UMA ANÁLISE COMPARATIVA ENTRE ABORDAGENS DEAPRENDIZADO DE MÁQUINA PARA DETECÇÃO DE AMEAÇAS CIBERNÉTICAS NA SEGURANÇA DA IOT

A COMPARATIVE ANALYSIS OF MACHINE LEARNING APPROACHES FOR CYBER THREAT DETECTION IN IOT SECURITY

Daniella Carolina Camargo Torelli[1]
Ricardo Petri Silva[2]

RESUMO

À medida que a *Internet* se torna essencial em nossas vidas, os ataques cibernéticos proliferam. Este artigo realiza uma análise comparativa entre duas pesquisas que utilizam técnicas de aprendizado de máquina para a detecção de ameaças cibernéticas. Exploramos a relevância da cibersegurança, abordando a crescente ameaça de ataques, incluindo exemplos como *phishing* e *ransomware*. Discutimos as aplicações do aprendizado de máquina na detecção de ameaças e revisamos conjuntos de dados críticos. Em seguida, propomos uma metodologia de avaliação com base na eficiência na detecção, tempo requerido para detecção e qualidade das bases de dados. Concluímos que ambos os artigos complementam-se e contribuem para uma compreensão abrangente da detecção de ameaças na In- ternet das Coisas (IoT). Este estudo fornece *insights* valiosos para pesquisadores e profissionais de segurança cibernética.

Palavras-chave: cibersegurança; aprendizado de máquina; detecção de ameaças; internet das coisas; análise comparativa.

ABSTRACT

As the Internet becomes essential in our lives, cyberattacks proliferate. This article conducts a comparative analysis between two research papers that employ machine learning techniques for cyber threat detection. We explore the relevance of cybersecurity, addressing the growing threat of attacks, includ- ing examples like phishing and ransomware. We discuss the applications of machine learning in threat detection and review critical datasets. Next, we propose an evaluation methodology based on detection efficiency, required time for detection, and database quality. We conclude that both papers complement each other and contribute to a comprehensive understanding of threat detection in the Internet of Things (IoT). This study provides valuable insights for researchers and cybersecurity professionals.

Keywords: cybersecurity; machine learning; threat detection; internet of things; comparative analy-sis.

[1] Centro Universitário Filadélfia de Londrina - UniFil
[2] Centro Universitário Filadélfia de Londrina - UniFil

1 INTRODUÇÃO

A segurança na *internet* tem sido objeto de extensa análise por parte de diversos pesquisadores. Com o crescente aumento no uso da *internet*, tanto para fins de lazer quanto profissionais, os ataques cibernéticos têm aumentado. Em 2022, de acordo com a Research (2022), empresa multinacional de cibersegurança *Trend Mi- cro*, foram bloqueadas incríveis 146.408.535.569 ameaças de variados tipos.

Conforme ressaltado por Gupta *et al*. (2017), um dos golpes que tem ganhado notoriedade progressiva e se tornado cada vez mais comum nos ataques cibernéticos é o *phishing*. Nesses incidentes, invasores empregam diversas abordagens para se apropriar de dados pessoais e/ou financeiros de terceiros, frequentemente por meio do uso de *malwareS*, como por exemplo os trojans, phishing e etc.

Sarker *et al*. (2020) aponta para um fenômeno preocupante: com a crescente migração para o mundo virtual, tem-se observado uma escalada de incidentes de segurança, que têm crescido exponencialmente nos últimos anos. Como destacado por Sun *et al*. (2019), diante do panorama marcado por uma variedade de ameaças e ataques na internet, a segurança e a integridade dos dados que circulam nesse ambiente tornaram-se indispensáveis. Como resultado, pesquisadores e empresas iniciaram investigações e propuseram esquemas de previsão e possíveis medidas de contenção, recorrendo a diversas fontes de dados, relatórios divulgados por organizações, informações presentes em redes, dados sintéticos, rastreamento de páginas da web e conteúdo de mídias sociais.

Como ressaltado por Dua e Du (2011), uma variedade de pesquisas tem sido conduzidas no campo da detecção de ataques, e uma das áreas que se destaca é uma vertente da Inteligência Artificial (IA) conhecida como Aprendizado de Máquina (AM). Por meio da análise de dados provenientes de bases de dados, é possível desenvolver algoritmos capazes de aprender com essas informações.

De acordo com Géron (2017), o AM é a ciência de programar computadores para que eles possam aprender com os dados. Seguindo a mesma ideia, Ceschin, Oliveira e Grégio (2019) apresentam um exemplo relacionado aos ataques cibernéticos, no qual *malware* (programas maliciosos) e *goodware* (programas benignos) constituem duas categorias de um ataque cibernético do tipo *malware*. Embora essas duas

categorias compartilhem semelhanças, suas distinções se baseiam nas ações executadas. O *malware* age com o intuito de prejudicar o sistema, enquanto o *goodware* realiza ações que não causam danos. Com base nesses dados coletados, algoritmos podem ser treinados para identificar a natureza de um determinado *software* (*malware* ou *goodware*).

Neste artigo, foi investigado diferentes técnicas de AM supervisionado e desenvolvido uma metodologia de classificação para cada uma delas. Nosso objetivo foi estabelecer um ranking dessas técnicas com base em três critérios relacionados à detecção de ameaças, tais como o tempo requerido para a detecção e treinamento do algoritmo, eficiência na detecção de ameaças, e a base de dados e cenário de avaliação.

Após a conclusão da pesquisa foi alcançado um resultado satisfatório, onde, apesar das diferenças nas pontuações em duas categorias, a pontuação total para ambos os artigos permanece igual. Isso sugere que os artigos possuem características complementares e são igualmente relevantes em nosso contexto de pesquisa em cibersegurança.

Nas seções seguintes, exploraremos os trabalhos relacionados ao nosso tópico de pesquisa, analisaremos a fundamentação teórica, onde discutiremos os conceitos relevantes para este artigo. Apresentaremos nossa proposta de estudo, detalharemos a metodologia empregada em nossa pesquisa, destacaremos os resultados obtidos e, por fim, concluiremos este estudo.

2 TRABALHOS RELACIONADOS

Esta seção traça um panorama abrangente das pesquisas recentes e relevantes que se relacionam diretamente com o tema abordado neste artigo. A cibersegurança tem sido um campo de estudo em constante expansão, e a literatura acadêmica reflete essa evolução com um conjunto crescente de contribuições significativas. Ao explorar as referências selecionadas, examinamos estudos que se debruçam sobre os desafios emergentes e as estratégias inovadoras utilizadas para fortalecer a segurança digital. Estes trabalhos fornecem *insights* valiosos que complementam nossa abordagem e aprimoram nossa compreensão do estado atual da cibersegurança, bem

como das oportunidades para futuras pesquisas.

2.1 Cibersegurança

Compreender o contexto e a importância da cibersegurança é essencial para a avaliação das técnicas de aprendizado de máquina aplicadas à detecção de ataques cibernéticos, conforme delineado neste estudo. Como destacado por Steinberg (2020), a cibersegurança é uma disciplina crítica que lida com a proteção de informações e sistemas eletrônicos, sendo um componente fundamental da segurança da informação. A preservação da integridade, confidencialidade e disponibilidade dos dados armazenados e processados eletronicamente é de suma importância, e a crescente ameaçade ataques cibernéticos requer estratégias eficazes de defesa.

Conforme ressaltado por von Solms e Van Niekerk (2013), a cibersegurança abrange uma ampla gama de elementos, desde políticas e conceitos de segurança até tecnologias, abordagens de gerenciamento de riscos, ações, treinamentos e melhores práticas. É crucial que as organizações estejam preparadas para proteger seu ciberespaço, infraestrutura, serviços de telecomunicações e, por extensão, as informações relevantes que residem nesses sistemas. Como observado por Pande (2017), a capacidade de resistir a ataques cibernéticos é um desafio atual para as organizações, o que torna a escolha das técnicas de aprendizado de máquina apropriadas ainda mais vital.

Portanto, ao avaliar as técnicas de AM no contexto da detecção de ataques cibernéticos, é fundamental considerar como essas abordagens contribuem para a segurança cibernética e para a proteção dos sistemas eletrônicos que são alvos potenciais de ameaças.

2.2 Malwares e Algumas de suas Categorias

Entender a natureza e a diversidade de *malwares*, conforme descrito por fontes como Kaspersky (2023a) e Oliveira (2018), é de extrema importância para o escopo deste estudo sobre a aplicação de técnicas de aprendizado de máquina na detecção de ataques cibernéticos. *Malwares* representam uma das principais ameaças

ao ciberespaço e às organizações, sendo responsáveis por uma variedade de ações maliciosas, desde a espionagem até a interrupção de sistemas.

Kaspersky (2023a) destaca a versatilidade dos *malwares*, que podem assumir várias formas, como vírus, cavalos de Troia e *ransomware*, e infectar computadores de diversas maneiras. O entendimento da natureza dessas ameaças é crucial para avaliar a eficácia das técnicas de AM na identificação desses ataques, uma vez que diferentes tipos de *malwares* podem exigir abordagens de detecção distintas.

Oliveira (2018) classifica *malwares* em categorias como ocultação, incluindo os notórios Trojans, e aqueles que visam a infectação, como os vírus, bem como aqueles que buscam benefícios próprios, como os *ransomwares*. Essas categorias refletem a complexidade das ameaças cibernéticas e ressaltam a importância de adotar abordagens de detecção capazes de lidar com essa diversidade.

Portanto, o conhecimento sobre a natureza dos *malwares* e suas categorias é fundamental para a análise das técnicas de AM na detecção de ataques cibernéticos. Isso demonstrará como a compreensão das ameaças subjacentes é relevante para a seleção e avaliação das técnicas de detecção de ataques cibernéticos abordadas neste estudo.

2.2.1 Trojans

Ao analisar a categoria de *malware* conhecida como *Trojans*, conforme abordada por Steinberg (2020), percebemos que eles são *malwares* disfarçados de *softwares* legítimos, muitas vezes adotando o nome "Cavalo de Tróia"devido à sua capacidade de se esconder em aplicações aparentemente inofensivas. Uma característica crucial dos *Trojans* é a necessidade de engenharia social para se propagar, o que frequentemente envolve enganar os usuários a clicar em links, instalar aplicativos ou abrir anexos de *e-mail* maliciosos.

Este entendimento da forma como os *Trojans* operam é fundamental para a análise das técnicas de AM utilizadas na detecção de ataques cibernéticos. A detecção eficaz de *Trojans* requer a identificação de comportamentos suspeitos e a análise de atividades que possam não ser facilmente detectadas por métodos convencionaisde segurança.

2.2.2 Ransomware

A análise da categoria de *malware* conhecida como *ransomware* é essencial para compreender a gravidade das ameaças cibernéticas enfrentadas por organizações e usuários, conforme ilustrado por Abraham e George (2019) e Eoco (2023). O *ransomware* tem se destacado como uma das formas mais prejudiciais de ataques cibernéticos, com alvos que variam desde pequenas empresas até grandes corporações. É interessante notar que, embora os ataques de *ransomware* tenham ganhado destaque nos últimos anos, eles não são uma novidade, tendo suas origens em 1989, como aponta Eoco (2023).

Steinberg (2020) fornece informações valiosas sobre o funcionamento do *ransomware*, destacando que esse tipo de *malware* criptografa os arquivos dos usuários, impedindo o acesso a menos que um resgate seja pago aos criminosos. A ameaça de excluir a chave de criptografia ou expor informações confidenciais adiciona um elemento de urgência a esses ataques.

O exemplo do ataque *WannaCry*, conforme descrito por Kaspersky (2023a), é uma demonstração de quão devastador um ataque de *ransomware* pode ser. A epidemia global de *ransomware* em 2017 afetou cerca de 230 mil computadores em todo o mundo e se espalhou por computadores com o *Microsoft Windows*. Os arquivos dos usuários eram mantidos como reféns e, para que fossem devolvidos, era exigido um resgate em *bitcoins*, destacando a necessidade crítica de identificar e mitigar esse tipo de ameaça.

2.2.3 Vírus

Ao abordar a categoria de *malware* denominada vírus, de acordo com as informações apresentadas por Gupta et al. (2017) e Steinberg (2020), torna-se evidente que essa classe de ameaças cibernéticas tem suas características distintivas. Os vírus são programas de computador que se propagam inserindo cópias de si mesmos em outros programas e arquivos, tornando-se uma parte integrante deles. Essa propagação requer a execução do programa ou arquivo hospedeiro, que serve como veículo para o vírus.

Como Steinberg (2020) destaca, os vírus geralmente residem em arquivos de dados, na parte especial de discos rígidos ou em unidades de estado sólido que contêm código e dados usados para inicializar um computador ou disco. Essa característica de esconder-se em locais específicos é crucial para a compreensão da detecção de vírus, pois implica que as técnicas de AM precisam identificar essas assinaturas e comportamentos específicos associados aos vírus.

2.2.4 Phishing

A análise do *phishing*, conforme descrito por Steinberg (2020), é fundamental para compreender as ameaças cibernéticas que visam enganar os usuários e obter informações pessoais e financeiras. O *phishing* representa uma ameaça persistente, na qual golpistas tentam se passar por entidades legítimas para induzir os usuários a revelar dados confidenciais.

Steinberg (2020) fornece um exemplo típico de como os ataques de *phishing* ocorrem, com um criminoso enviando *e-mails* fraudulentos que aparentam ser de instituições legítimas, como bancos. Esses *e-mails* frequentemente solicitam que o destinatário clique em *links* e forneça informações confidenciais, como senhas e dados pessoais e de contas bancárias. O entendimento de como esses ataques operam é crucial para a identificação e prevenção eficaz de ameaças de *phishing*.

2.3 Aprendizado de Máquina

De acordo com o autor Géron (2021), o AM é uma disciplina que combina a ciência e a arte da programação de computadores, permitindo que eles adquiram a capacidade de aprender com dados. No âmbito do AM, vários modelos se destacam, sendo os mais comuns o Aprendizado Supervisionado, como apontado pelo autor. Nesse modelo, os algoritmos utilizam conjuntos de dados previamente classificados para treinamento. Um exemplo notável é o conjunto de dados *KDDCUP'99*, empregado na Terceira Competição Internacional de Ferramentas de Descoberta de Conhecimento e Mineração de Dados em conjunto com o evento *KDD-99 (Knowledge Discovery in Databases)*.

Matheus, Raphaell e Calanca (2023) destacam outro modelo amplamente empregado, o Aprendizado Não Supervisionado, no qual os algoritmos recebem conjuntos de dados sem nenhuma classificação e buscam compreender as relações entre os dados, agrupando-os e realizando previsões com base nessa compreensão. Além desses dois modelos predominantes, existe o Aprendizado Semi-Supervisionado, uma combinação dos modelos mencionados anteriormente, em que uma parte do conjunto de dados é classificada e a outra não, permitindo economia de custos e otimização de tempo.

Por último, de acordo com Matheus, Raphaell e Calanca (2023), o Aprendizado por Reforço é o mais distinto de todos, caracterizado por iniciar os testes aleatori- amente e, posteriormente, ajustar o algoritmo com base nas "recompensas"recebidas em função dos acertos e erros. Dessa forma, o algoritmo aprimora suas capacidades ao longo dos testes para atingir a melhor solução possível.

É fundamental compreender os princípios e modelos do AM, conforme abordado nesta seção, uma vez que esse conhecimento constitui a base para o entendimento e a aplicação das técnicas de AM na detecção de ataques cibernéticos, objetivo central deste estudo. O AM desempenha um papel crucial na capacidade de identificar ameaças cibernéticas em constante evolução, oferecendo métodos eficazes para analisar, classificar e responder a possíveis ataques. A compreensão dos diferentes modelos, como o Aprendizado Supervisionado, Não Supervisionado, Semi-Supervisionado e por Reforço, fornece as ferramentas necessárias para selecionar as abordagens mais adequadas à detecção de ataques cibernéticos, levando em consideração as nuances e as complexidades das ameaças. Portanto, a familiaridade com os conceitos e modelos de AM é crucial para a pesquisa em cibersegurança, pois permite a criação de estratégias eficazes de defesa e aprimoramento contínuo das técnicas de detecção de ataques.

2.4 Datasets sobre Ataques Cibernéticos

Compreender os Sistemas de Detecção de Intrusão (IDS), como mencionado por Özgür e Erdem (2016), é de extrema importância para a pesquisa de detecção e

mitigação de ataques cibernéticos. Os *IDS* desempenham um papel crítico na identificação de atividades suspeitas ou maliciosas em sistemas de computador e redes, ajudando a proteger organizações e indivíduos contra ameaças cibernéticas em constante evolução. Como destacado por Obeidat *et al*. (2019), os *IDS* são divididos em diferentes categorias e técnicas, cada uma adequada para lidar com tipos específicos de ameaças:

> O *IDS* combina hardware e software para detectar ataques em redes, a fim de garantir a proteção do sistema contra acessos não autorizados. O *IDS* pode ser dividido em duas classificações principais com base no método de detecção do ataque. O primeiro é o uso indevido e o segundo é a detecção de anomalias. A detecção de anomalias pode ser utilizada de diversas formas para detectar qualquer comportamento estranho do usuário dentro do tráfego da rede (OBEIDAT *et al*., 2019).

2.4.1 DARPA

O entendimento do papel desempenhado pela *DARPA (Defense Advanced Research Projects Agency)*, conforme apontado por Squeff e Negri (2017), é fundamental para a compreensão do desenvolvimento e pesquisa em tecnologias de defesa. A *DARPA* é a agência do Departamento de Defesa dos Estados Unidos responsável por realizar investimentos iniciais cruciais no desenvolvimento de tecnologias de defesa, desempenhando um papel vital na garantia da segurança nacional.

Como ilustrado por Silva (2020), a colaboração da *DARPA* com o Laboratório *Lincoln Labs*, afiliado ao *MIT*, resultou na criação de Sistemas de Detecção de Intrusão *(IDSs)* direcionados à segurança de redes. Esse esforço conjunto visava proteger redes contra ameaças cibernéticas em constante evolução, sendo exemplificado por um conjunto de dados originado da simulação de uma rede militar em uma base aérea norte-americana. Esse ambiente foi monitorado durante 7 semanas e sofreu uma série de ataques cibernéticos. Os dados coletados ao longo desse período foram posteriormente consolidados em um único conjunto de dados, originando o renomado conjunto de dados *DARPA*.

2.4.2 KDD99

O estudo conduzido por Obeidat et al. (2019) lança luz sobre uma das bases de dados mais renomadas e talvez uma das mais antigas no campo da cibersegurança, a *Knowledge Discovery in Databases* (KDD). Esta base de dados é um repositório online que abrange uma ampla variedade de tipos de ataques cibernéticos, englobando desde ataques de negação de serviço (DOS) até ataques de R2L *(Remote to Local)*, U2R *(User to Root)* e *PROBE* (sondagem).

Conforme destacado por Silva (2020) e reforçado por Obeidat et al. (2019), em 1998, o conjunto de dados DARPA passou por modificações substanciais para torná-lo mais adequado para competições de extração de conhecimento e mineração de dados. Essas adaptações deram origem à KDD99, que se tornou uma valiosa fonte de dados para aplicações de aprendizado de máquina.

A compreensão da evolução do conjunto de dados DARPA para a KDD99 e seu uso generalizado em pesquisa e competições de aprendizado de máquina é crucial. Isso ocorre porque a KDD99 representa um recurso valioso para avaliar e aprimorar técnicas de detecção de ataques cibernéticos, com sua variedade de tipos de ataques e cenários, além de mostrar um início das pesquisas em cibersegurança, propiciando futuramente a criação de *datasets* atuais e concisos.

2.4.3 NSL-KDD

Como apontado por Silva (2020) e reiterado por Obeidat *et al*. (2019), a KDD99, embora mais acessível em comparação com o conjunto de dados DARPA, apresentava algumas deficiências significativas. Entre essas limitações estavam a presença de instâncias redundantes e duplicadas, bem como um tamanho considerável que podia dificultar a criação de Sistemas de Detecção de Intrusões (IDS), resultando em resultados insatisfatórios ou distorcidos.

Para mitigar os efeitos indesejados decorrentes das falhas presentes na KDD99, foi desenvolvido o conjunto de dados NSL-KDD. Esse novo conjunto de dados, criado por meio da exclusão dos registros problemáticos, é de tamanho reduzido em compa- ração com a KDD99. A criação do NSL-KDD visou fornecer

um recurso mais equili- brado e de alta qualidade para o desenvolvimento e avaliação de IDS, eliminando as instâncias redundantes e duplicadas que poderiam afetar negativamente a precisão e a eficácia das técnicas de detecção de intrusões.

2.4.4 STRATOSPHERE IPS

O grupo de segurança cibernética sediado no Centro de Inteligência Artificial da Faculdade de Engenharia Elétrica da Universidade Técnica Tcheca em Praga, conhecido como Stratosphere (2015), tem se destacado nas pesquisas que abordam a segurança cibernética e o AM. Sua atuação abrange uma ampla gama de projetos, englobando desde o desenvolvimento de Sistemas de Detecção de Intrusão (IDS) baseados em AM até investigações detalhadas sobre propaganda computacional.

Como ressaltado por Stratosphere (2021), uma característica excepcional das pesquisas conduzidas por esse grupo é a utilização de dados reais de tráfego de *malware*. Essa abordagem prática desempenha um papel crucial na garantia da qualidade e eficácia dos modelos desenvolvidos no contexto da segurança cibernética. Ao longo dos anos, o grupo tem desempenhado um papel de liderança ao disponibilizar para a comunidade de pesquisa um impressionante acervo de mais de 300 conjuntos de dados de tráfego de *malware* de longo prazo.

> O *Stratosphere IPS* se alimenta de modelos criados a partir de capturas reais de tráfego de *malware*. Ao usar e estudar como o *malware* se comporta na realidade, garantimos que os modelos que criamos sejam precisos e que nossas medições de desempenho sejam reais. Nosso projeto irmão, *Malware Capture Facility Project*, é responsável por monitorar continuamente o cenário de ameaças em busca de novas ameaças emergentes, recuperando amostras maliciosas e executando-as em nossas instalações para capturar o tráfego (STRATOSPHERE, 2015).

A Tabela 1 fornece um resumo abrangente dos tipos de captura de *malware* utilizados pelo *Stratosphere Lab*. Esses métodos de captura desempenham papéis específicos na obtenção de dados para alimentar os conjuntos de dados de tráfego de *malware*.

Tabela 1 – Tipos de Captura de *Malware* Utilizados pela *Stratosphere Lab*

CAPTURAS DE MALWARE	CAPTURAS NORMAIS	CAPTURAS MISTAS
Faz capturas de malware de longo prazo.	É fundamental para calcular com precisão os valores verdadeiros de Falsos Positivos e Verdadeiros Negativos.	Fornecem um cenário real onde uma máquina não está infectada, depois é infectada e depois de algum tempo a infecção é eliminada.
Obtém continuamente malware e dados normais para alimentar os datasets.	-	Facilita o teste dos algoritmos e modelos de aprendizado de máquina StratosphereIPS.

Fonte: Stratosphere (2015)

A atualização trazida por Stratosphere (2021) em 2021 é de significativa importância para a comunidade de pesquisa em segurança cibernética. O novo índice de conjunto de dados apresentado facilita consideravelmente a localização e o acesso a uma variedade de conjuntos de dados valiosos. Esses conjuntos de dados estão categorizados de maneira a permitir a busca por tipos de *malware* específicos, datas de infecção e outros critérios relevantes.

Neste contexto, essa pesquisa se concentra na utilização dessas bases de dados desenvolvidas pelo grupo Stratosphere (2015). Esses conjuntos de dados de tráfego de *malware* representam uma fonte valiosa para avaliar e aprimorar técnicas de detecção de ataques cibernéticos com base em aprendizado de máquina. Portanto, os artigos escolhidos para estudo nesse artigo tem como padrão a utilização desse *dataset* disponibilizado pelo grupo Stratosphere (2015) e o acesso a esses conjuntos

de dados realistas e desafiadores são essenciais para nossas avaliações.

3 FUNDAMENTAÇÃO TEÓRICA

A seção de fundamentação teórica deste artigo desempenha um papel crucial ao proporcionar uma base sólida para a compreensão dos tópicos relevantes à pesquisa. É aqui que exploraremos as definições e conceitos essenciais relacionados a esse campo dinâmico e em constante evolução, a cibersegurança. A principal intenção desta seção é estabelecer um alicerce sólido que servirá de fundamento para as discussões subsequentes e a apresentação dos resultados. Nela, apresentamos os princípios e terminologias fundamentais para a compreensão completa do conteúdo deste artigo.

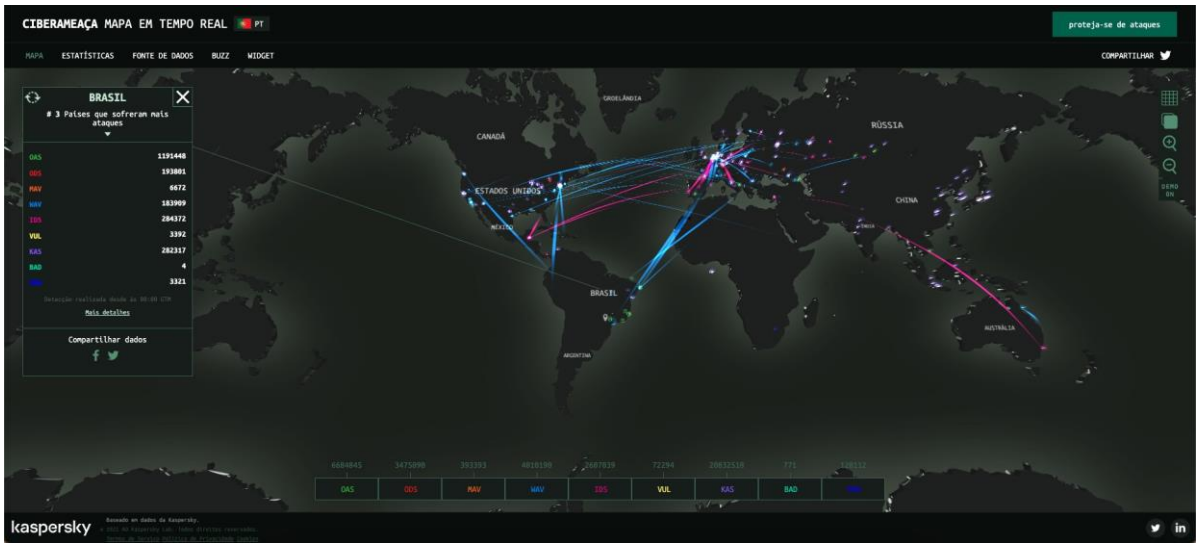
3.1 Cibersegurança

Em sua obra intitulada ”Cibersegurança Para Leigos: Os Primeiros Passos Para o Sucesso!” Steinberg (2020), enfatiza que, embora o termo ”cibersegurança” possa ser aparentemente simples de conceituar, sua interpretação e aplicação variam significativamente de acordo com o contexto. Aponta ainda a multifacetada natureza dessa disciplina, que se reflete em políticas, práticas e procedimentos diversificados. Por meio de exemplos, o autor ilustra essa diversidade de interpretações; Para pequenos empresários, a cibersegurança pode englobar a proteção de informações sensíveis, como dados de cartões de crédito, juntamente com a correta implementação de padrões de segurança de dados nos registros de pontos de venda. Para empresas que operam no cenário digital, o conceito de cibersegurança inclui a proteção de servidores que mantêm interações regulares com terceiros não confiáveis. Provedores de serviços compartilhados enfrentam o desafio de proteger inúmeros data centers que abrigam inúmeros servidores virtuais, cada qual pertencente a diferentes entidades. Para o setor governamental, o espectro da cibersegurança pode abranger a categorização de dados em diferentes níveis, cada um deles com seu próprio conjunto de legislações, políticas, procedimentos e tecnologias correlatas.

A empresa de segurança cibernética Kaspersky (2023b) disponibiliza um Mapa

em Tempo Real de Ameaças Cibernéticas que coleta informações sobre ataques de todo o mundo por meio de seus produtos de segurança globalmente distribuídos e pode ser acessado através do link https://cybermap.kaspersky.com/pt . A Figura 1 abaixo ilustra o funcionamento dessa ferramenta, exibindo a atual paisagem de ataques cibernéticos em tempo real.

Figura 1 – Mapa em Tempo Real de Ameaças Cibernéticas



Fonte: Captura da tela dos autores no site (KASPERSKY, 2023b). Acesso em: 24 out. 2023.

3.2 *Malware*

Com base no conteúdo apresentado por Melo (2023) em seu livro ”Análise de *Malwares*,” o termo *"Malware"* engloba uma ampla variedade de programas meticulosamente concebidos para executar ações prejudiciais em sistemas computacionais. No contexto da internet, observamos uma extensa diversidade de tipos de *malwares*, incluindo vírus, cavalos de Troia (trojans), *ransomware, phishing* e outras variantes bem reconhecidas.

Para a empresa de segurança digital McAfee (2023), *malware* é um termo genérico para qualquer tipo de *software* malicioso projetado para prejudicar ou explorar qualquer dispositivo, serviço ou rede programável. Os criminosos cibernéticos costumam usá-lo para extrair dados que podem ser utilizados das vítimas para obter ganhos financeiros, para que forneça dados pessoais para roubo de identidade, para assumir

controle de múltiplos computadores para lançar ataques de negação de serviço contra outras redes, para infectar computadores e usá -los para minar *bitcoin* ou outros moedas virtuais, etc. Eles vão de dados financeiros, registros médicos a *e-mails* e senhas pessoais, as possibilidades de que tipo de informação pode ser comprometidase tornaram infinitas.

3.3 Aprendizado de Máquina

Conforme apontado por Mitchell (1997), o campo do aprendizado de máquina concentra-se na elaboração de programas de computador capazes de aprimorar suas capacidades automaticamente por meio da experiência. Essa melhoria é alcançada por meio de treinamento com conjuntos de dados, fazendo com que os programas aprimorem suas habilidades, culminando em previsões mais precisas com base nas informações disponíveis. Em sua maioria, esses conjuntos de dados consistem em informações do mundo real.

No livro ”Mãos à Obra: Aprendizado de Máquina com *Scikit-Learn, Keras e TensorFlow* : Conceitos, Ferramentas e Técnicas para a Construção de Sistemas Inteligentes” de Géron (2021), o AM é definido como a disciplina que combina a ciência e a arte da programação de computadores, permitindo que eles adquiram a capacidade de aprender com dados. O autor explora aplicações do AM, destacando sua eficácia em lidar com problemas nos quais as soluções convencionais demandam ajustes minuciosos, extensas listas de regras ou a compreensão de problemas complexos que envolvem grandes volumes de dados. Géron (2021) ainda cita que o AM encontra aplicações em diversos domínios, como a detecção de tumores a partir de imagens de exames cerebrais, a identificação de fraudes em transações de cartão de crédito e a classificação automática de artigos de notícias, entre outros cenários desafiadores.

4 METODOLOGIA

A finalidade deste artigo é realizar um estudo de caso comparativo envolvendo dois artigos relacionados à detecção de ataques cibernéticos por meio de técnicas de aprendizado de máquina. O objetivo principal é desenvolver três critérios de avaliação

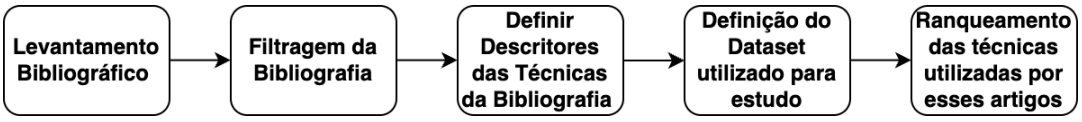
que permita determinar qual dos dois estudos foi mais eficaz em termos de detecção e precisão.

Para preparar este estudo, conduzimos uma pesquisa bibliográfica minuciosa sobre cibersegurança e AM. Nosso objetivo principal era realizar uma análise comparativa entre dois artigos que aplicavam técnicas de AM supervisionado na detecção eprevenção de ataques cibernéticos.

Entretanto, encontramos desafios significativos durante o processo de escrita. A dificuldade inicial surgiu na busca por artigos que se concentrassem exclusivamente em AM supervisionado. Isso levantou questões sobre a compatibilidade dos conjuntos de dados utilizados, uma vez que todos os artigos examinados deveriam usaro mesmo conjunto de dados para garantir a imparcialidade em nossa análise. Além disso, descobrimos que o conjunto de dados inicialmente selecionado estava desatualizado e talvez não atendesse às necessidades de nossa pesquisa.

Diante deste impasse, foi adotada uma nova abordagem. Inicialmente, foram identificados conjuntos de dados confiáveis e atualizados. Em seguida, os esforços se concentraram na busca de artigos que fizessem uso desses *datasets*, independentemente do tipo de AM empregado. Após uma seleção criteriosa, dois artigos foram escolhidos como apropriados para prosseguir com a análise comparativa. Com base nesse ponto de partida, o escopo e os objetivos iniciais foram adaptados, culminando na redação do restante do artigo, que incluiu a análise e a elaboração de um *ranking* que ilustra a eficácia dos métodos empregados. A Figura 2 representa um diagrama resumido da metodologia empregada na construção deste artigo.

Figura 2 – Metodologia Utilizada para a Construção desse Artigo



Fonte: Autora

A importância dessa metodologia reside não apenas na superação dos desafios iniciais, mas também na capacidade de refinar o trabalho, proporcionando ideias valiosas e resultados significativos. Assim, a abordagem adotada desempenhou um papel crucial no alcance dos objetivos estabelecidos para este

estudo.

5 RESULTADOS E DISCUSSÕES

Em busca de prosseguir com o propósito deste artigo, realizamos uma análise minuciosa de dois artigos relacionados à cibersegurança, ambos fazendo uso do mesmo conjunto de dados fornecido pelo laboratório *Stratosphere, o IoT-23*. A seguir, abordaremos em detalhes essas pesquisas e seus resultados, com o intuito de oferecer uma compreensão mais profunda e comparativa dessas contribuições para o campo da segurança na Internet das Coisas (IoT).

5.1 Artigo 1: Predicting Malicious Software in IoT Environment Ba- sed on Machine Learning and Data Mining Techniques

Alharbi, Hamid e Lahza (2022) abordam em seu artigo a crescente ameaça cibernética no contexto da Internet das Coisas (IoT) e como a utilização de técnicas de AM pode ser aplicada para detectar e prever ataques maliciosos nessa área. A *IoT* refere-se a dispositivos conectados à internet que podem transferir dados por meio de uma rede. No entanto, esses dispositivos estão vulneráveis a ataques cibernéticos, como *DDoS, ransomware* e ataques de *botnet IoT*, que visam desativar sistemas e redes.

Os autores mencionam pesquisas anteriores que usaram modelos de aprendizado de máquina para avaliar o desempenho na detecção de ataques cibernéticos em dispositivos *IoT*. Essas pesquisas utilizaram conjuntos de dados específicos, como o *IoT-23*, que continha informações rotuladas sobre tráfego benigno e malicioso. Várias técnicas de classificação foram testadas, incluindo *Support Vector Machine, Random Forest, Naive Bayes*, Regressão Logística e Árvore de Decisão. Os resultados dessas pesquisas indicaram que o algoritmo *Random Forest* alcançou a maior precisão na detecção de *malware*, com uma precisão média ponderada de 100%. Outros algoritmos também foram eficazes, embora com taxas de precisão variáveis. Além disso, alguns estudos se concentraram na detecção de ameaças, enquanto este artigo ressalta a importância de prever ataques maliciosos para reduzir as vulnerabilidades dos

dispositivos *IoT*.

Apontam ainda os diferentes tipos de ameaças cibernéticas, como *malware*, ataques de *ransomware*, ataques de negação de serviço distribuído *(DDoS)*, espionagem cibernética e ataques de *botnet IoT*. É destacado que os ataques de *botnet IoT* representam uma parte significativa das ameaças cibernéticas, causando danos financeiros e prejudicando a integridade dos dispositivos *IoT*.

Como proposta de estudo, os autores conduziram uma pesquisa com o objetivo de prever o tráfego de rede na Internet das Coisas (IoT), com foco na detecção detráfego malicioso e benigno. Para isso, foram utilizadas técnicas de AM e Mineração de Dados. Como dito anteriormente, o estudo se baseou em um conjunto de dados chamado *IoT-23*, pertencente ao *Stratosphere Lab*, que continha tráfego de rede rotu- lado como malicioso e benigno, coletado ao longo de seis meses. A metodologia do estudo envolveu várias etapas essenciais.

Primeiramente, Alharbi, Hamid e Lahza (2022) coletaram os dados de tráfego de rede da *IoT-23* e realizaram um processo de pré-processamento para remover dados irrelevantes e normalizar as informações, tornando os dados adequados para análise. Em seguida, os dados foram divididos em dois conjuntos: um conjunto de treinamento e um conjunto de teste. O conjunto de treinamento foi utilizado para treinar os modelos de AM, enquanto o conjunto de teste serviu para avaliar o desempenho desses modelos. Uma parte crucial desse estudo foi a seleção de recursos. Para isso, o autor aplicou a Análise de Componentes Principais (PCA) no conjunto de dados *IoT-23*. Isso permitiu reduzir a dimensionalidade do conjunto de dados, mantendo os recursos mais relevantes e eliminando redundâncias. O método PCA contribuiu para melhorar o desempenho dos algoritmos de classificação, evitando o *overfitting*.

O treinamento e o teste dos modelos de AM foram realizados com o auxílio de ferramentas como o *Weka*, que fornece uma coleção de algoritmos de aprendizado de máquina, e o *Orange*, que é uma plataforma de código aberto para análise de dados e mineração de dados. Os resultados do estudo demonstraram que o classificador *Random Forest (RF)* obteve o melhor desempenho, com uma precisão de classificação de 97,14% e uma Área sob a Curva *(AUC)* de 96,44%. Outros algoritmos, como *Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM) e Naive Bayes (NB)*, também apresentaram resultados satisfatórios, embora o *SVM* tenha tido um

desempenho ligeiramente inferior.

5.2 Artigo 2: Data Science in Cybersecurity: Evaluating the Use of Machine Learning in an IOT-IDS

Em seu artigo, Prazeres (2022) apresenta uma solução de cibersegurança voltada para ambientes de Internet das Coisas (IoT), especialmente em cenários de cidades inteligentes. Para isso, o autor realiza uma análise das quatro camadas que compõem a arquitetura de uma cidade inteligente. Começando pela camada de percepção, onde os sensores desempenham um papel fundamental, o autor propõe a segmentação do tráfego de rede da *IoT*. Essa segmentação é baseada em protocolos de mensagens de aplicação, como *MQTT ou CoAP*, ou na categorização por serviço. Essa abordagem resulta na criação de redes virtuais que permitem a criação de pontos de observação dedicados a cada serviço, simplificando a análise da rede e a compreensão do seu comportamento.

A camada de rede é responsável por encaminhar os dados para a camada de suporte, utilizando tecnologias como *Wi-Fi, Ethernet* ou conexões de rádio 4G/5G. Na camada de suporte, o autor sugere a implementação da computação de borda *(fog computing)* como parte central da infraestrutura. Nesse contexto, a camada de suporte desempenha um papel crucial, pois lida com todo o tráfego gerado, não apenas pela camada de percepção, mas também pela camada de aplicação. A solução proposta pelo autor envolve a implementação de um sistema de detecção de intrusões (IDS) baseado em aprendizado de máquina (AM) na camada de suporte. Os nós de computação de borda *(fog nodes)* atuam como pontos de observação da rede, capturando o tráfego de forma passiva e fornecendo os dados ao IDS. O IDS utiliza modelos de AM, Supervisionado, Árvore de Decisão, Naive Bayes, entre outros; treinados para distinguir entre o comportamento normal da rede e o comportamento anormal, que pode indicar um possível ataque.

Além disso, destaca a importância do treinamento prévio dos modelos de AM, o qual pode ser realizado com conjuntos de dados de treinamento construídos em ambientes controlados de laboratório ou em redes de pré-produção que se assemelham às futuras implantações em uma cidade inteligente. Como parte de sua avali-

ação, o autor utilizou o conjunto de dados ”IoT23,” que é um conjunto de dados do *Stratosphere Lab*, que contém registros de tráfego de dispositivos IoT rotulados como benignos ou maliciosos. Esse conjunto de dados é valioso para o desenvolvimento de algoritmos de aprendizado de máquina voltados à detecção de ameaças.

O autor detalha os procedimentos e descobertas obtidas ao aplicar modelos de Aprendizado de Máquina (ML) ao conjunto de dados, com o objetivo de identificar relações existentes nos dados e avaliar a eficácia da seleção de características na identificação de fluxos maliciosos em sua amostra.

5.2.1 Resultados de Classificação Binária

Prazeres (2022) utilizou a linguagem de programação *Python*, juntamente com a biblioteca *scikit-learn*, para implementar os algoritmos de AM, como Regressão Logística, *Random Forest* e *Naive Bayes*, para a classificação binária. Os procedimentos incluíram a preparação dos dados, a divisão do conjunto de dados em treinamento e teste, treinamento dos modelos e avaliação do desempenho.

Os resultados mostraram que os três modelos não são classificadores aleatórios e têm uma alta porcentagem de verdadeiros positivos. A análise da curva *ROC (Receiver Operating Characteristic)* demonstrou que os modelos apresentaram bom desempenho, com curvas próximas ao canto superior esquerdo, indicando alta eficácia na detecção de fluxos maliciosos. O modelo de *Random Forest* se destacou, apresentando os melhores resultados em termos de *Recall* e Precisão, obtendo um resultado de 99,95% em ambos, indicando um desempenho superior em comparação com os outros modelos.

5.2.2 Resultados de Classificação Multiclasse:

Além da classificação binária, o autor também abordou a classificação multiclasse, tentando classificar os dados em sete classes diferentes de ataques. No entanto, os modelos de AM tradicionais, como Regressão Logística e *Naive Bayes*, enfrentaram dificuldades na classificação precisa de todas as classes minoritárias. A precisão e *recall* diminuíram significativamente para essas classes menos represen-

tadas.

Os modelos de Redes Neurais Artificiais, tanto o modelo base quanto o modelo de Perceptrons Multicamadas (MLP), também não conseguiram identificar com precisão a classe de ataque menos comum (ataque tipo 4). No entanto, o MLP demonstrou uma vantagem significativa em termos de *Recall* (99,70%) e Precisão (99,57%), em comparação com o modelo de base (99,02% e 95,29% em *recall* e precisão res- pectivamente).

5.2.3 Resultados de Classificação Não Supervisionada:

O autor também explorou a classificação não supervisionada com base em técnicas de *clustering*, como o algoritmo *K-Means*. Para a classificação binária, o algoritmo foi configurado para identificar dois *clusters*. Os resultados mostraram um bom desempenho na identificação dos *clusters*, com base em uma análise da pontuação de silhueta.

Os resultados foram considerados bastante satisfatórios, com o modelo não supervisionado mostrando excelentes resultados na classificação binária. No entanto, para a classificação multiclasse, a relação entre os rótulos de *clusters* e os tipos de ataques apresentou desafios adicionais, exigindo análises mais detalhadas.

A seguir, vamos analisar os resultados de ambos os artigos e classificá-los, com o objetivo de criar um ranking de classificação.

5.3 Avaliação e Ranqueamento dos Artigos com Base em Critérios de Detecção e Mitigação de Ameaças

Nesta seção, avaliamos detalhadamente dois artigos: ”Predicting Malicious Software in IoT Environment Based on Machine Learning and Data Mining Techniques” e ”Data Science in Cybersecurity: Evaluating the Use of Machine Learning in an IOT-IDS.” Concentramos nossa análise na eficácia da detecção e mitigação de ameaças na Internet das Coisas (IoT). Avaliamos a eficiência na detecção de ameaças, o tempo de detecção e a qualidade dos conjuntos de dados e cenários propostos em cada artigo. Essa análise busca identificar o artigo de destaque na área de segurança

cibernética em ambientes IoT.

1. Eficiência na Detecção de Ameaças (aqui iremos focar nos resultados do algoritmo *Random Forest*, já que ambos realizaram testes com ele):
   - Artigo 1: Embora tenha alcançado excelentes resultados na classificação utilizando o *Random Forest*, 97,14% de precisão, o Artigo 2 superou-o em termos de precisão, o que é um indicador crucial na eficácia da detecção de ameaças.
   - Artigo 2: Demonstrou um desempenho notável ao utilizar o algoritmo *Ran- dom Forest*, alcançando uma precisão de classificação de 99,95%. Essa precisão é essencial na detecção eficaz de ameaças em ambientes IoT.

2. Tempo Requerido para a Detecção e Treinamento do Algoritmo:
   - Artigo 1: No artigo faltou detalhar um pouco o tempo necessário para treinamento do algoritmo e o tempo necessário para a detecção de *malware*.
   - Artigo 2: Aqui também não foi detalhado o tempo utilizado para treinar o algoritmo e para o algoritmo trazer resultados.

3. Base de Dados e Cenário de Avaliação:
   - Artigo 1: O autor usou o conjunto de dados IoT23 e aplicou a Análise de Componentes Principais (PCA) para reduzir a dimensionalidade do conjunto de dados, mantendo apenas os recursos mais relevantes e eliminando redundâncias, contribuindo para melhorar o desempenho dos algoritmos de classificação, evitando o *overfitting*.
   - Artigo 2: O autor também usou o conjunto de dados IoT23, que é uma fonte valiosa de dados rotulados para desenvolver algoritmos de detecção de ameaças. Além disso, o artigo propôs uma abordagem interessante de segmentação do tráfego, considerando as camadas de uma cidade inteligente,o que enriquece o cenário de avaliação.

Com base nessas análises, criamos uma escala de 0 a 10, onde 0 é um resultado insatisfatório e 10 é um resultado excelente, para atribuir aos artigos. A Tabela 2

traz o resultado obtido:

Tabela 2 – Quadro de Ranqueamento dos Artigos

Categoria	**Artigo 1**	**Artigo 2**
Eficácia na Detecção de Ameaças	7	9
Tempo Requerido para a Detecção e Treinamento do Algoritmo	3	3
Base de Dados e Cenário de Avaliação	10	8
TOTAL	20	20

Fonte: Autora

Ao analisar os resultados apresentados na Tabela 2 e as pontuações atribuídas a cada artigo, é evidente que o ”Artigo 2” supera o ”Artigo 1” no quesito ”Eficácia na Detecção de Ameaças.” No entanto, ao considerar a categoria ”Base de Dados e Cenário de Avaliação,” o ”Artigo 1” se destaca devido ao tratamento minucioso aplicado. No que diz respeito ao ”Tempo Requerido para a Detecção e Treinamento do Algoritmo,” ambos os artigos obtiveram pontuações que deixam margem para melhorias.

É notável que, apesar das diferenças nas pontuações em duas categorias, a pontuação total para ambos os artigos permanece a mesma. Isso sugere que os artigos possuem características complementares e são igualmente relevantes para o nosso contexto de pesquisa em cibersegurança. Eles contribuem de maneira significativa para a compreensão das técnicas de detecção de ameaças na Internet das Coisas (IoT) e, juntos, oferecem uma visão abrangente sobre o assunto.

6 CONCLUSÃO

Este estudo realizou uma análise comparativa detalhada de duas pesquisas que utilizam técnicas de aprendizado de máquina para a detecção de ameaças cibernéticas na segurança da Internet das Coisas (IoT). A crescente importância da IoT e a dependência cada vez maior da internet em nossas vidas destacam a necessidade de reforçar a cibersegurança para proteger sistemas e dados sensíveis.

Nossa análise enfatizou a eficiência na detecção de ameaças, o tempo neces-

sário para detecção e a qualidade das bases de dados como critérios essenciais na avaliação das abordagens de aprendizado de máquina. Ambos os artigos, apesar das diferenças em pontuações nesses critérios, mantiveram pontuações totais iguais. Isso sugere que eles possuem características complementares e são igualmente valiosos no contexto de pesquisa em cibersegurança.

Nosso estudo ressaltou a importância de uma abordagem multifacetada ao lidar com ameaças cibernéticas na IoT. Cada pesquisa ofereceu *insights* valiosos e técnicas únicas que, quando combinadas, podem proporcionar uma visão abrangente da detecção de ameaças. As abordagens de aprendizado de máquina continuam a desempenhar um papel crucial na proteção da IoT contra ameaças cibernéticas em constante evolução.

Para pesquisadores e profissionais de segurança cibernética, este estudo oferece uma compreensão mais profunda das estratégias e técnicas disponíveis para a detecção de ameaças na IoT, auxiliando na seleção das abordagens mais adequadas às suas necessidades. À medida que as ameaças cibernéticas evoluem, a pesquisa contínua e a colaboração são essenciais para manter a segurança na IoT e na ciberespaço em geral.

REFERÊNCIAS

ABRAHAM, J. A.; GEORGE, S. M. A survey on preventing crypto ransomware using machine learning. *In*: *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*. [S.l.: s.n.], 2019. v. 1, p. 259–263.

ALHARBI, A.; HAMID, M. A.; LAHZA, H. Predicting malicious software in iot environment based on machine learning and data mining techniques. *International Journal of Advanced Computer Science and Applications*, v. 13, 2022.

CESCHIN, F.; OLIVEIRA, L. S.; GRéGIO, A. Aprendizado de máquina para segurança: Algoritmos e aplicações. *XIX Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais - SBSeg*, p. 41–90, 2019.

DUA, S.; DU, X. *Data mining and machine learning in cybersecurity*. 1. ed. [S.l.: s.n.], 2011. v. 1. 16 p.

EOCO. *WHAT IS RANSOMWARE AND 15 EASY STEPS TO KEEP YOUR SYSTEM PROTECTED*. 2023. 5

GUPTA, B. B. et al. Fighting against phishing attacks: state of the art and future challenges. *Neural Computing and Applications*, v. 28, n. 12, p. 3629–3654, 2017.

GéRON, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'reilly media. [S.l.: s.n.], 2017. v. 1. 3-5 p.

GéRON, A. *Mãos A Obra: Aprendizado De Máquina Com Scikit-Learn, Keras TensorFlow: Conceitos, Ferramentas e Técnicas Para a Construção de Sistemas Inteligentes*. 1. ed. [S.l.: s.n.], 2021. v. 1. 1226 p.

KASPERSKY. *Aprenda sobre malware e como proteger todos os seus dispositivos contra eles*. 2023.

KASPERSKY. *CIBERAMEAÇA MAPA EM TEMPO REAL*. 2023. Acesso em: 24 de outubro de 2023.

MATHEUS, Y.; RAPHAELL, B.; CALANCA, P. *Quais são os 4 tipos de aprendizagemna IA, algoritmos e usos no dia a dia*. 2023. Acesso em: 23 de outubro de 2023.

MCAFEE. *O que é malware?* 2023. Acesso em: 26 de outubro de 2023. 13MELO, S. *Análise de Malware*. 1. ed. [S.l.: s.n.], 2023. v. 1. 177 p.

MITCHELL, T. M. *Machine Learning*. [S.l.]: McGraw-Hill Science/Engineering/Math;, 1997.

OBEIDAT, I. et al. Intensive pre-processing of kdd cup 99 for network intrusion classification using machine learning techniques. *International Journal of Interactive Mobile Technologies (iJIM)*, v. 13, p. 70, 01 2019.

OLIVEIRA, J. C. de. Ransomware - laboratório de ataque do wannacry. *Universidade de Brasília - UnB*, p. 81, Novembro 2018.

ÖZGÜR, A.; ERDEM, H. A review of kdd99 dataset usage in intrusion detection and machine learning between 2010 and 2015. *PeerJ Preprints*, v. 4, p. e1954v1, abr. 2016.

PANDE, J. Introduction to cyber security. *Uttarakhand Open University*, p. 152, 2017.

PRAZERES, N. A. G. d. Data science in cybersecurity: Evaluating the use of machine learning in an iot-ids. *https://iconline.ipleiria.pt/handle/10400.8/7214?mode=full*, p. 101, 2022.

RESEARCH, T. M. *Rethinking Tactics: 2022 Annual Cybersecurity Report*. [S.l.], 2022.

SARKER, I. H. et al. Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*, v. 7, n. 1, p. 41, 2020.

SILVA, J. V. V. Análise estatística sobre o conjunto de dados kdd-99 para o desenvolvimento de sistemas de segurança de rede usando aprendizado de máquina. *https://app.uff.br/*, p. 1–68, 2020.

SQUEFF, F. de H. S.; NEGRI, F. D. CiÊncia e tecnologia de impacto: Uma anÁlise do caso darpa. *https://portalantigo.ipea.gov.br/*, p. 1–30, 7 2017.

STEINBERG, J. *Cibersegurança para Leigos*. 1. ed. [S.l.: s.n.], 2020. v. 1. 368 p.

STRATOSPHERE. *Stratosphere Laboratory Datasets*. 2015. Acessado em 11 de Outubro de 2023 , em url https://www.stratosphereips.org/datasets- visão geral. Acesso em: 11 de outubro de 2023.

STRATOSPHERE. *Stratosphere Datasets Update: Quickly Browse and Search!* 2021. Acesso em: 11 de outubro de 2023.

SUN, N. et al. Data-driven cybersecurity incident prediction: A survey. *IEEE Communications Surveys Tutorials*, v. 21, n. 2, p. 1744–1772, 2019.

von Solms, R.; van Niekerk, J. From information security to cyber security. *Computers Security*, v. 38, p. 97–102, 2013.