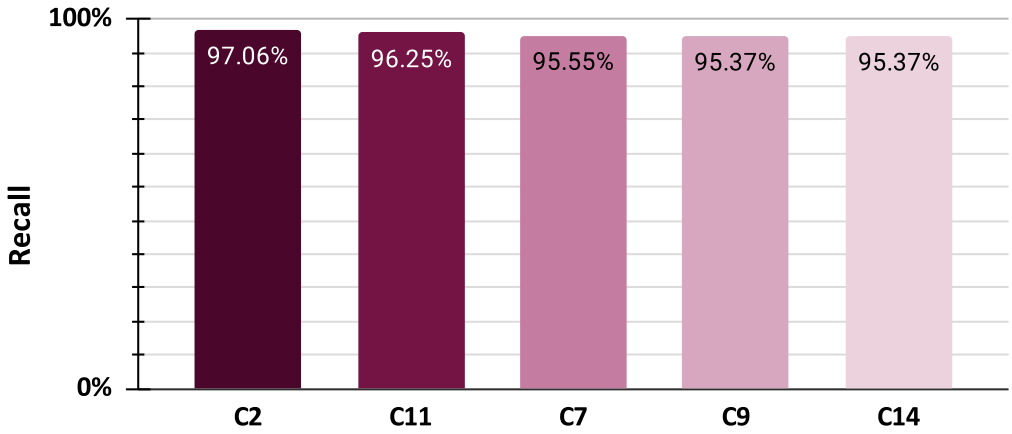

Figura 6 – *Recall* dos 5 melhores conjuntos de características.

O *recall*, representado na Figura 6, é uma métrica especialmente importante em contextos de segurança, pois indica a capacidade do modelo de identificar verdadeiros positivos. Os conjuntos de características com os melhores *recall* sugerem uma habilidade superior para capturar instâncias de ataque, o que é essencial para minimizar a taxa de falsos negativos.

4.5.1.3 Precisão

A precisão, mostrada na Figura 7, é outra métrica chave, pois informa sobre a proporção de identificações positivas do modelo que foram efetivamente corretas. Conjuntos de características que resultam em alta precisão são cruciais para reduzir falsos positivos, o que é crítico para evitar alarmes desnecessários e sobrecarga nos times de resposta a incidentes.

4.5.1.4 F1-Score

Finalmente, o *F1-Score*, uma média harmônica entre precisão e *recall*, é mostrado na Figura 8. Esta métrica é útil quando busca-se um equilíbrio entre precisão e *recall*, o

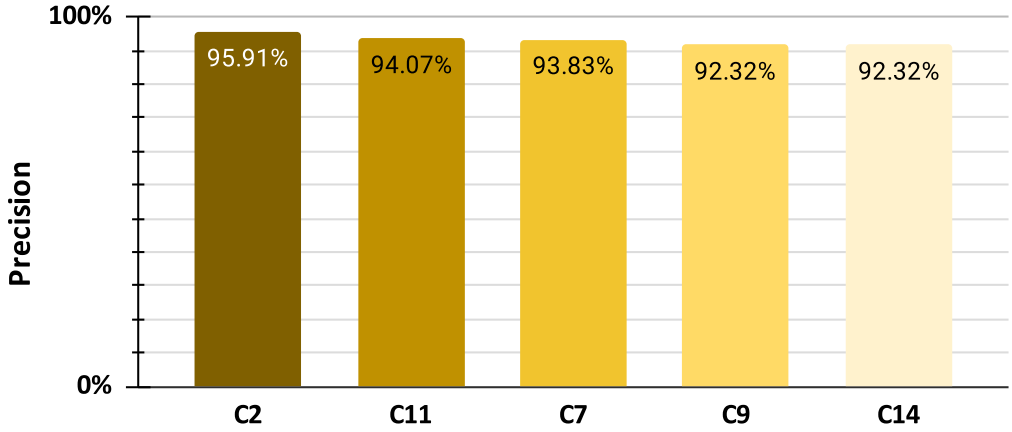

Figura 7 – Precisão dos 5 melhores conjuntos de características.

que é frequentemente o caso em sistemas de detecção de intrusões, onde ambos, identificar corretamente as intrusões e evitar a classificação errônea de atividades normais, são igualmente importantes.

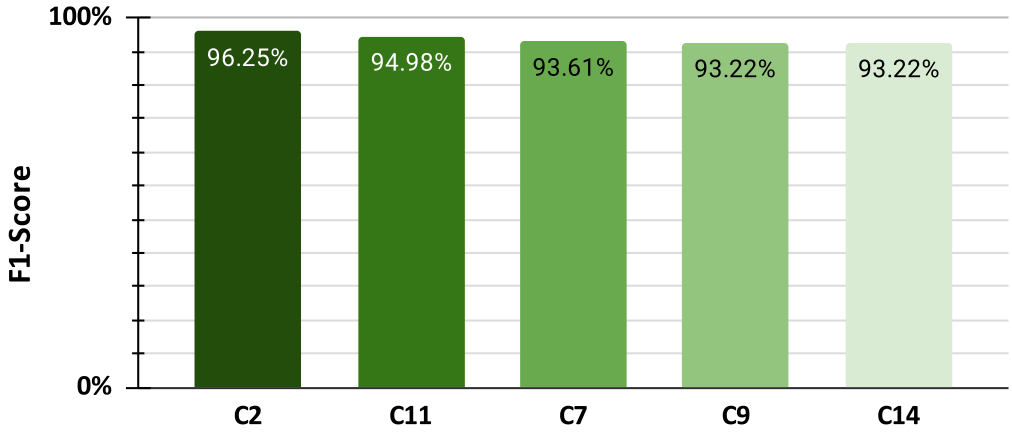

Figura 8 – F1-Score dos 5 melhores conjuntos de características.

Esses resultados reforçam a importância de uma cuidadosa seleção e avaliação de conjuntos de características, e estabelecem um forte argumento para a seleção de características como uma etapa crítica na melhoria dos sistemas de detecção de intrusões.

4.5.2 Explicabilidade das características

Nesta seção, será explorado o uso da técnica SHAP para elucidar a contribuição individual de cada característica na performance de modelos preditivos aplicados à detecção dos ataques de personificação estudados. Em particular, a aplicação de técnicas de XAI proporcionou uma visão detalhada sobre a contribuição de cada característica na performance dos modelos, destacando a importância da seleção e interpretação adequada das características na otimização dos IDS.

Através de gráficos intuitivos gerados pelo SHAP, será demonstrado a seguir como diferentes características influenciam a decisão final do modelo, proporcionando *insights* valiosos sobre a dinâmica e potenciais vulnerabilidades do sistema. As análises detalhadas a seguir baseiam-se nas características previamente discutidas na Subseção 3.2.2.

A aplicação de técnicas interpretáveis de aprendizado de máquina, como a análise SHAP, surge como um recurso valioso para aprimorar sistemas de detecção de intrusões, em particular no diagnóstico de ataques de personificação. Com relação a segurança de redes sem fio, a personificação refere-se à prática maliciosa de imitar dispositivos confiáveis para ganhar acesso não autorizado ou perturbar as operações de rede.

4.5.2.1 Summary Plot: Classe Normal

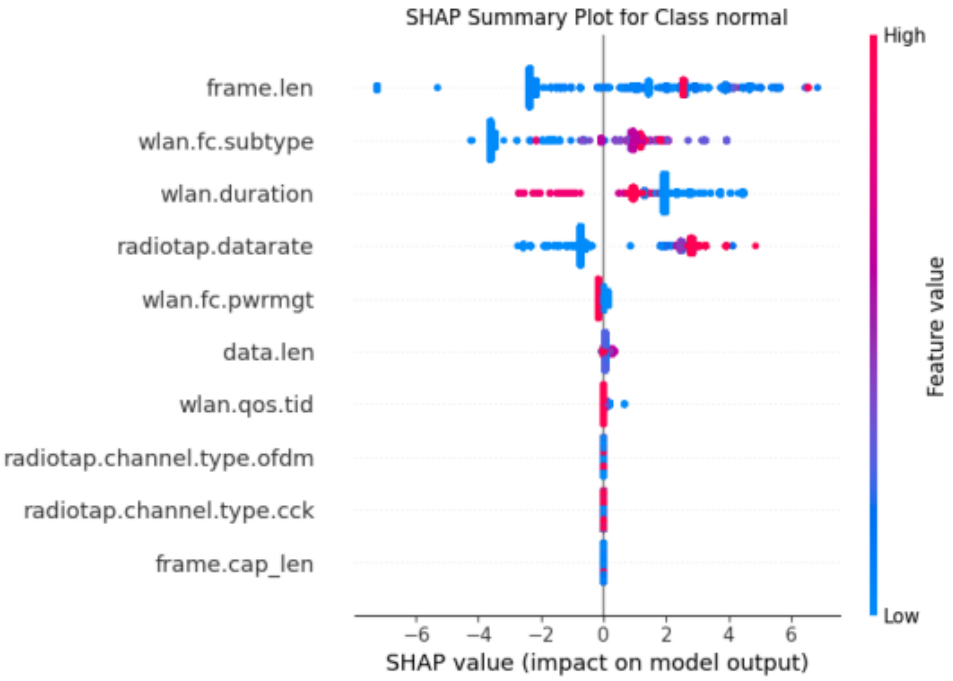

Figura 9 – Gráfico *Summary Plot* com resultados de ataques para a classe normal.

O gráfico representado pela Figura 9, fornece informações detalhadas sobre como características especificas influenciam a identificação de atividades legitimas x maliciosas. O gráfico mostra o impacto SHAP de cada característica na saída do modelo para a classe "normal". O eixo x SHAP mostra como a presença ou ausência de uma característica afeta a classificação do modelo. Valores à esquerda do zero indicam um impacto negativo (ou uma associação com a classe não-normal), enquanto valores à direita indicam um impacto positivo (ou uma associação com a classe normal).

Os pontos representam a distribuição dos valores SHAP para cada característica em muitas instâncias do conjunto de dados. A cor dos pontos (do azul ao vermelho) representa o valor real da característica (do baixo ao alto, respectivamente). A presença de

pontos tanto no lado negativo quanto no positivo para uma característica sugere que essa característica tem um efeito variável sobre a saída do modelo, dependendo do contexto ou do valor da característica em uma dada observação.

Características como *frame.len* e *data.len*, que representam comprimentos dos quadros e dos dados, podem indicar uma anomalia quando desproporcionalmente grandes ou pequenos, divergindo dos padrões usuais de tráfego. O *wlan.fc.subtype* e o *wlan.qos.tid* indicam o propósito e a prioridade do quadro, que divergem dos padrões esperados, podem sugerir tentativas de ataques de personificação. As técnicas de modulação *radiotap.channel.type.ofd* e *radiotap.channel.type.cck* também fornecem informações valiosas, pois os invasores podem manipular esses parâmetros para imitar dispositivos legítimos ou efetuar ataques de negação e serviço.

4.5.2.2 Histograma: Classe Evil Twin

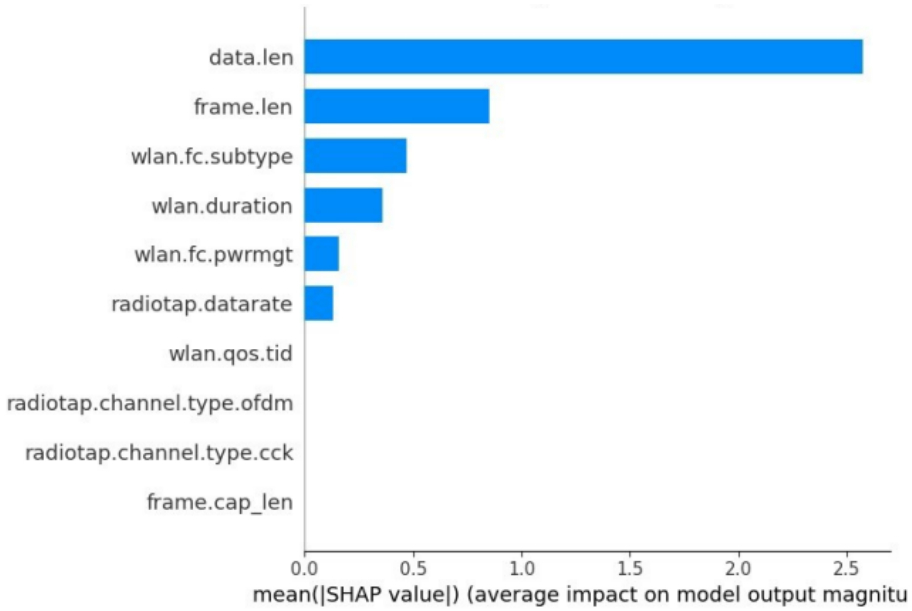

Figura 10 – Histograma com resultados de ataques para a classe *Evil Twin*.

O histograma SHAP para a classe 'evil_twin',representado na Figura 10, oferece informações quantitativas sobre a média do impacto absoluto (valor médio de SHAP) de cada característica no resultado de um modelo de ML destinado a identificar ataques gêmeos 'evil_twin' em redes WLAN.

Os ataques gêmeos são uma forma sofisticada de ataque de personificação, onde um ponte de acesso sem fio malicioso imita o **SSID** e **MAC** de um ponto de acesso legítimo, enganando os usuários a se conectarem a ele.

Percebe-se na Figura 10 um destaque para o *data.len* e *frame.len* como características mais influentes, sugerindo que a quantidade de dados e o tamanho dos quadros são in-

dicativos significativos de atividade maliciosa. Uma irregularidade nesses valores podem indicar tentativas de interceptação ou transmissão de *payloads* maliciosos.

Na sequência observa-se o *wlan.fc.subtype* mostra a relevância dos tipos específicos de quadros na identificação de ataques anormais, onde quadros de associação e reautenticação que ocorrem com frequência em ataques gêmeos.

Parâmetros como *wlan.duration* e *wlan.fc.pwrmgt* nos informam a duração e o gerenciamento de energia dos quadros, cujas anormalidades podem sinalizar atividades suspeitas.

A taxa de transmissão de dados *(radiotap.datarate)* e os tipos de canais utilizados *(radiotap.channel.type.ofdm / radiotap.channel.type.cck)* completam as características, pois variações atípicas nestas características podem indicar a presença de um atacante tentando replicar ou perturbar uma comunicação regular.

4.5.2.3 Summary Plot: Classe Evil Twin

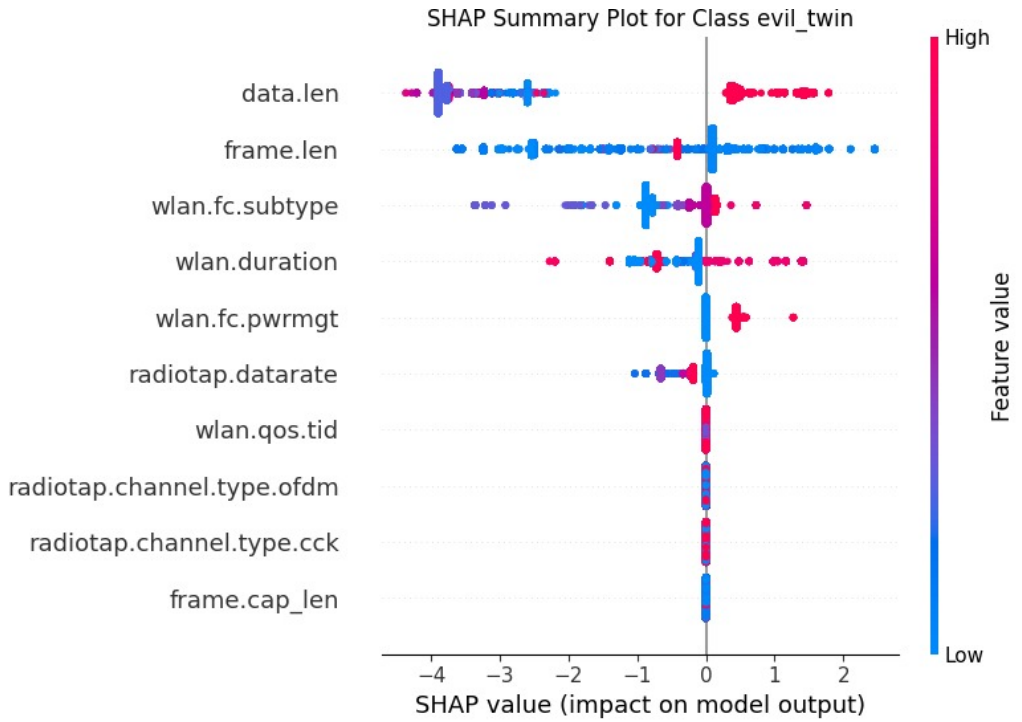

Figura 11 – *Summary Plot* com resultados de ataques para a classe *Evil Twin*.

"O ataque *Evil Twin* se baseia na clonagem de AP a fim de que a vítima se conecte ao APmalicioso ao invés do legítimo. Para tanto, é importante que o AP clonado ofereça um sinal superior ao AP legítimo. Desse modo, a vítima passa a se comunicar com o ponto de acesso malicioso, possibilitando ao atacante a interceptação e manipulação de toda a comunicação sem o conhecimento da vítima. Portanto, basicamente, o atacante age como um Homem-

> No-Meio, do inglês, *Man-In-The-Middle* (MITM) [17]."(QUINCOZES; KAZIENKO; COPETTI, 2018)

O *Summary Plot* para a classe *Evil Twin*, representado na Figura 11, reflete o impacto individual de cada característica na detecção de ataques gêmeos, uma forma de ataque de personificação em redes sem fio. No gráfico, cada ponto representa o efeito de uma característica na predição do modelo, com o eixo x mostrando o valor SHAP, que quantifica este impacto. Valores positivos indicam uma tendência da característica de aumentar a probabilidade de uma observação ser classificada como um ataque gêmeo, enquanto valores negativos sugerem contrário.

As características com maior distanciamento horizontal, como *data.len* e *frame.len*, possuem influência variada, indicando que dependendo do valor que assumem, podem tanto aumentar quanto diminuir a probabilidade de identificação de um ataque. A variação dos pontos que vai do azul ao vermelho, demonstra o valor real da característica, com o azul representando valores baixos e o vermelho, altos. Essa variação de cores permite identificar padrões nos dados: por exemplo, valores mais altos de *data.len* podem estar frequentemente associados a ataques gêmeos 'evil_twin'.

Seguindo analisando o gráfico, percebe-se algumas características com relação ao tamanho e tipo dos quadros *(frames)*, taxa de dados e protocolos de modulação que desempenham um papel importante na classificação dos ataques. A compreensão detalhada dessas relações é crucial para o aprimoramento de sistema de detecção de intrusões, permitindo a implementação de medidas de segurança mais direcionadas e eficazes.

Ao captar a complexidade e as variações das contribuições das características, o modelo oferece uma visão mais rica do que constitui um comportamento de rede anormal, guiando os analistas a focarem em aspectos específicos do tráfego de rede para identificar e neutralizar ataques gêmeos.

4.5.2.4 Summary Plot: Cafe Latte

(QUINCOZES; KAZIENKO; COPETTI, 2018) (Avaliação de Conjuntos de Atributos para a Detecção de Ataques de Personificação na Internet das Coisas)

> "Ao executar o ataque *Cafe Latte*, o atacante não precisa estar no raio de alcance de pontos de acessos da rede alvo, pois diferentemente do *Evil Twin*, as vítimas desse ataque são clientes isolados de redes protegidas pelo protocolo WEP. Isso acontece porque muitos dispositivos costumam armazenar as credenciais de acesso a fim de estabelecer conexão de forma automática quando o cliente estiver dentro do alcance do AP[18]. Assim, ao forjar requisições ARP para a vítima, é possível a quebra da chave WEP [19]." (QUINCOZES; KAZIENKO; COPETTI, 2018)

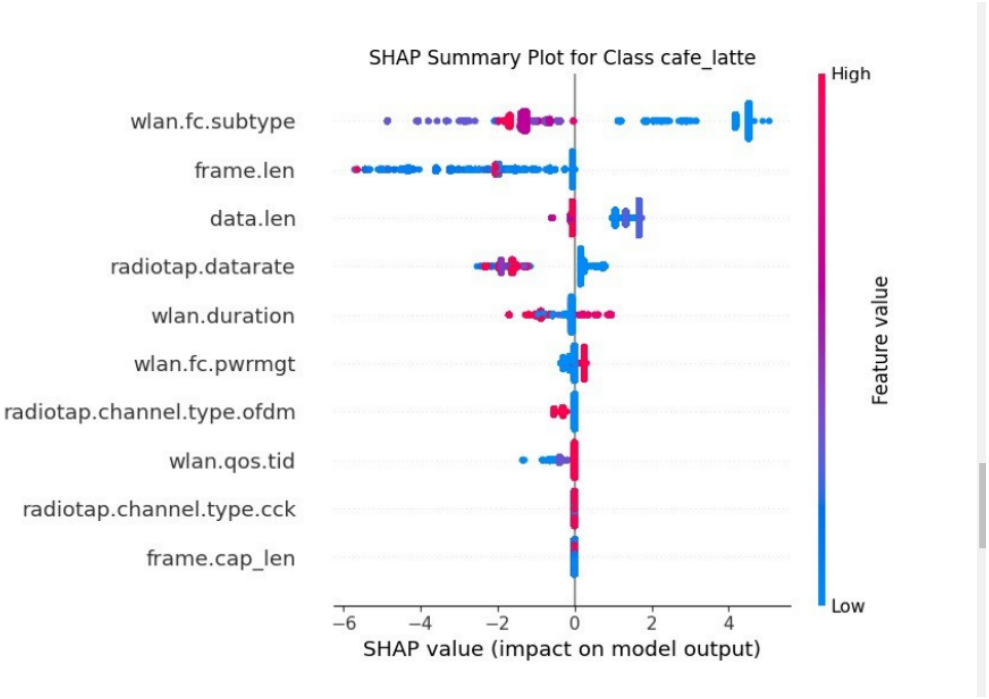

Figura 12 – *Summary Plot* com resultados de ataques para a classe *Cafe Latte*.

O gráfico SHAP *Summary Plot* para a classe *(Cafe Latte)*, representado na Figura 12, oferece uma análise interpretativa das características influenciando a classificação de atividades de rede associadas a um ataque de personificação específico, conhecido como ataque *Cafe Latte*. Este tipo de ataque explora vulnerabilidades em redes WLAN para interceptar informações de autenticação.

As características listadas no gráfico proporcionam ideias sobre o impacto de cada uma na decisão do modelo de identificar uma atividade como sendo parte desse ataque.

A característica *wlan.fc.subtype* indica o subtipo do quadro de controle WLAN, fornecendo pistas sobre a natureza da atividade de rede; por exemplo, subtipos associados a processos de autenticação podem ser manipulados em um ataque *Café Latte*.

Frame.len e *data.len* refletem, respectivamente, o comprimento total do quadro e o comprimento dos dados transmitidos, onde desvios significativos dos padrões típicos podem sugerir uma tentativa de ataque, como o envio de pacotes de dados estranhamente grandes ou pequenos para confundir ou desorganizar processos de autenticação.

A *radiotap.datarate*, ou taxa de transmissão de dados, oferece indicações da velocidade de comunicação; variações inesperadas podem ser sintomas de um atacante tentando se ajustar para se passar por um ponto de acesso legítimo.

A *wlan.duration* pode revelar tentativas de ocupar o canal por tempos anormais, o que pode ser um artefato de ataques. Por outro lado, *wlan.fc.pwrmgt*, associado à gestão de energia, pode indicar o uso de estratégias para prolongar a vida útil da bateria em dispositivos comprometidos.

As características *radiotp.channel.type.ofdm* e *radiotap.channel.type.cck* representam

a utilização de métodos de modulação específicos, com variações atípicas podendo indicar alterações na comunicação pertinentes a ataques.

A característica *wlan.qos.tid* mostra a identificação de tráfego para qualidade de serviço, que, se manipulada, pode interferir na priorização do tráfego e ser explorada em ataques. Por fim, *frame.cap.len* apresenta o tamanho de captura, sendo importante na análise de padrões de tráfego para a detecção de anomalias. Cada ponto no gráfico SHAP representa o efeito de uma observação no modelo, com a posição no eixo x refletindo o impacto na probabilidade de ser um ataque *Cafe Latte*.

A distribuição dos pontos em torno do zero e a codificação de cores (do azul para baixos valores de característica ao vermelho para altos valores) ressaltam a complexidade e a heterogeneidade dos dados. Este gráfico desvenda o relacionamento entre as características e a classificação de ataques, permitindo ao ser humano compreender e aprimorar os mecanismos de detecção e prevenção de ameaças em redes WLAN.

4.5.2.5 Discussão

A aplicação da análise SHAP em nossos estudos destaca-se como uma ferramenta interpretativa robusta, oferecendo vislumbres essenciais sobre a dinâmica entre as características de uma rede e a presença de atividades maliciosas. Através do exame preciso de gráficos SHAP *Summary Plot* para diversas classes de *tráfego-normal*, 'evil_twin' e 'cafe_latte' emergem padrões distintos de comportamento das características que compõem a teia de comunicação em redes sem fio. As variáveis como *frame.len*, *data.len*, *wlan.fc.subtype*, e *wlan.qos.tid* provaram ser de valor precioso na distinção entre o tráfego legítimo e potenciais ataques de personificação, revelando anomalias quando comparadas com os padrões de tráfego habituais.

Técnicas de modulação como *radiotap.channel.type.ofdm* e *radiotap.channel.type.cck* e parâmetros como *wlan.fc.pwrmgt* e *radiotap.datarate* fortalecem ainda mais o poder de detecção, pois desvios nestas características podem sinalizar tentativas de intrusão e interferência no fluxo de comunicação. O modelo SHAP auxilia não apenas na identificação de ameaças, mas também serve como uma ponte para uma compreensão mais profunda dos métodos utilizados pelos invasores, abrindo caminho para o desenvolvimento de estratégias de prevenção mais precisas e proativas.

Portanto, entende-se que a interpretabilidade proporcionada pelo SHAP é fundamental para o desenvolvimento de sistemas de segurança cibernética mais eficazes. Ela permite aos especialistas em segurança de rede não só detectar ataques de personificação com maior precisão, mas também entender os mecanismos por trás destes ataques, facilitando a implementação de soluções específicas para a proteção da integridade e confidencialidade das informações em redes WLAN. Com a crescente sofisticação das ameaças cibernéticas, a clareza fornecida pela análise SHAP é um recurso valioso na frente da defesa cibernética,

habilitando profissionais a se anteciparem e combaterem os desafios de segurança em um mundo cada vez mais conectado.

Capítulo 5

Análise dos Resultados

5.1 Avaliação dos Resultados

Este segmento do documento destaca o impacto e a significância de variadas características analisadas nos modelos preditivos, usando gráficos de SHAP para aprofundar a compreensão das relações entre atributos e classificações de tráfego de rede, é feito uma leitura dessas características para cada contexto de rede dos ataques selecionados.

5.1.1 Características para a classe Normal

O gráfico de resumo SHAP da Figura 9 ilustra o impacto de cada característica no resultado do modelo.

- As características mais influentes foram identificadas como `frame.len`, `wlan.fc.subtype`, e `wlan.duration`, conforme evidenciado por sua posição superior no gráfico.
- Altos valores de `frame.len`, `wlan.fc.subtype` e `wlan.duration` estão significativamente relacionados à classificação do tráfego como normal.
- A distribuição dos valores SHAP para características abaixo do `wlan.fc.pwrmgt` mostraram-se mais concentradas ao meio do gráfico, indicando um efeito mais consistente na predição do modelo e pouco efetivas quando se trata de um ambiente anormal de rede.

5.1.2 Características para o *Cafe Latte*

Com base no sumário SHAP fornecido na Figura 12 , procede-se a uma análise imparcial das características relevantes para a classificação do ataque "cafe_latte". Cabe destacar que os pontos azuis no sumário SHAP indicam valores baixos da característica, enquanto os pontos vermelhos representam valores altos. Uma disposição assimétrica destes pontos em relação ao eixo vertical sugere que o impacto da característica no modelo

varia com seu valor. Da mesma forma que, valores concentrados à esquerda do gráfico nos indicam maior predominância de um possível ataque, logo, ao lado direito representa o oposto, um cenário de normalidade.

- A característica **wlan.fc.subtype** demonstra uma distribuição ampla dos valores SHAP, sugerindo uma influência variável no modelo. Essa característica exibe tanto associação negativa quanto positiva em relação à classe "cafe_latte", indicando uma diversidade na sua contribuição para a classificação do ataque, no entanto, a presença de pontos mais avermelhados ao lado esquerdo do gráfico, indicam sua predominância no acontecimento do ataque.

- Para **frame.len**, observa-se uma predominância de valores SHAP no espectro negativo, implicando que, em geral, comprimentos de *frame* são associados ao ataque "cafe_latte".

- A característica **data.len** apresenta uma distribuição de valores SHAP menos extensa que **frame.len**, e por concentrar seus valores perto do zero, indica um possível impacto menor na previsão.

- Com relação a **radiotap.datarate**, a maioria dos valores SHAP está situada no lado negativo, o que pode ser interpretado como uma taxa de dados capturada pelo *radiotap* geralmente favorável à classificação como "cafe_latte". Por outro lado, valores no lado positivo sugerem que taxas de dados presentes também podem ser um indicativo de um falso positivo.

- A distribuição dos valores SHAP para **wlan.duration** é mais concentrada negativamente, insinuando que durações mais breves de WLAN estão frequentemente associadas ao ataque "cafe_latte". Valores positivos, embora menos frequentes, indicam que em certas ocasiões, a duração pode ser um fator contribuinte para um ambiente normal da rede.

- **wlan.fc.pwrmgt** exibe uma distribuição de valores SHAP relativamente equilibrada, com uma leve inclinação para o lado negativo, revelando uma associação ambivalente com a classificação do ataque "cafe_latte", visto que os valores à direita são altos também.

- As características **radiotap.channel.type.ofdm** e **radiotap.channel.type.cck** mostram valores SHAP negativos ou concentrados ao meio do gráfico, indicando que tais tipos de canais são menos prováveis de estar associados ao ataque "cafe_latte".

- A característica **wlan.qos.tid** apresenta uma distribuição de valores SHAP quase igual entre os lados negativo e positivo, considerando o peso de cada ponto, implicando um papel variável na determinação da classe "cafe_latte".

- Similarmente, **frame.cap_len** revela um padrão próximo ao do **radiotap.channel.type.cck**, predominando seus valores à zero.

5.1.3 Características para a classe *Evil Twin*

O SHAP *Summary Plot* da Figura 11 descreve a contribuição de cada característica na predição da classe 'evil_twin'.

- Características como `data.len` e `frame.len` mostraram ter a maior importância na classificação do modelo. Altos valores dessas características tendem a impactar positivamente a predição da presença de um ataque *Evil Twin*.
- As características `wlan.fc.subtype` e `wlan.duration` também são relevantes, apresentando uma distribuição de impactos tanto positivos quanto negativos, indicando influências variáveis na predição da classe 'evil_twin'.
- A característica `wlan.fc.pwrmgt` apresentou uma distribuição de valores SHAP principalmente positiva para o acontecimento do ataque, sugerindo uma associação de seus valores positivo e altos estão presentes no contexto do ataque do 'evil_twin'.
- Os recursos `radiotap.channel.type.ofdm`, `radiotap.channel.type.cck`, `frame.cap_len` e `wlan.qos.tid` possuem uma distribuição de impacto mais estreita, concentrando-se no ponto zero, indicando um efeito mais previsível no resultado do modelo.

Em geral, os gráficos indicam uma clara divisão de contribuições entre as características, com algumas tendo uma influência mais direta e outras apresentando uma gama de efeitos mais ampla.

A avaliação dos resultados obtidos neste estudo revela *insights* significativos sobre a eficácia dos conjuntos de características na identificação de ataques de personificação em redes Wi-Fi (802.11). Entre os conjuntos analisados, o Conjunto 2 se destacou consistentemente em todas as métricas de desempenho avaliadas: precisão, *recall*, *F1-Score* e acurácia, estabelecendo-se como a escolha mais relevante e pontual para a detecção desses tipos de ataques.

Esta seção destaca os acertos ao identificar que o Conjunto 2, pela sua composição única de características, foi capaz de capturar com eficiência os indicativos de comportamentos mal-intencionados, mesmo em meio a tráfegos de falsos positivos. Esta constatação valida a hipótese inicial de que uma seleção cuidadosa e estratégica de características pode melhorar significativamente a capacidade de um sistema de detecção de intrusões.

Por outro lado, é crucial reconhecer as limitações identificadas durante a avaliação. Por exemplo, a característica *frame.len* mostrou uma tendência a gerar falsos positivos em um ambiente sem ataques, apontando para uma área que requer atenção adicional e

refinamento. Em cenários de ataque real, no entanto, o Conjunto 2 perpetua sua robustez, identificando com precisão características-chave que não eram tão proeminentes quando em condições normais.

Esta análise enfatiza a importância de equilibrar a sensibilidade do sistema à detecção de ataques com a necessidade de minimizar alertas falsos, um desafio crítico no desenvolvimento de sistemas de detecção de intrusões eficazes. A habilidade do Conjunto 2 em se sobressair em contextos de ataque reitera seu valor, sugerindo que, apesar das dificuldades inerentes à identificação incorreta de comportamentos legítimos como maliciosos, uma análise detalhada e um enfoque minucioso na seleção de características são essenciais para fortalecer a segurança contra ataques de personificação em redes Wi-Fi (802.11).

Portanto, a avaliação dos resultados corrobora a tese central do estudo, demonstrando que a seleção cuidadosa de conjuntos de características é fundamental para a eficácia dos sistemas de detecção de intrusões em identificar ataques de personificação, equilibrando precisão e confiabilidade em um ambiente de rede dinâmico e complexo.tação para quebrar chaves criptográficas em comunicações WEP. Essa abordagem se baseia na divisão da requisição em múltiplos fragmentos e na manipulação do comprimento do primeiro fragmento para alterar o IP de origem durante o processo de remontagem pelo cliente. Esta manipulação facilita significativamente a obtenção da chave WEP, tornando o ataque particularmente eficaz.

A técnica envolve o envio de pacotes que são intencionalmente fragmentados de maneira que, quando reagrupados pelo algoritmo de *reassemble* do dispositivo alvo, resultam em uma mensagem que parece ser de uma fonte confiável. Esta estratégia explora vulnerabilidades no protocolo WEP que não foi projetado para lidar de forma segura com pacotes fragmentados que podem ser manipulados durante a transmissão.

O ataque *Hirte* é uma demonstração da necessidade de protocolos mais robustos e seguros, como WPA2 ou WPA3, que oferecem mecanismos mais eficazes para proteger contra alterações no tráfego de rede.

5.1.4 Resumo dos Resultados

Os ataques *Cafe Latte* e *Evil Twin* são exemplos de ameaças cibernéticas em redes Wi-Fi (802.11), e cada um apresenta características distintas que podem ser identificadas e utilizadas em modelos de aprendizado de máquina para detecção de intrusões. Juntamente com estes, o tráfego de rede normal também é analisado para estabelecer uma linha de base de comportamento típico e inofensivo, crucial para diferenciar atividades legítimas de maliciosas.

5.1.4.1 Classe Normal

Para a classe normal, características como *frame.len*, *wlan.fc.subtype*, e *wlan.duration* são identificadas como altamente influentes. Altos valores das primeiras duas características citadas são consistentemente favoráveis para classificação do tráfego e aprendizado do modelo, no entanto, valores menores do *wlan.duration* também ajuda positivamente o modelo, indicando padrões de uso regular da rede com menores indícios de atividades mal-intencionadas.

5.1.4.2 Classe *Cafe Latte*

No ataque *Cafe Latte*, características como altos valores de *wlan.fc.subtype* , a ausência do *frame.len* e valores baixos de *data.len* são cruciais. Uma ampla distribuição nos valores SHAP destas características sugere um uso estratégico de diferentes subtipos de *frames* e comprimentos variados para efetuar a interceptação ou injeção de dados na rede.

5.1.4.3 Classe *Evil Twin*

No *Evil Twin*, valores altos de *data.len*, *frame.len* e*wlan.fc.subtype* mostram-se importantes para a avaliação do modelo de aprendizado de máquina, sendo essas características com valores mais extremados para a detecção de um ataque.

Capítulo 6

Conclusão

Este estudo focou na eficácia de diferentes conjuntos de características na detecção de ataques de personificação em redes Wi-Fi (802.11), com a utilização de técnicas avançadas de aprendizado de máquina. Entre as ferramentas empregadas, destaca-se o algoritmo XGBoost pela sua robustez e eficiência em modelagem preditiva e a biblioteca SHAP pela sua capacidade de proporcionar explicações detalhadas e intuitivas sobre a influência de cada característica no modelo. A escolha destas ferramentas deve-se não apenas à sua eficácia demonstrada em estudos anteriores, mas também à necessidade de entender profundamente o comportamento dos modelos em contextos de segurança.

Os resultados mostraram que o Conjunto 2 de características foi consistentemente superior em várias métricas, sendo considerado ideal para a detecção de ataques de personificação em ambientes de rede.

6.1 Contribuições do Estudo

Este trabalho ofereceu contribuições valiosas para a segurança em redes Wi-Fi (802.11):

- Demonstrou-se a eficácia do Conjunto 2 de características na detecção de ataques de personificação, reforçando a importância de uma seleção cuidadosa de características.
- Empregaram-se técnicas de XAI para proporcionar uma maior compreensão e precisão das características que indicam a presença de ataques, aumentando a transparência e confiança nos modelos de detecção.
- Evidenciou-se a relevância de métodos interpretáveis, que têm sido explorados em trabalhos anteriores, mas não com o foco e a profundidade proporcionados pela combinação de XGBoost e SHAP.
- Evidenciou-se como a combinação de IA e ML podem trazer resultado mais confiáveis e pontuais.

- ❏ Provou-se como as características escolhidas em cada conjunto possuem sua relevância no contexto de cada ataque que aconteceu.
- ❏ Evidenciou-se que para cada tipo de ataque, houveram características específicas e destacantes para a concepção do contexto de cada ataque ou ambiente normal de rede.

6.2 Limitações e Sugestões para Pesquisas Futuras

Este estudo possui várias limitações, que sugerem direções para futuras pesquisas:

- ❏ Aprofundamento na investigação de conjuntos de características e outros algoritmos que possam proporcionar resultados comparáveis ou superiores.
- ❏ Utilização da mesma metodologia para versões novas do *dataset* AWID.
- ❏ Aprofundamento das técnicas em outros padrões e protocolos de redes Wi-Fi
- ❏ Expansão do modelo para abranger um espectro mais amplo de tipos de ataques e ameaças em redes Wi-Fi (802.11).
- ❏ Integração de modelos de detecção com sistemas de resposta a incidentes para uma mitigação mais eficaz.
- ❏ Utilização do mesmos recursos voltados para outros ataques fora do escopo de redes Wi-Fi (802.11).

6.3 Contribuições em Produção Bibliográfica

Os resultados deste trabalho serão submetidos ao Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais (SBSeg) e estendidos para submissão ao *International Journal of Information Security* (IJIS), visando contribuir significativamente para a literatura em segurança cibernética.

Estas contribuições destacam a importância de continuar explorando e aprimorando as técnicas de detecção de intrusões em redes Wi-Fi (802.11), com um foco particular na explicabilidade e na eficácia dos modelos empregados.

Referências

AHMAD, M.; RAMACHANDRAN, V. Cafe latte with a free topping of cracked wep retrieving wep keys from road warriors. In: **Proc. Conf. ToorCon**. [S.l.: s.n.], 2007. Citado na página 19.

ALOTAIBI, B. **Empirical techniques to detect rogue wireless devices**. Tese (Doutorado) — University of Bridgeport, 2016. Citado 2 vezes nas páginas 33 e 34.

ALOTAIBI, B.; ELLEITHY, K. A majority voting technique for wireless intrusion detection systems. In: IEEE. **Systems, Applications and Technology Conference (LISAT), IEEE Long Island**. [S.l.], 2016. p. 1–6. Citado na página 33.

AMINANTO, M. E. et al. Deep abstraction and weighted feature selection for wi-fi impersonation detection. **IEEE Transactions on Information Forensics and Security**, IEEE, v. 13, n. 3, p. 621–636, 2018. Citado 4 vezes nas páginas 17, 19, 33 e 34.

AMINANTO, M. E.; KIM, K. Detecting impersonation attack in wifi networks using deep learning approach. **Cryptology and Information Security Lab, School of Computing, Korea Advanced Institute of Science and Technology (KAIST)**, Daejeon, Republic of Korea, 2016. Citado 4 vezes nas páginas 12, 15, 24 e 28.

AMINANTO, M. E.; KIM, K.-H. Weighted feature selection techniques for detecting impersonation attack in wi-fi networks. In: THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS. **2017 Symposium on Cryptography and Information Security (SCIS)**. [S.l.], 2017. Citado 7 vezes nas páginas 12, 15, 17, 24, 28, 33 e 34.

AMINANTO, M. E. et al. Weighted feature selection techniques for detecting impersonation attack in wi-fi networks. In: **Proc. Symp. Cryptogr. Inf. Secur.(SCIS)**. [S.l.: s.n.], 2016. p. 1–8. Citado 2 vezes nas páginas 33 e 34.

AMINANTO, M. E.; TANUWIDJAJA, H. C.; YOO, e. a. P. D. Wi-fi intrusion detection using weighted-feature selection for neural networks classifier. In: IEEE. **2017 International Workshop on Big Data and Information Security (IWBIS)**. [S.l.], 2017. p. 99–104. Citado 2 vezes nas páginas 31 e 33.

ANTONIOLI, D.; TIPPENHAUER, N. O.; RASMUSSEN, K. Bias: Bluetooth impersonation attacks. In: IEEE. **2020 IEEE symposium on security and privacy (SP)**. [S.l.], 2020. p. 549–562. Citado 2 vezes nas páginas 13 e 27.

BARBEAU, M.; HALL, J.; KRANAKIS, E. Detecting impersonation attacks in future wireless and mobile networks. In: **Secure Mobile Ad-hoc Networks and Sensors**. [S.l.]: Springer, 2006. p. 80–95. Citado na página 17.

BROECK, G. Van den et al. On the tractability of shap explanations. **Journal of Artificial Intelligence Research**, v. 74, p. 851–886, 2022. Citado na página 22.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. [S.l.: s.n.], 2016. Citado na página 22.

GOGA, O.; VENKATADRI, G.; GUMMADI, K. P. The doppelgänger bot attack: Exploring identity impersonation in online social networks. In: **Proceedings of the 2015 internet measurement conference**. [S.l.: s.n.], 2015. p. 141–153. Citado na página 26.

GOUVEIA, A.; CORREIA, M. Network intrusion detection with xgboost. **Recent Advances in Security, Privacy, and Trust for Internet of Things (IoT) and Cyber-Physical Systems (CPS)**, 2020. Citado na página 21.

KALEEM, D.; FERENS, K. A cognitive multi-agent model to detect malicious threats. Citado na página 33.

KALEEM D.; FERENS, K. A cognitive approach for attribute selection in internet dataset. In: IEEE. **IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC)**. [S.l.], 2017. p. 319–328. Citado na página 33.

KOLIAS, C. et al. Intrusion detection in 802.11 networks: empirical evaluation of threats and a public dataset. **IEEE Communications Surveys & Tutorials**, IEEE, v. 18, n. 1, p. 184–208, 2016. Citado 2 vezes nas páginas 17 e 31.

QUINCOZES, S. E.; KAZIENKO, J. F.; COPETTI, A. Avaliaçao de conjuntos de atributos para a detecçao de ataques de personificaçao na internet das coisas. In: SBC. **Anais Estendidos do VIII Simpósio Brasileiro de Engenharia de Sistemas Computacionais**. [S.l.], 2018. Citado 9 vezes nas páginas 17, 18, 19, 21, 31, 32, 33, 34 e 49.

TAMILSELVAN, L.; SANKARANARAYANAN, D. V. Prevention of impersonation attack in wireless mobile ad hoc networks. **International Journal of Computer Science and Network Security (IJCSNS)**, v. 7, n. 3, p. 118–123, 2007. Citado na página 25.

TU, S. et al. Reinforcement learning assisted impersonation attack detection in device-to-device communications. **IEEE Transactions on Vehicular Technology**, v. 70, n. 2, p. 1474–1479, 2021. Citado 2 vezes nas páginas 13 e 26.

_____. Security in fog computing: A novel technique to tackle an impersonation attack. **IEEE Access**, v. 6, p. 74993–75001, 2018. Citado 2 vezes nas páginas 24 e 25.

University of the Aegean. **AWID - A Wireless Intrusion Dataset**. 2024. <https://icsdweb.aegean.gr/awid>. Acessado em: data de acesso. Citado na página 35.

ZHU, J.; CAO, Z. Cryptanalysis of one authentication and key agreement scheme for internet of vehicles. Citado na página 24.

ZOVI, D. A. D.; MACAULAY, S. A. Attacking automatic wireless network selection. In: IEEE. **Proceedings from the Sixth Annual IEEE SMC. Information Assurance Workshop. IAW'05.** [S.l.], 2005. p. 365–372. Citado na página 18.