COORDENADORIA DE ENGENHARIA DA COMPUTAÇÃO

ANDRÉ CHIERIGHINI

TIAGO MACHADO MALDONADO

APRENDIZADO DE MÁQUINA PARA ANÁLISE DE TRÁFEGO DE REDE E DETECÇÃO DE INTRUSÃO

Sorocaba/SP

2021

André Chierighini

Tiago Machado Maldonado

APRENDIZADO DE MÁQUINA PARA ANÁLISE DE TRÁFEGO DE REDE E DETECÇÃO DE INTRUSÃO

Trabalho de conclusão de curso apresentado ao Centro Universitário Facens como exigência parcial para obtenção do diploma de graduação em Engenharia da Computação.
Orientador: Prof. Johannes Von Lochter

Sorocaba/SP

2021

FICHA CATALOGRAFICA

ELABORADA PELA “BIBLIOTECA FACENS”

C533a
Chierighini, André.
Aprendizado de máquina para análise de tráfego de rede e detecção de intrusão / por André Chierighini, Tiago Machado Maldonado. - Sorocaba, SP: [s.n.], 2021.
90f.

Trabalho de Conclusão de Curso (Graduação) – Centro Universitário Facens – Coordenadoria de Engenharia Computação – Curso de Engenharia Computação, 2021.
Orientador: Prof. Johannes Von Lochter.

1. segurança da informação, 2. segurança de redes. 3. inteligência computacional. I. Maldonado, Tiago Machado. II. Centro Universitário Facens. III. Título.

CDD 621.39

André Chierighini

Tiago Machado Maldonado

APRENDIZADO DE MÁQUINA PARA ANÁLISE DE TRÁFEGO DE REDE E DETECÇÃO DE INTRUSÃO

Trabalho de conclusão de curso apresentado ao Centro Universitário Facens como exigência parcial para obtenção do diploma de graduação em Engenharia da Computação. Orientador: Prof. Johannes Von Lochter

Sorocaba, 24 de novembro de 2021

BANCA EXAMINADORA

_____

Prof. Dr. Johannes Von Lochter

_____

Prof. Me. Jones Artur Gonçalves

_____

Prof. Me. Lucas Nunes Monteiro

AGRADECIMENTOS

Os principais agradecimentos desse trabalho são para a comunidade *Open Source*, sem a qual este trabalho não seria possível. Em seguida ao Professor Johannes von Lochter, o qual orientou este trabalho e foi crucial para sua execução.

Os demais agradecimentos são para todos aqueles que ajudaram com a revisão ao ler e reportar erros e problemas textuais deste trabalho.

“O insucesso é apenas uma oportunidade para recomeçar com mais inteligência.”

Henry Ford

RESUMO

Devido ao crescente uso da internet nas últimas duas décadas, as tecnologias evoluíram para possuir capacidade de automação em quase todos os sentidos, de testes automatizados à robôs de respostas automáticas aos clientes de um serviço. No entanto, muitas instituições atualmente ainda não possuem medidas de prevenção significativas e automatizadas contra-ataques realizados via internet. No momento atual de pandemia e isolamento, o mundo vive uma situação na qual ataques cibernéticos estão em constante crescente, e é necessário que a segurança dos dados, serviços e ativos, tanto da instituição, quanto dos clientes seja garantida. Em sua grande maioria, ataques à segurança de informação são facilmente mitigados por medidas disciplinares aplicadas pelos membros da instituição, porém alguns desses ataques são apenas detectados pela análise dos dados de tráfego de rede. A escala na qual esses dados são acumulados torna impossível a análise por olhos humanos, e assim requer soluções automatizadas para garantir a segurança da sua aplicação e serviços oferecidos. A proposta do projeto é criar, com a utilização de técnicas de Inteligência computacional, um modelo capaz de identificar anomalias no tráfego de rede, podendo assim facilitar o trabalho de analistas de segurança e, consequentemente, contribuir para a privacidade em geral, dos dados, serviços e ativos da instituição na qual o modelo está aplicado. Assim, o modelo é capaz de identificar ataques à rede na qual estará implementada, permitindo assim uma resposta apropriada e rápida ao incidente. A precisão do modelo, embora alta, poderia ser melhorada para casos reais e utilizando métodos mais refinados de aprendizado de máquina. O código produzido para esse experimento está disponível em: https://github.com/chierighini/TCC

Palavras-Chave: segurança da informação, segurança de redes, inteligência computacional

ABSTRACT

Due to the rising use of the internet in the past two decades, technologies have evolved to be automated in every way possible, from automated tests to chatbots. However, a large number of institutions still lack significant automated countermeasures against digital attacks. During the current pandemic and isolation period, the world stands in a spot where cyber attacks are on a rise, and it is imperative for organizations ensure that their data, clients and assets are secure. Even though most cyber security threats can be mitigated through disciplinary actions involving employees, there is a share of attacks only detectable through network traffic analysis. Since this data is too large for analysis by humans, it requires an automated solutions to ensure protection to any services at play. The concept presented by this paper is to create, with computational intelligence techniques, a model capable of identifying anomalies in network traffic, making the jobs of security professionals easier and contributing to the overall safety where the model may be applied. The model is capable of identifying attacks and anomalies in the network, allowing for quick and appropriate responses to incidents. Even though the model accuracy is high, it could be improved by the use of more refined machine learning techniques. The code produced for this experiment is available at: https://github.com/chierighini/TCC

Key-words: Information security. Network security. Computational Inteligence.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
BR	Brasil
BN	Biblioteca Nacional
CCPA	Lei de privacidade do consumidor da California
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DARPA	Defense Advanced Research Projects Agency
DoS	Negação de Serviço
DDoS	Negação de Serviço Distribuída
GDPR	Regulamento geral da proteção de dados
IBGE	Instituto Brasileiro de Geografia e Estatística
IDS	*Intrusion Detection System*
KDD	*Knowledge Discovery and Data Mining*
LGPD	Lei Geral de Proteção de Dados
MIT	*Massachusetts Institute of Technology*
OMS	Organização Mundial da Saúde
TCP	Tag Distribution Protocol
UDP	User Datagram Protocol

SUMÁRIO

1 INTRODUÇÃO

Ao longo do tempo, o conceito de segurança vem se moldando de acordo com o desenvolvimento de novas tecnologias e demandas do mercado de trabalho, deixando assim práticas que uma vez garantiram a prevenção de problemas para trás e adotando novas medidas para garantir a proteção de seu trabalho, funcionários e dados.

Com a curva exponencial da evolução tecnológica, alavancada durante o século 19 até os dias atuais, as grandes companhias intensificaram o uso da internet em seus serviços internos e externos adotando o uso de práticas como *Cloud Computing*, ferramentas de Big Data, entre outros recursos que possibilitam o escalonamento e o alcance de seus processos, “a dependência de aplicações web são essenciais em nossas atividades diárias como transferências bancárias, compras ou apenas o compartilhamento de dados com outros” (*Clincy*, 2018). Com os problemas encontrados na sociedade atual como por exemplo o isolamento social, tais práticas tiveram uma ênfase maior, fazendo com que em algumas empresas mudassem permanentemente o jeito de gerenciar seus funcionários e serviços, tornando as conexões totalmente online.

Em paralelo a esses acontecimentos, indivíduos com intenções maliciosas enxergaram oportunidades em invadir tais sistemas e gerar diversos transtornos para as empresas, que não estão necessariamente ligados ao roubo de dados ou a exposição deles como por exemplo desestabilizar empresas concorrentes, ataques ativistas etc. Existem técnicas que sobrecarregam o tráfego de informações, ocasionando falhas na comunicação entre os serviços e eventualmente ocorrendo a interrupção da execução do mesmo.

Um dos exemplos mais conhecidos deste tipo de prática criminosas é o ataque DoS (*Denial of Service*) que consiste na saturação de uma banda ou equipamento através de diversas requisições oriundas de uma única máquina. Devido a maior facilidade de localização do computador responsável pelo ataque, originou-se uma variante chamada DDoS (*Distributed Denial of Service*) que consiste no uso de diversas máquinas para realizar as requisições ao serviço aumentando a capacidade do impacto e dificultando ainda mais a localização do

IP de origem, “os serviços online de um cliente podem ser seriamente danificados por ataques DDoS e de acordo com o uso de IoT (*Internet of Things*) e serviços na nuvem, o ambiente na internet pode ter um número enorme de tráfego de dados no futuro” (Hyun, 2017).

Antes do ataque ser consolidado, é possível observar mudanças na rede que possibilitam os profissionais de segurança a agirem de forma preventiva e evitar que possíveis problemas venham a surgir. Tais mudanças consistem em perda de performance da rede, indisponibilidade temporária e aumento de tráfego inesperado.

A solução estudada retrata o uso de inteligência computacional e *machine learning* para automação e análise de mudanças no tráfego de rede, sendo capazes de identificar anomalias prévias a fim de ser implementado, futuramente, em um sistema capaz de notificar os profissionais antecipadamente para agirem conforme o necessário. A necessidade de treinar esses modelos previamente a futuros ataques é fundamental para uma melhor distinção e evitar falsos alarmes, mesmo entendendo que é melhor ser realizado um falso alarme do que um ataque ser realizado e não haver notificações.

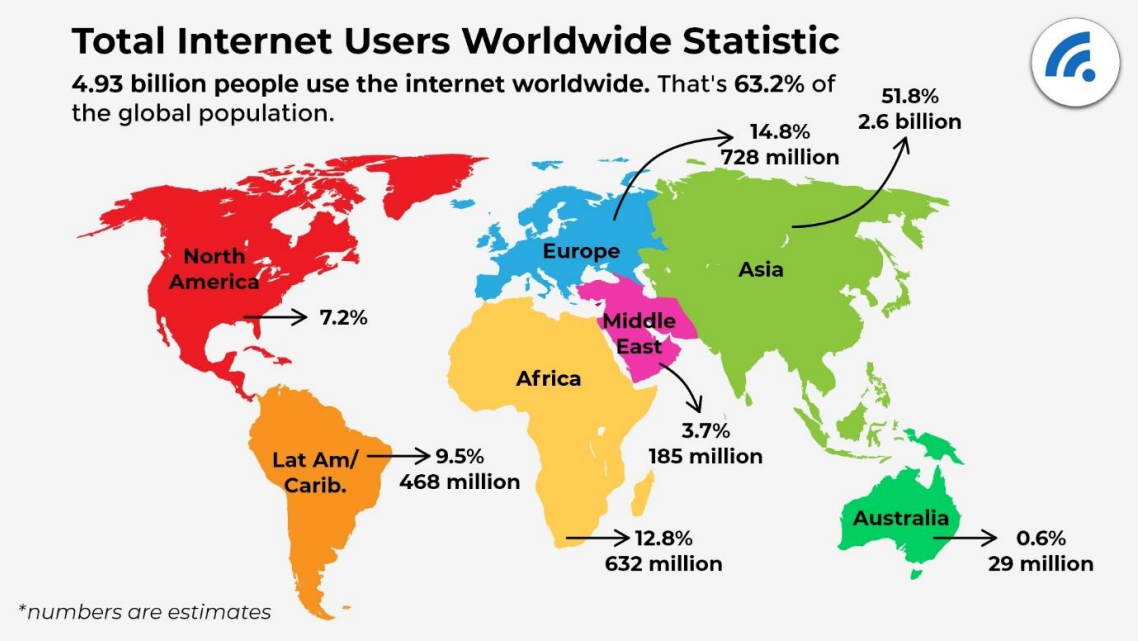
2 USO DA INTERNET E AMEAÇAS DIGITAIS

Durante os últimos anos com a modernização de serviços como compras, comunicações, entretenimento, entre outras que tornaram suas funcionalidades *online*, a sociedade observou uma curva de crescimento acentuada no uso da internet em seu dia a dia, principalmente pelo agravante da pandemia causada pelo Coronavírus que forçou a todos a permanecerem em casa e utilizarem funções remotas (SANCHES, 2021). Em paralelo a isso, as corporações enxergaram a capacidade dos serviços *online* e iniciaram um processo de adaptação virtual para seus serviços, emergindo assim diversos produtos e técnicas de implementação para essas situações.

2.1 Uso da internet entre a população mundial

A internet teve um enorme impacto na sociedade, principalmente após o início do século 21, contando com uma parcela importante no processo de globalização que disseminou padrões na maneira em que as pessoas iriam consumir produtos e conteúdos ao redor do mundo (UOL, 2021). Apesar dos grandes centros conterem em sua maioria uma população conectada virtualmente, as regiões periféricas têm majoritariamente problemas de acessibilidade a uma rede estável e condições de uso durante o decorrer do dia. Segundo um relatório produzido pela *Internet World Stats*, o mundo alcançou um total de 4,93 bilhões de pessoas conectadas a internet no final de 2020, sendo equivalente a 63,2% da população mundial (BROADBAND SEARCH, 2021) conforme mostra a Figura 1.

Figura 1 – Usuários da internet ao redor do mundo.



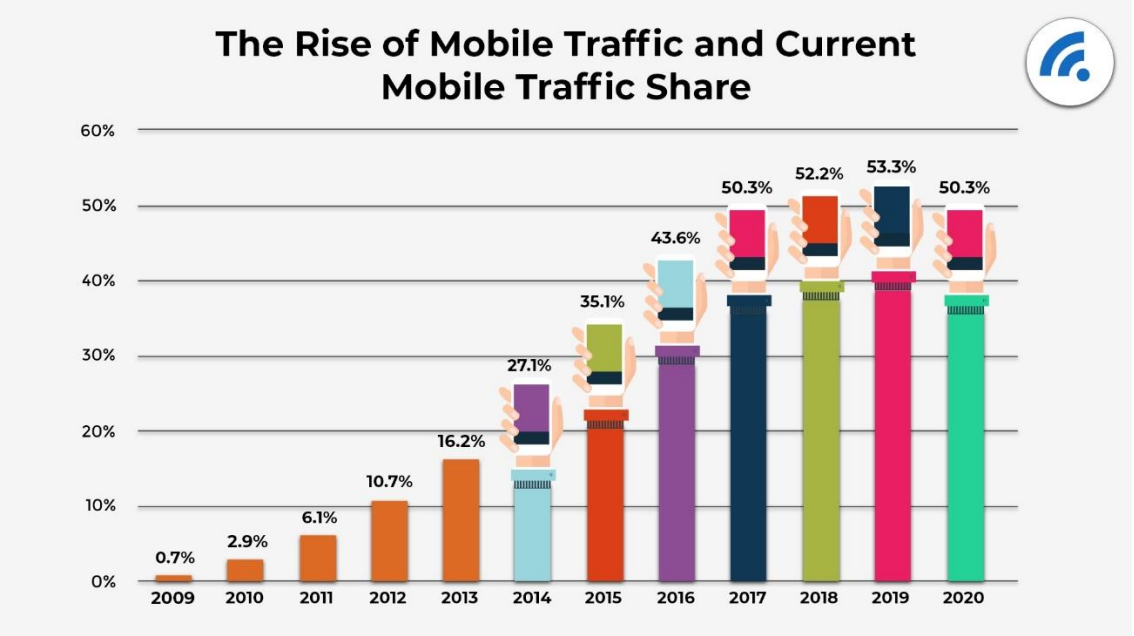
[1]Fonte: Disponível em: https://www.broadbandsearch.net/. Acessado em: 30 abr. 2021.

Apesar da África apresentar índices maiores de usuários conectados à internet que o subcontinente Norte-Americano, seus números estão diretamente ligados à sua proporção populacional que conta com o dobro de pessoas residindo no continente (COUNTRYMETERS, 2021). A América do Norte, juntamente com a Europa apresentam dados que indicam que cerca de 90% da população tem acesso à internet diariamente, sendo consideradas as regiões mais adaptadas para a conexão ao ambiente virtual. O acesso à tecnologia vem se tornando cada vez mais comum graças a redução de custos que facilitaram em países emergentes o crescimento tecnológico.

O acesso à internet por dispositivos móveis vem aumentando drasticamente nos últimos anos que, por motivos de praticidade e oportunidades, a população encontrou maneiras de gerar rendas, comprar, se entreter e principalmente compartilhar informações, conforme apresentado na Figura 2.

[1] Imagem retirada de: https://www.broadbandsearch.net/blog/internet-statistics.

Figura 2 – Crescimento do uso e tráfego de dados mobile.



[2]Fonte: Disponível em: https://www.broadbandsearch.net/. Acessado em: 30 abr. 2021.

Através de celulares e *tablets,* o mercado de *e-commerce* cresceu consideravelmente desde 2017 devido a possível fonte de renda extra para alguns com o desenvolvimento de propagandas e produtos digitais, assim como a praticidade de compra de produtos e serviços online para outros, que estimaram a geração total de 4.28 trilhões de dólares em 2020, dos quais os dispositivos móveis são responsáveis por aproximadamente 70% de toda a renda gerada pelo e-commerce (COPPOLA, 2020).

Com a população conectada a internet, as empresas encontraram maneiras de distribuir seus produtos em larga escala, gerando novas formas de entretenimento e serviços que, em sua maioria, foram bem recebidos pelo público. Somente o continente americano gera a cada minuto cerca de 3.14 milhões de GB de dados em tráfego de rede (BROADBAND SEARCH, 2021), fazendo com que diversas informações inclusive sensíveis circulem pela internet e assim gerando preocupações para as empresas manterem os dados seguros.

2.2 Uso da Internet no meio Corporativo

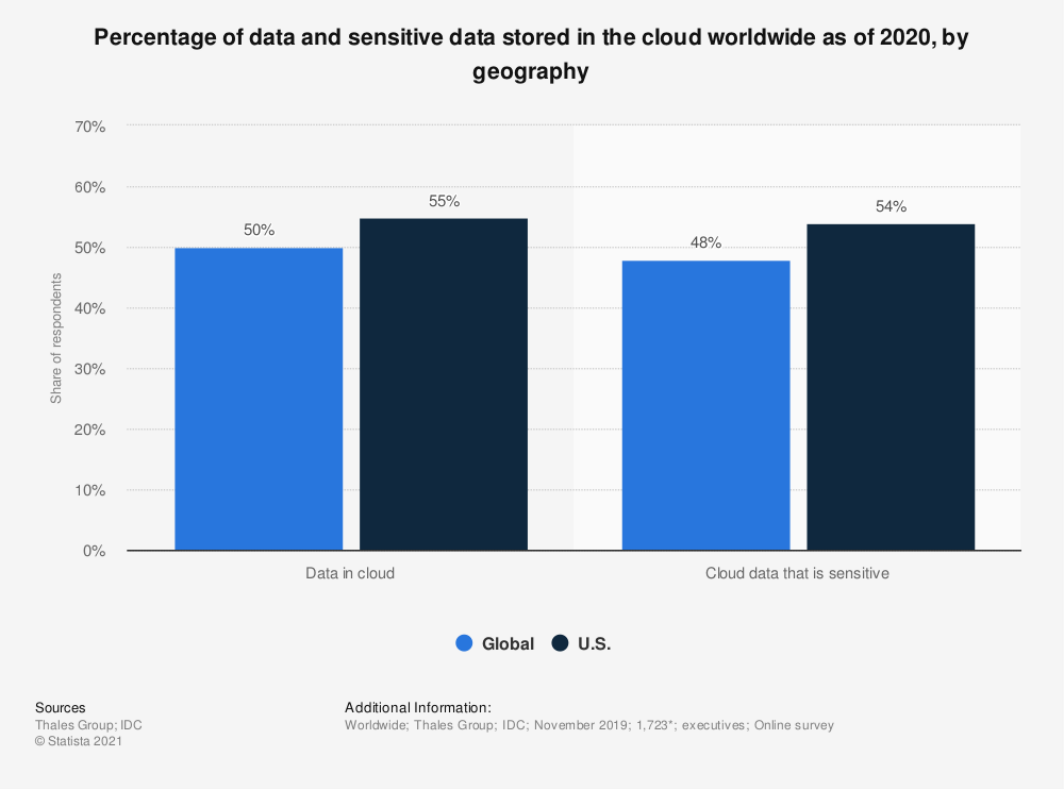
[2] Imagem retirada de: https://www.broadbandsearch.net/blog/internet-statistics.

Com a diversa gama de dados circulando através da rede e dos sistemas internos das empresas, foi observado a necessidade da criação de métodos preventivos a ataques e normas de proteção aos dados do público em geral, deste modo, fazendo com que surgisse normas como a GDPR (*General Data Protection Regulation*) na Europa em 2018, CCPA (*California Consumer Privacy Act*) no Estados Unidos em 2018 e a LGPD (Lei Geral de Proteção de Dados) no Brasil, entrando em vigor em 2020 e garantindo com que, através da justiça, as empresas desenvolvessem diversas técnicas para adaptação da maneira em que os dados alheios fossem visualizados, alterados e salvos.

Técnicas de *cloud computing* vêm se tornando cada vez mais comuns dentro do âmbito corporativo devido ao custo reduzido para a realização de serviços internos e armazenamento de dados. Empresas de grande porte como Amazon, Microsoft e Google realizaram grandes investimentos na adaptação de sua estrutura interna para além de seu uso próprio, possibilitar o uso de serviços na nuvem para empresas de pequeno e médio porte e assim fazendo com que, comparado com anteriormente quando tais empresas realizavam suas operações em seu próprio hardware, obtivessem gastos reduzidos em suas operações (MLITZ, 2021).

Para os dados armazenados na nuvem foi realizado uma pesquisa pela Statista que determinou que 50% dos dados mundiais estão salvos na nuvem dos quais 48% deles são sensíveis, conforme mostra a Figura 3.

Figura 3 – Proporção dos dados salvos em nuvem.



[3]Fonte: Disponível em: https://www.statista.com/. Acessado em: 30 abr. 2021.

Materiais armazenados na nuvem devem receber os devidos cuidados para não terem seus dados corrompidos ou vazados, principalmente quando se trata de informações particulares que podem comprometer a integridade não somente da companhia, como também do cliente para a exposição de hackers.

Da mesma maneira funciona para os serviços que funcionam através da rede, para que não haja ataques com o intuito de interromper o processo de execução, as empresas tiveram que entender como os ataques funcionam e assim gerar estratégias de prevenção.

2.3 Ataques e prevenções

Para a prevenção de incidentes que ameaçariam o pleno funcionamento de serviços *online*, amplamente utilizados atualmente e, especialmente, no momento de pandemia, os profissionais de segurança da informação possuem

[3] Imagem retirada de: https://www.statista.com/statistics/1202541/sensitive-data-cloud-location/.

diversas ferramentas à disposição. Desde detectores automáticos de vulnerabilidades, até *softwares* de antivírus.

Entretanto, há diversas categorias de ataques e inúmeros métodos de se detectar e prevenir tais ataques. Enquanto um detector de vulnerabilidades pode apontar onde há falhas exploráveis num *software*, pouco ele fará para detectar um ataque feito pela internet, cuja detecção seria feita analisando o tráfego recebido num intervalo de tempo nos *firewalls* de uma empresa, por exemplo.

2.4 Princípios da Segurança da Informação

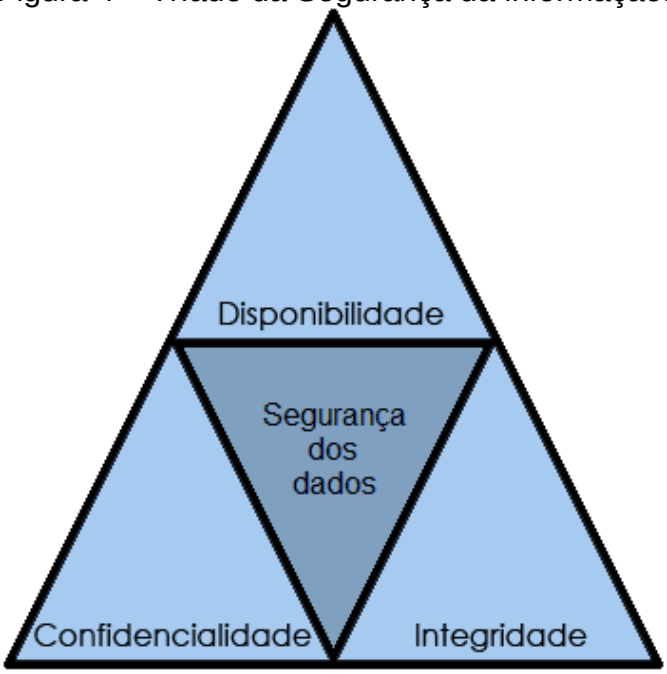
Quando se fala de segurança de informação, é necessário esclarecer os três pilares que devem ser mantidos para o pleno funcionamento de aplicações e serviços. Exemplificados na Figura 4, eles são:

a) Confidencialidade, o princípio de que a informação deve ser acessada apenas por partes que tem autorização para acessá-la;

b) Integridade, o princípio de que a informação deve se manter íntegra, ou seja, inalterada por partes não autorizadas ou sem a geração de registros da alteração. Isso acarreta que a informação deve ser modificada apenas pelas partes que tem autorização para fazer alterações e o fazem de forma explícita, de forma a registrar as mudanças feitas;

c) Disponibilidade, o princípio de que a informação deve estar disponível para o acesso por partes autorizadas quando necessário.

Figura 4 – Tríade da Segurança da informação.



[4]Fonte: Disponível em: https://www.gta.ufrj.br/ . Acesso em: 30 abr. 2021.

Além dos princípios citados acima, há também a necessidade de autenticidade e o princípio de não repúdio. Ambos estão interligados no fato de que se deve, para toda e qualquer transação feita não pode ser negada por nenhuma das partes envolvidas. Para cada linha alterada num banco de dados, por exemplo, deve estar registrado com clareza o que foi alterado e por quem.

2.5 Gerenciamento De Risco

Para todo e qualquer ativo de uma empresa, há um risco atrelado (LALONDE; BOIRAL, 2012). Seja um servidor que é responsável pelo funcionamento de uma aplicação, um roteador ou uma impressora. Em cada ativo há a possibilidade de problemas e falhas, um custo para sua reposição e um prejuízo atrelado ao tempo de reposição do ativo. Uma impressora quebrada causará uma fração do prejuízo de um servidor parado e custará muito menos para ser reposta.

Os problemas envolvendo ativos da empresa são o que criam a necessidade de manter planos de gerenciamento de risco e resposta a

[4] Imagem retirada de: https://www.gta.ufrj.br/grad/12_1/seg_smartgrid/possiveisataques.html.

incidentes, uma vez que podem gerar desde prejuízo, até problemas legais para a empresa, no caso de vazamentos de dados sensíveis, como números de cartões de crédito.

A prioridade dada a cada possível incidente é determinada pelos fatores de risco atrelados ao ativo e o custo de reposição deste (LALONDE; BOIRAL, 2012). Se o ativo possui altas chances de gerar um incidente, e o impacto causado pela sua perda é alto, este deve ser prioridade quando se faz um plano de gerenciamento de risco.

Embora qualquer incidente que atrapalhe o funcionamento de um ativo possa ser considerado uma ameaça, há apenas a necessidade de se preparar para os eventos mais plausíveis. Numa região sem ocorrência de terremotos não há a necessidade de se preparar para o eventual impacto de um terremoto, porém se o contrário for verdade, deve se levar em consideração tudo o que pode ser feito para evitar a indisponibilidade de seu serviço.

A melhor forma de manter o ambiente interno de uma empresa seguro é estar preparado previamente para ataques, já que isso elimina a necessidade de responder a incidentes. Além disso, a prevenção pode impedir que a empresa sofra ações legais tomadas pelos afetados por um vazamento de dados sensíveis, por exemplo.

Atrelado a isso, há o conceito de defesa em profundidade, o qual consiste em implementar camadas redundantes de defesa para manter o ambiente da empresa seguro e em funcionamento.

Entre os tipos de controles de segurança há:

a) controles físicos, os quais consistem em prevenir o acesso aos sistemas de TI fisicamente;
b) controles administrativos, aplicados por meio de políticas empresariais que guiem como os processos internos devem ser executados;
c) controles técnicos, que são implementações de software e hardware responsáveis por proteger os ativos da empresa.

O método discutido neste trabalho cobre uma das possibilidades de proteção preventiva, na qual o tráfego de rede é analisado para detectar e prevenir incidentes antes que aconteçam.

2.6 Resposta a Incidentes

Caso todas as medidas discutidas anteriormente falhem, é necessário haver um plano de resposta a incidentes para minimizar o impacto causado (CYNET, 2021). Isso é essencial quando se trata de perda de lucros por indisponibilidade de serviço ou vazamento de dados sensíveis.

É necessário que a resposta a incidentes seja feita de forma adequada, tanto do ponto de vista da empresa, cujo lucro e reputação depende do pleno funcionamento de seus serviços, assim como do cliente, cuja confiança garante sua lealdade aos serviços oferecidos.

O gerenciamento de risco e a preparação para respostas a incidentes necessitam que se entenda os riscos atrelados a cada ativo da empresa. Os riscos mais comuns são os riscos de falhas, ataques e desastres naturais.

Todo equipamento pode apresentar falhas. E isso é um risco para o pleno funcionamento de uma empresa, por exemplo, quando um roteador que liga todos os computadores a internet para de funcionar, por exemplo.

Ataques costumam ser mitigados por meio de controles técnicos e físicos. Ataques vindos por meio da internet são prevenidos com o uso de firewalls, criptografia e mecanismos de autenticação, por exemplo. Os controles físicos garantem que ninguém poderá entrar na empresa e roubar equipamentos ou ter acesso aos sistemas internos.

A proteção contra desastres naturais é de difícil execução, costuma ser limitada a planos para manter o os serviços funcionando em casos de falta de energia e situações atípicas ao dia a dia, como enchentes por exemplo.

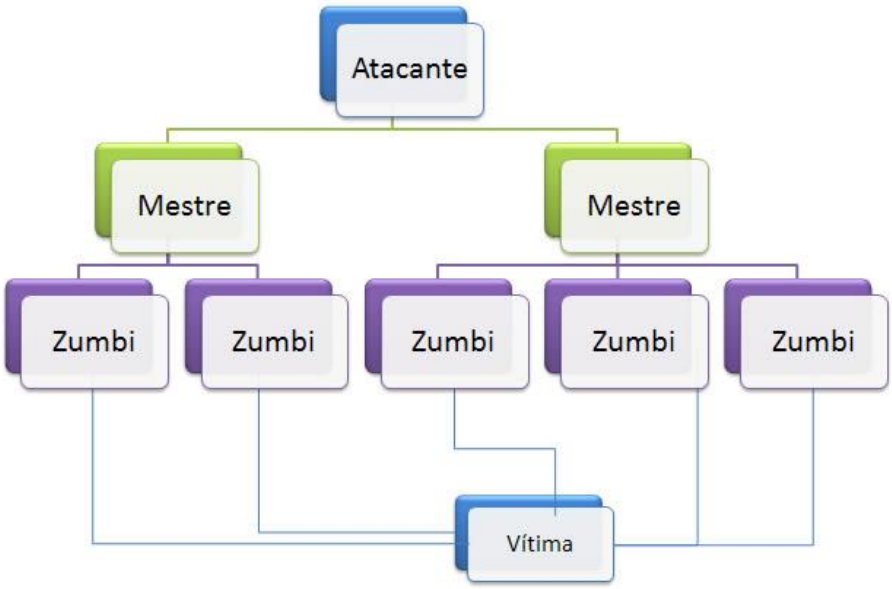
2.7 Ataques Cibernéticos

Ataques são caracterizados como explorações de uma vulnerabilidade presente em algum ativo de uma rede de computadores conectada a internet.

Sempre que há uma vulnerabilidade a ser explorada, ela pode ser utilizada para operações maliciosas que possam tirar um serviço do ar, roubar dados e danificar os ativos da empresa.

É considerado um ataque ativo, quando a intenção é afetar as operações de um sistema, como por exemplo um ataque DoS, no qual inúmeras requisições são feitas a um mesmo sistema com a intenção de sobrecarregar o tráfego da rede, assim impedindo o uso por outros usuários conforme mostra a Figura 5.

Figura 5 – Ataque DDoS.



[5]Fonte: Disponível em: https://www.techtudo.com.br/ . Acesso em: 30 abr. 2021.

Ataques passivos tem a intenção de adquirir informações de um sistema, mantendo seu funcionamento inalterado, e são mitigados por meio de prevenção, já que são muito difíceis de serem detectados. Um exemplo disso são os chamados *keyloggers* que registram cada tecla digitada numa máquina, sendo assim capaz de descobrir senhas e outras informações confidenciais.

É de grande importância que ataques sejam prevenidos, mitigados e detectados para qualquer serviço ou aplicação, já que interferem diretamente nos princípios básicos da segurança da informação, e causam grandes prejuízos para os afetados.

[5] Imagem retirada de: https://www.techtudo.com.br/artigos/noticia/2012/01/entenda-o-que-sao-os-ataques-dos-e-ddos.html.

Analisar o tráfego de rede, como proposto neste trabalho, é uma maneira de detectar tráfego anormal e tomar as medidas necessárias para mitigar quaisquer ataques que estejam sendo executados.

2.8 Sistemas de Detecção de Intrusão

Os mecanismos de detecção de intrusão estão presentes em redes de computadores pelo mundo, e tem a finalidade de garantir a segurança das redes, detectando anomalias no comportamento comum de eventos na rede (UFRJ, 2021).

As anomalias detectadas normalmente são decorrentes das ações de hackers e precisam ser detectadas para evitar roubos, alterações e destruição de dados e sistemas.

2.8.1 Honey Pots

Honey Pots são sistemas utilizados para atrair os hackers tentando entrar num sistema, que consiste de um local sem riscos para o hacker agir e ter suas atividades catalogadas a fim de usar esses dados para mitigar suas tentativas de intrusão (LEOBONS, 2021).

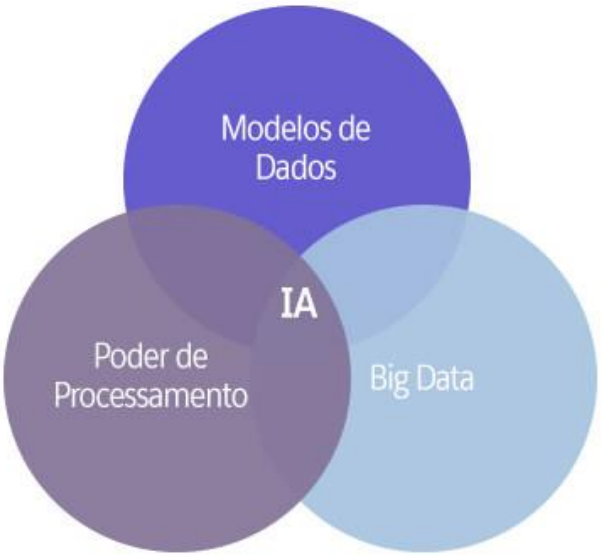
3 INTELIGÊNCIA ARTIFICIAL

Apesar dos conceitos de inteligência artificial começarem a surgir a partir da década de 50 após a conferência de Dartmouth (ZIZU, 2018), foi somente nos últimos anos que a sociedade foi capaz de observar a presença de sua utilização em quase todas as áreas de atuação, desde a geração de entretenimento até segurança (TODOROV, 2021) devido aos avanços tecnológicos que melhoram as condições em que os pilares de IA funcionam.

3.1 Conceitos Básicos

Essencialmente, uma inteligência artificial é dar a capacidade de pensar à uma máquina com a possibilidade de aprender, raciocinar, perceber e decidir a partir de uma sequência de análises. Atualmente, os pilares que possibilitam tais atividades a uma máquina alcançaram condições suficientes para escalar esses conceitos a funções importantes na sociedade, que antigamente eram inviáveis visto a infraestrutura tecnológica da época. Tais pilares são modelos de dados bem estruturados para a análise dos dados, a grande densidade de dados disponíveis, também conhecidos como Big Data e por fim a alta capacidade de processamento que as máquinas atuais possuem.

Figura 6 – Pilares da Inteligência Artificial.



[6]Fonte: Disponível em: https://www.salesforce.com/br. Acesso em: 11 set. 2021.

Devido principalmente a grande evolução da internet ao redor do mundo, possibilitou-se então a captação em massa de dados em todos os segmentos possíveis e a computação em nuvem que fez com que as máquinas alcançassem um outro nível em seu poder de processamento.

3.2 Aprendizado de Máquina

Sendo o principal impulsionador para a inteligência artificial, *Machine learging*, ou em português aprendizado de máquina, busca realizar um método diferente para o alcance dos resultados esperados ao qual diferente do método tradicional que é criado regras para se obter um resultado desejado, em aprendizado de máquina as regras são criadas a partir da análise automática dos dados, tornando a cada iteração o modelo de dados melhor e mais preciso, alcançando assim resultados de maneira autônoma. De modo geral, seu objetivo é utilizar algoritmos mínimos que sejam capazes de organizar dados, reconhecer padrões e assim fazer com que as máquinas sejam capazes de aprender e

[6] Imagem retirada de: https://www.salesforce.com/br/products/einstein/ai-deep-dive/

gerarem seus próprios resultados, devido a quantidade de dados massiva encontrada hoje em dia, é possível que seja aprimorado cada vez mais os modelos, fazendo com que os sistemas consigam identificar cada vez mais padrões e encontrarem respostas mais assertivas (IBM, 2020).

Segundo Tamir (2020), um típico algoritmo de aprendizado de máquina é composto por três etapas:

a) Um processo de decisão, neste passo, o algoritmo utiliza os dados alimentados para criar uma predição do resultado.
b) Uma função de erro, ou seja, uma comparação com modelos já conhecidos para avaliar a predição.
c) Um processo de atualização ou otimização, para verificar se o modelo pode se adequar melhor aos dados utilizados e fazer essa melhoria de forma autônoma até um limite de precisão.

Os algoritmos de aprendizado de máquina podem ser classificados como supervisionados e não supervisionado, possuindo algumas características diferentes para cada abordagem.

3.2.1 Algoritmos supervisionados

A principal característica deste grupo é ser capaz de realizar predições utilizando conjuntos de dados rotulados, e ajudam organizações a resolver problemas no mundo real (IBM, 2020).

Alguns exemplos de métodos utilizados em aprendizado supervisionado são: redes neurais, naïve bayes, regressão linear, regressão logística, dentre outros.

3.2.2 Algoritmos não supervisionados

No aprendizado não supervisionado, são analisados conjuntos de dados não rotulados, e todas as variáveis são utilizadas no processo de classificação. Os algoritmos não supervisionados podem ser usados para realizar a rotulagem dos dados para a implementação de um algoritmo supervisionado subsequente (UNSUPERVISED AND SEMI-SUPERVISED LEARNING, 2020).

3.2.3 Algoritmos semi-supervisionados

Algoritmos semi-supervisionados são a mistura dos dois métodos e utilizam de conjuntos de dados rotulados, proveniente de conjuntos maiores e não rotulados, sendo a melhor abordagem quando não se há dados rotulados suficientes para realizar um aprendizado supervisionado (IBM,2020)

3.3 Aprendizagem profunda

Considerado uma evolução do aprendizado de máquina, *Deep learning*, ou em português aprendizagem profunda, visa os mesmos objetivos de *Machine learning* porém necessitando de uma quantidade ainda maior de dados e abordando de forma diferente o processo de aprendizado tentando replicar a rede neural humana.

Similar ao funcionamento dos neurônios no cérebro, os algoritmos de aprendizagem profunda criam camadas hierárquicas que controlam a propagação, conexões e a direção que o dado irá seguir no processo de análise tornando a cada iteração um pouco mais abstrato e composto a informação para a próxima etapa, fazendo assim gerar resultados mais complexos acima de cada dado (MARBLESTONE; WAYNE; KORDING, 2016).

O processo de conexão entre as camadas, desde a entrada até a saída do dado é conhecido como CAP (*Credit assigment path*) podendo haver ilimitadas camadas e tem como objetivo apresentar possíveis semelhanças entre a entrada e a saída alcançada, não existe uma definição especifica que diferencie um aprendizado profundo mas é de comum senso entre os pesquisadores que é necessário ter mais de 2 camadas internas de profundidade, considerando ainda a camada de saída (SUGIYAMA, 2019).

Graças ao seu desenvolvimento, está sendo possível com que outras áreas relacionadas se desenvolvam tais como processamento de linguagem natural, diagnósticos por imagens, detecção de fraudes, análises de sentimento, sistemas de recomendação, visão computacional, previsão de falhas etc. (EQUIPE TOTVS, 2020).

Deng (2014) cita definições importantes para a compreensão do conceito de *Deep Learning* que se resumem em uma classe de técnicas de *machine learning* que utiliza vários níveis de representação para modelar relacionamentos mais complexos entre os dados do conjunto em questão.

3.4 Algoritmos de aumento de gradiente

Segundo Friedman (2002), um algoritmo de aumento de gradiente em árvores de decisão constrói modelos de regressão aditivos ao encaixar sequencialmente uma função parametrizada simples aos atuais “pseudo”-resíduos por quadrados mínimos a cada iteração. Em outras palavras, *Gradient Boosting*, como é conhecido em inglês, treina vários modelos de maneira gradual, aditiva e sequencial, na medida que cria várias árvores de decisão, sendo que cada nova árvore é encaixada numa versão modificada do conjunto de dados original (SINGH, 2018).

3.5 Ensembles

Opitz e Maclin (1999) definem *Ensembles* como um conjunto de classificadores treinados individualmente cujas predições são combinadas para classificar uma instância específica.

Ensembles são métodos eficazes para obter predições precisas ao combinar outras menos precisas, já que fazem uma “votação” ponderada das predições de todos os modelos combinados (DIETTERICH, 2000).

3.6 Big data

Com o crescente uso da internet, foi possível a geração e captação em massa de dados oriundos de todas as regiões e áreas da sociedade fazendo com que surgisse o conceito que conhecemos hoje como *Big data*. Dividido em em três V’s sendo eles volume, velocidade e variedade (ORACLE, 2021), é importante uma quantidade imensa de dados para assim as máquinas terem uma visão maior do escopo com que estão trabalhando, a necessidade de uma velocidade alta no recebimento dos dados e análise deles fazem com que

serviços na internet tenham que operar quase ou em tempo real no processamento dos dados e pôr fim à grande variedade de dados não estruturados e semiestruturados que podem variar suas origens como textos, áudios e imagens e assim necessitar de um pré-processamento.

O processo de Big data acontece essencialmente em três etapas que se inicia na integração através da captação dos dados de aplicativos e sistemas em geral que geram terabytes ou petabytes de dados e assim serem processados para que haja a possibilidade de armazenamento e trabalho das regras de negócio acima das informações recebidas, o próximo passo é o gerenciamento para armazena-los, deste modo houveram crescentes aumentos no uso da nuvem devido a sua facilidade no controle de recursos e acessos remotos, por fim a última etapa é a de analise que são gerados modelos de dados para *machine learning* e inteligência artificial encontrarem padrões, realizarem predições e assim descobrir novas demandas.

3.7 Métricas

Ao tratar de resultados de aprendizado de máquina é necessária a apresentação de métricas que deem o embasamento necessário para a explicação do que foi obtido.

3.7.1 Feature importance

Feature importance é uma métrica que diz a importância de uma variável no conjunto de dados para a predição do modelo. A seleção das variáveis do *dataset* pode impactar de forma significativa o modelo de *machine learning*, e por isso deve ser tratada com atenção (SHAIKH, 2018), (ZIEN; KRÄMER; SONNENBURG; RÄTSCH, 2009).

3.7.2 Log loss

Log loss diz a quão próxima é a probabilidade de predição para o valor verdadeiro, e aumenta conforme a probabilidade difere do valor real (DEMBLA, 2020).

Segundo Vovk (2015) a função *Log loss* corresponde à teoria clássica da aleatoriedade e vários casos é a forma preferida de avaliar funções de perda para aplicações de *machine learning,* ou seja, funções que avaliam a qualidade da predição realizada.

3.7.3 Matriz de confusão

Matrizes de confusão são tabelas que apresentam a performance de classificação de um modelo, e é importante para entender a quantidade de acertos e erros encontrados, na medida que mostra os falsos positivos e negativos obtidos (BROWNLEE, 2016), (MIDWEST ARTIFICIAL INTELLIGENCE AND COGNITIVE SCIENCE CONFERENCE, 2011).

Para a classificação binária, a matriz de confusão tem como formato duas linhas e duas colunas dos quais as coordenadas (0,0) representa os verdadeiros negativos, (0,1) representa falsos positivos, (1,0) representa os verdadeiros positivos e (1,1) representa os falsos positivos (SCIKIT-LEARN, 2021).

3.7.4 F-score

F-score também chamado de *F1-score* mede a acurácia de um modelo no seu respectivo conjunto de dados, sendo uma métrica de maior importância quando os casos de falso positivo e falso negativo são cruciais para a análise dos modelos de classificação binária (WOOD, 2021).

Seu valor é resultado da relação entre as métricas de *Precision* e *Recall* dada pela seguinte equação:

$$F1 = 2 * \frac{(precision * recall)}{(precision + recall)} \quad (1)$$

O qual *Recall* é referente a todos os casos que de fato deveriam ser verdadeiros, dado pela relação dos casos de verdadeiros positivos com a soma casos verdadeiros positivos e falsos negativos, apresentado na formula a seguir:

$$Recall = \frac{Verdadeiros\ positivo}{(Falsos\ negativo + Verdadeiros\ positivo)} \quad (2)$$

Já a métrica de *Precision* indica a relação dos casos de verdadeiros positivos pela soma de casos verdadeiros positivos e falsos positivos, apresentado pela seguinte formula:

$$Precision = \frac{Verdadeiros\ positivos}{(Verdadeiros\ positivos + Falsos\ positivos)} \quad (3)$$

Quanto mais próximo o valor de F-score for próximo de 1, melhor é a relação do modelo com o caso analisado (HUILGOL, 2019).

3.7.5 Acurácia

A Acurácia é a métrica utilizada para avaliação da qualidade do modelo quando todas as classes, como falsos e verdadeiros, positivos e negativos são igualmente importantes e seu resultado é dado pela seguinte equação:

$$Acurácia = \frac{(VP + VN)}{(VP + VN + FN + FP)} \quad (4)$$

Sendo VP os casos de Verdadeiros positivos, VN os casos de Verdadeiros negativos, FN os casos de Falsos negativos e FP os casos de Falsos positivos (HUILGOL, 2019).

4 MÉTODO

Durante a realização das partes práticas deste trabalho, houveram poucas complicações, sendo o pré-processamento o que necessitou maior atenção para a realização correta e, consequentemente, os resultados obtidos

4.1 Dados

O conjunto de dados KDD Cup 1999 surgiu a partir do conjunto de dados DARPA1998 criado pela agência americana DARPA e pelo Laboratório de Pesquisas da Força Aérea dos Estados Unidos, o qual foi trabalhado pelos laboratórios de pesquisa do MIT e utilizado na Terceira Competição Internacional de Ferramentas de Mineração de Dados e Descoberta de Conhecimento (*The Third International Knowledge Discovery and Data Mining Tools Competition*), da qual recebeu o nome KDD Cup 1999 (Gomes, 2019).

Tavallaee, Bagheri, Lu e Ghorbani (2009) sugeriram melhorias no conjunto de dados KDD Cup 1999, apresentando o conjunto NSL-KDD, o qual resolvia algumas redundâncias presentes no conjunto original e é o conjunto de dados utilizado para este trabalho.

O conjunto NSL-KDD possuía 44 colunas com diversas informações do tráfego de rede separadas em linhas para cada conexão estabelecida. Entre as informações contidas no conjunto é possível encontrar o tipo de conexão estabelecida, como UDP ou TCP, número de tentativas de login que falharam, se a mensagem é urgente ou não, entre outros. Posteriormente as colunas foram reduzidas de acordo com sua correlação com a variável alvo, nomeada *attack_flag*. Todas as informações em relação ao conjunto de dados foram esclarecidas no trabalho de Choudhary e Kesswani (2020).

4.2 Ferramentas

As ferramentas utilizadas consistem principalmente de elementos presentes na linguagem de programação *Python*. O projeto de código fonte aberto *Jupyter*, que permite a criação de scripts com células independentes umas das outras, facilitando a realização de experimentos relacionados a ciência

de dados e aprendizado de máquina, por exemplo, foi a base para a escrita do código utilizado.

Para trabalhar com os *datasets* foi necessário o uso da biblioteca *pandas* que permite a leitura, escrita e manuseio dos dados do conjunto, e foi crucial para a análise exploratória dos dados.

A biblioteca *sweetviz* foi utilizada para visualizar as relações entre as variáveis disponíveis e a variável alvo, permitindo assim a análise de cada variável individualmente.

Após realizado a análise exploratória e o pré-processamento, foi utilizada a biblioteca *mljar-supervised* para testar diferentes algoritmos e modelos de aprendizado de máquina e definir quais obtiveram o melhor resultado de predições, apresentando relatórios detalhados para justificá-los.

Os resultados percentuais adquiridos para cada execução dos algoritmos foram obtidos a partir da biblioteca *sklearn* a qual realiza os testes ao comparar as predições do modelo com os dados já existentes para os conjuntos de dados de teste.

O projeto foi escrito inicialmente utilizando o *Google Colab*, aonde foram obtidos os primeiros resultados. Em seguida foi movido para o *Visual Studio Code* e salvo em um repositório no *GitHub*, contendo o código fonte, os arquivos de texto do conjunto de dados e os relatórios mais recentes dos algoritmos testados pela biblioteca *mljar-supervised*.

Tabela 1 – Softwares utilizados no protótipo.

SOFTWARE	VERSÃO
Python	3.8.10
Jupyter	4.7.1
sweetviz	2.1.3
mljar-supervised	0.11.0
sklearn	0.24.2
pandas	1.3.3

Fonte: elaborada pelo autor.

A tabela 1 apresenta apenas os softwares instalados manualmente, e não contém todas as dependências do projeto, as quais podem ser acessadas por via do repositório no *GitHub*.

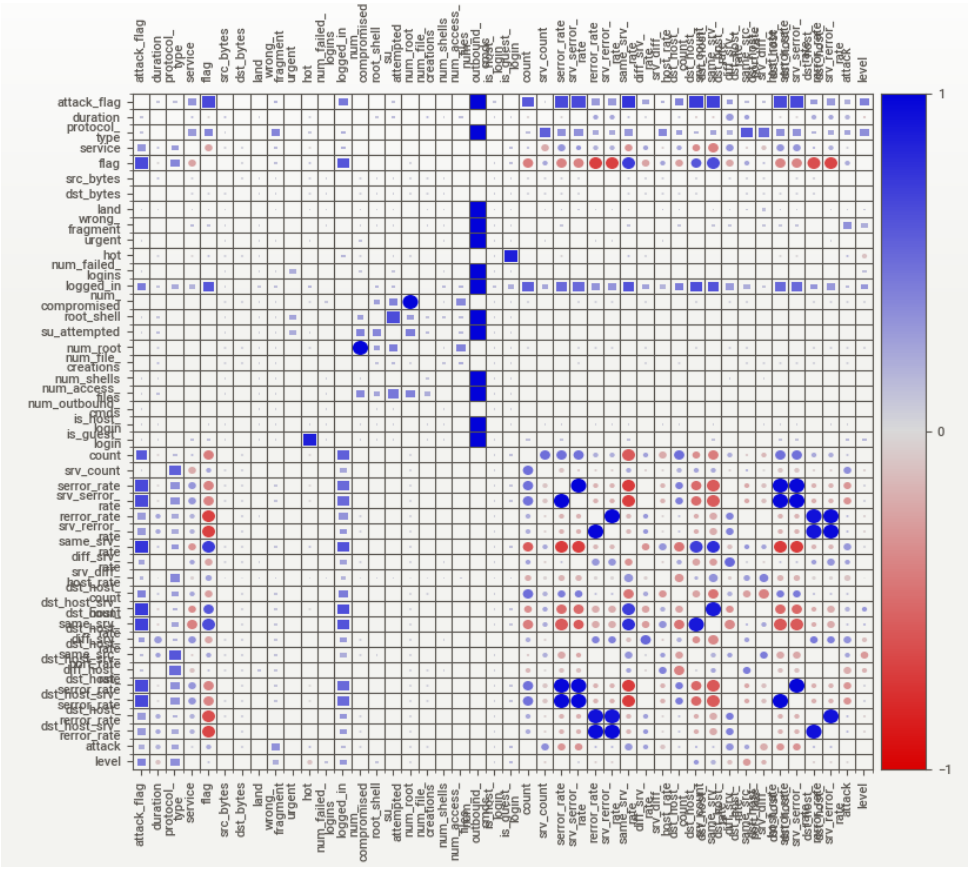
4.3 Pré-processamento

Inicialmente não haviam rótulos no *dataset*, com exceção dos formatos *Excel* e ARFF, e após complicações para ler estes formatos de arquivo com a biblioteca *pandas*, a solução foi encontrar os rótulos na página original do conjunto no site *kaggle*.

O conjunto de dados possuía 44 colunas, com diversas informações referentes ao tráfego de rede, e cada conexão estabelecida. O primeiro passo realizado foi tratar as variáveis categóricas referentes ao protocolo de rede, as *flags* da mensagem, o serviço de rede e o tipo de ataque realizado. As colunas foram transformadas em números inteiros, representando cada valor possível.

Em seguida, com o uso da biblioteca *sweetviz*, foram gerados relatórios das relações entre as variáveis do conjunto com a coluna alvo, incluindo a tabela de correlações das variáveis, demonstrada na figura 7.

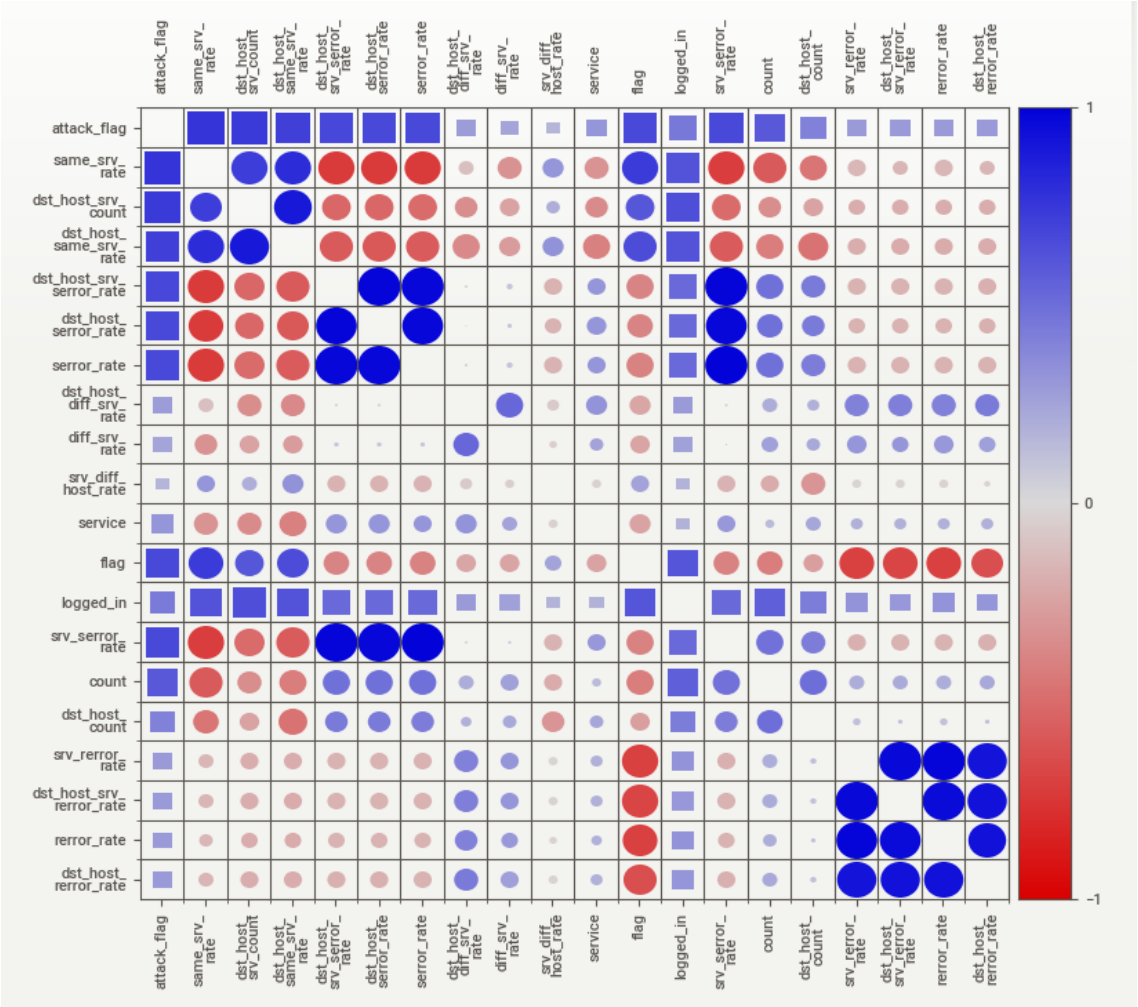
Figura 7 – Tabela de correlações inicial do conjunto de dados.



Fonte: Elaborada pelo autor.

Ao utilizar essas informações, foi possível extrair as colunas menos significativas para a coluna alvo, de acordo com as correlações, e obter um conjunto de dados mais enxuto e significativo, como demonstrado na tabela da figura 8.

Figura 8 – Tabela de correlações após o pré-processamento.



Fonte: Elaborada pelo autor.

Entre as colunas remanescentes vale notar que as variáveis “*service”, “count”* e “*dst_host_srv_serror_rate”* possuem grande correlação com a variável alvo, já que estas colunas se referem, respectivamente, ao serviço utilizado pela rede de destino, ao número de conexões realizada por um mesmo *host* e ao número de conexões destinadas a um mesmo *host*.

4.4 Algoritmos

Terminadas as etapas de pré-processamento, foi iniciado o treinamento dos modelos pela biblioteca *mljar*, resultando em alguns algoritmos sendo escolhidos de acordo com seu resultado, a partir do procedimento de treino utilizando a divisão 25/75% do conjunto de dados.

Os algoritmos testados durante a fase inicial foram: *DecisionTree, Xgboost* e *Random Forest*. A métrica de avaliação utilizada para os modelos foi *Log Loss* e apontou os algoritmos de *Xgboost* e *Random Forest* como os mais precisos nesta etapa, apresentados na tabela 2.

Em seguida foram selecionados os dois melhores métodos avaliados para serem analisados com maior profundidade. Devido a melhor performance do algoritmo *Xgboost*, foi escolhido outro algoritmo de aumento de gradiente: *LightGBM*, a fim de avaliá-lo para o caso de estudo.

A partir desse ponto foi esclarecido que os modelos de aumento de gradiente geraram resultados melhores. A fim de observar melhor os resultados, foi gerado um *Ensemble,* a partir de uma lista de algoritmos, na qual os três anteriores estavam incluídos, para verificar qual o melhor candidato.

4.5 Trabalhos Correlatos

Detecção de ameaças e intrusões é uma parte fundamental de infraestruturas de segurança, e refinar as técnicas utilizadas para isso garante que seja possível manter-se protegido de um número cada vez maior de ameaças.

Almseidin (2020) analisa diferentes modelos de aprendizado de máquina com o conjunto de dados KDD, o mesmo utilizado neste trabalho, e analisa os resultados de forma a classificar os mais adequados para cada situação.

Debar (1992) fez sua classificação de ataques utilizando redes neurais. Seu trabalho é feito com base nos padrões de comportamento de usuários, e se utiliza disso para verificar o tráfego de rede e detectar intrusões.

Ambedkar e Babu (2015) definiram regras características de cara ataque para auxiliar os algoritmos em sua classificação. Estas regras se baseiam no comportamento do tráfego de rede ocorrido quando algum destes ataques é realizado, e assim permite observar anomalias a partir disso.

Javaid, Sun, Niyaz e Alam (2016) utilizaram a técnica de *Deep Learning* para o desenvolvimento de um sistema de detecção de intrusão, tendo resultados muito positivos, em comparação a trabalhos correlatos anteriores, incluindo todas as métricas utilizadas em seu projeto.

A leitura dos trabalhos citados foi crucial para o embasamento teórico deste trabalho e esclareceu vários conceitos teóricos utilizados.

5 RESULTADOS

A partir da metodologia proposta, foram realizados os experimentos em diferentes cenários, os quais estão listados a seguir:

a) Comparação de *Log Loss* dos algoritmos iniciais a partir do modo *Explain* da biblioteca mljar.
b) O algoritmo *Xgboost*.
c) O algoritmo *Random Forest*.
d) O algoritmo *LightGBM*.
e) Um *Ensemble* comparando os algoritmos.

Utilizando o modo de explicação disponível na biblioteca mljar, foi possível observar a possível prevalência dos resultados para o algoritmo de otimização (aumento de gradiente) em relação aos algoritmos de busca e decisão tais como *Random Forest* e *Decision Tree* a partir da métrica de Log Loss apresentada na tabela a seguir:

Tabela 2 – Valores de *Log Loss*.

Valores de *Log Loss*	
Decision Tree	0,16
Xgboost	0,01
Random Forest	0,11

Fonte: Elaborada pelo autor.

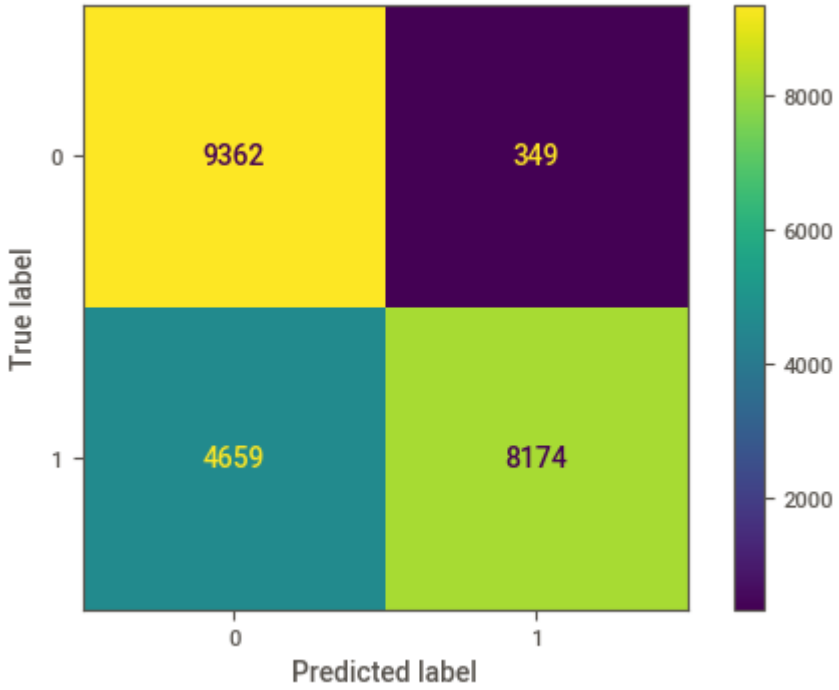
Contudo, o modo apresentado não utiliza de todos os algoritmos de aumento de gradiente, no caso faltando os testes com o algoritmo *LightGBM* e também não serve como cenário ideal para a decisão prematura do melhor algoritmo visto a métrica utilizada, que fizeram com que surgisse a necessidade de testar separadamente os algoritmos a fim de verificar a captação de melhores resultados, apresentados a seguir:

Tabela 3 – Métricas do algoritmo *Xgboost*.

Métricas *Xgboost*	
LogLoss	0,01
Acurácia	0,78
F-score	0,76
Precisão	0,96
Recall	0,64

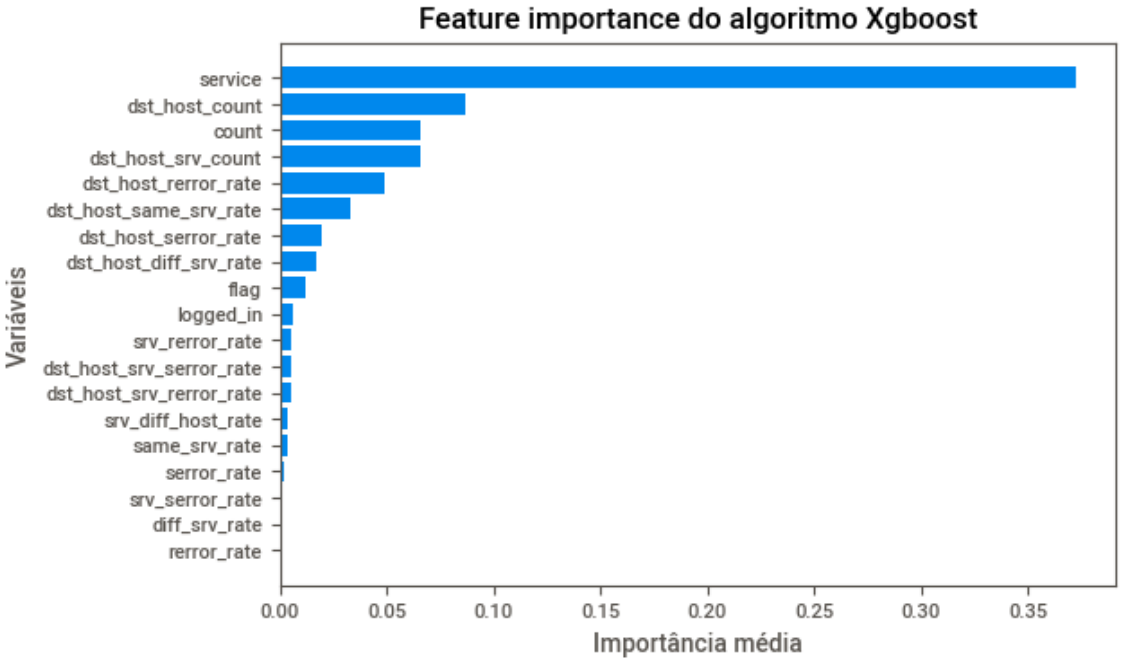
Fonte: Elaborada pelo autor.

Figura 9 – Matriz de confusão do algoritmo *Xgboost*.



Fonte: Elaborada pelo autor.

Figura 10 – *Feature Importance* do algoritmo *Xgboost*.



Fonte: Elaborada pelo autor.

Como é possível observar, as variáveis mais importantes para o modelo foram: *service, dst_host_count, count, dst_host_srv_count* e *dst_host_rerror_rate.* A variável *service* é referente ao serviço de rede utilizado pela rede de destino, e *count* se refere ao número de requisições de um mesmo *host*. A variável *dst_host_count* se refere ao número de requisições destinadas a um mesmo *host*. As outras duas variáveis: *dst_host_srv_count* e *dst_host_rerror_rate* se referem respectivamente ao número de conexões destinadas a uma mesma porta e ao número de conexões que ativaram uma *flag* de erro entre as conexões realizadas. Estas são muito importantes pois se um determinado serviço recebe inúmeras requisições de um único endereço, destinadas a apenas um *host*, numa mesma porta de rede, com um grande número de erros, há grandes chances de ser um ataque.

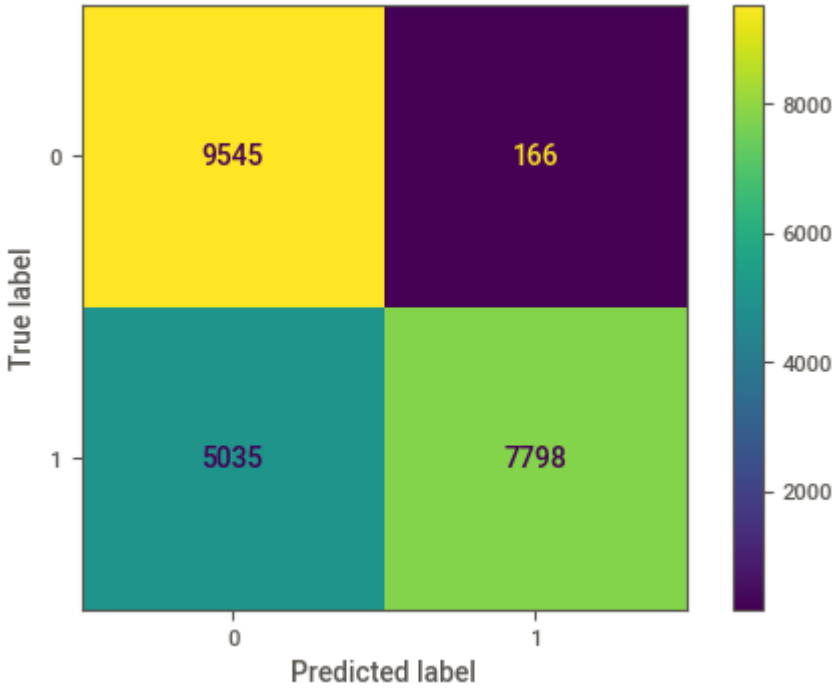
A fim de comparação, foi realizado a captação das métricas isoladas do algoritmo Random Forest apresentadas a seguir para confirmação de que sua estratégia de implementação não é a melhor neste cenário.

Tabela 4 – Métricas do algoritmo *Random Forest*.

Métricas *Random Forest*	
LogLoss	0,11
Acurácia	0,77
F-score	0,75
Precisão	0,98
Recall	0,61

Fonte: Elaborada pelo autor.

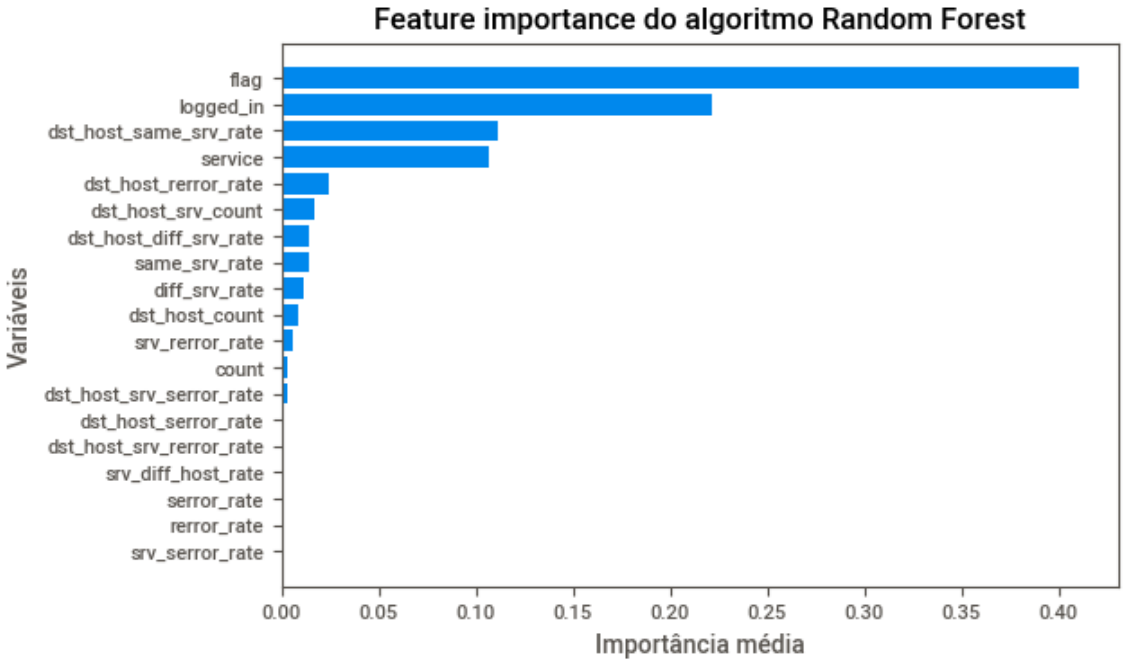
Figura 11 – Matriz de confusão do algoritmo *Random Forest*.



Fonte: Elaborada pelo autor.

Como é possível observar, ao utilizar o algoritmo de *Random Forest,* o número de falsos positivos foi reduzido, porém, há uma quantia ligeiramente maior de falsos negativos, em relação ao *Xgboost*, oposto ao objetivo de redução destes, os quais podem ser muito mais prejudiciais à infraestrutura de rede do que os falsos positivos.

Figura 12 – *Feature Importance* do algoritmo *Random Forest*.



Fonte: Elaborada pelo autor.

As variáveis selecionadas pelo modelo de *Random Forest* trazem uma interpretação diferente, já que foram utilizados principalmente *flag, logged_in, dst_host_same_srv_rate* e *service*. Elas se referem, respectivamente, a *flags* de erros na mensagem, ao estado de *login* da mensagem, ao número de conexões destinadas a um mesmo serviço de um mesmo *host* e ao serviço a qual a conexão é destinada. Se há um número grande de conexões destinadas a um mesmo serviço, de um mesmo *host,* com muitas *flags* de erro e falhas de *login*, é um sinal de que talvez esteja havendo um ataque.

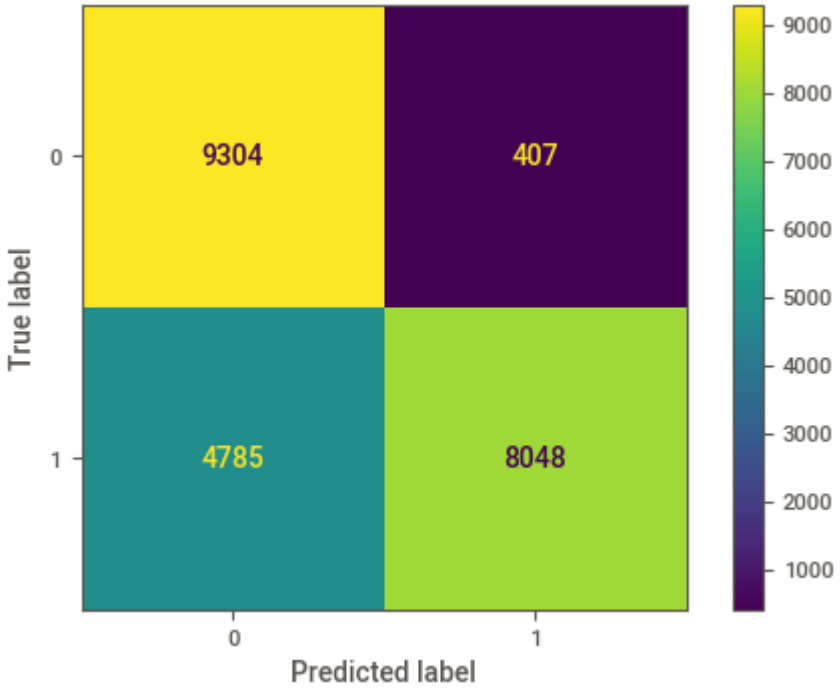
Deste modo, foi realizado o último teste isolado com o LightGBM para verificar qual dos algoritmos de aumento de gradiente apresentam melhores resultados em suas métricas.

Tabela 5 – Métricas do algoritmo *LightGBM*.

Métricas *LightGBM*	
LogLoss	0,01
Acurácia	0,77
F-score	0,76
Precisão	0,95
Recall	0,63

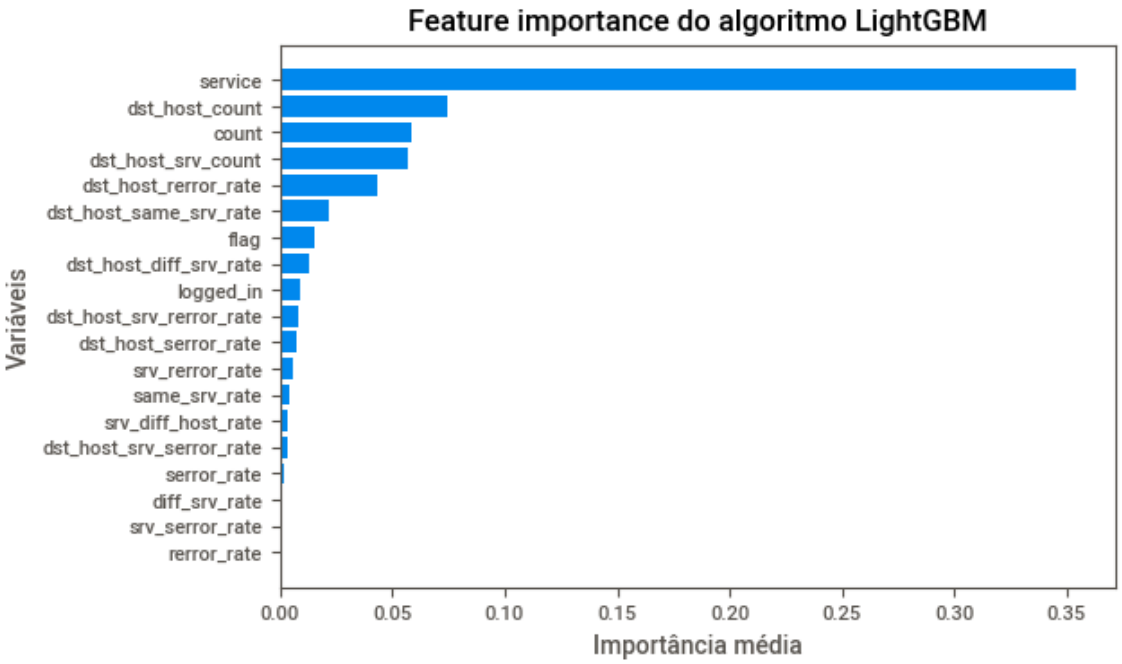
Fonte: Elaborada pelo autor.

Figura 13 – Matriz de confusão do algoritmo *Light GBM*.



Fonte: Elaborada pelo autor.

Figura 14 – *Feature Importance* do algoritmo *LightGBM*.



Fonte: Elaborada pelo autor.

Assim como no teste do *Xgboost,* as variáveis de maior importância foram *service, dst_host_count, count, dst_host_srv_count* e *dst_host_rerror_rate*. Como ambos são algoritmos de aumento de gradiente, possuem processos similares de seleção de variáveis, como estudado e descrito por Xu, Huang, Weinberger e Zheng (2014).

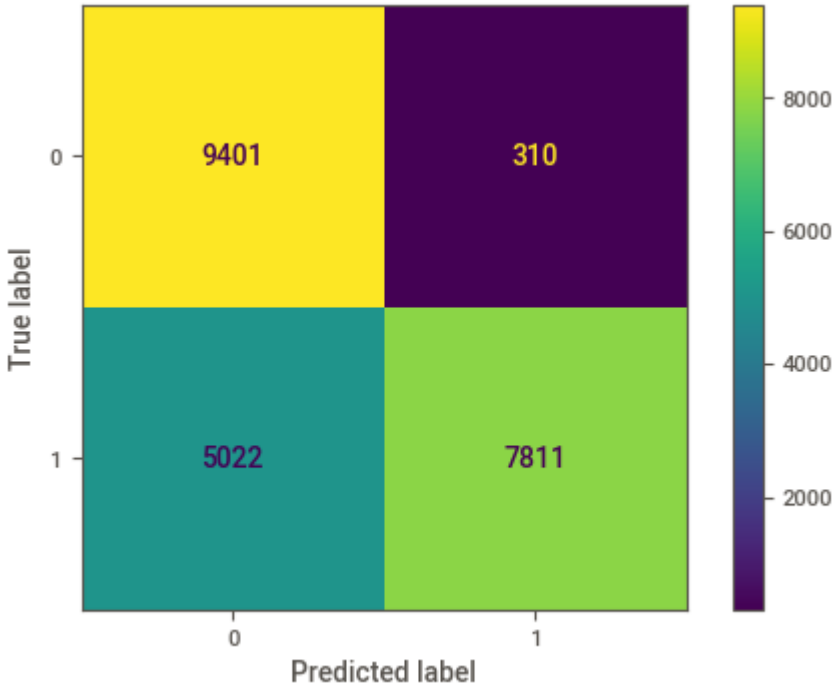
A fim de comparar os algoritmos anteriores, estes foram inseridos numa lista de algoritmos para serem votados em um *Ensemble*:

Tabela 6 – Métricas do *Ensemble*.

Métricas *Ensemble*	
LogLoss	0,01
Acurácia	0,76
F-score	0,74
Precisão	0,96
Recall	0,61

Fonte: Elaborada pelo autor.

Figura 15 – Matriz de confusão do *Ensemble*.



Fonte: Elaborada pelo autor.

As métricas *F-score* e *Recall* são os fatores decisivos entre os modelos, já que levam em consideração os valores de falsos negativos e falsos positivos, os quais representam os casos mais perigosos, quando se trata de segurança de redes. Os valores considerados na tabela 6, mostrada a seguir, correspondem às métricas em questão:

Tabela 7 – Comparação das métricas dos algoritmos testados.

Algoritmo	Métrica	Valor
Xgboost	*F-score*	0,76
	Recall	0,64
Random Forest	*F-score*	0,75
	Recall	0,61
LightGBM	*F-score*	0,76
	Recall	0,63
Ensemble	*F-score*	0,74
	Recall	0,61

Fonte: Elaborada pelo autor.

Os algoritmos citados neste capítulo, foram os que obtiveram os melhores resultados em suas predições. Embora os resultados estejam mais próximos de 75% de acerto, isso não é ideal para uma análise de algo tão sensível, tratando-se de um risco para a segurança de ativos de uma organização. O objetivo principal era a redução de falsos negativos, já que é melhor lidar com um alarme falso do que com uma invasão despercebida.

O candidato selecionado como concluinte deste trabalho é o modelo *Xgboost*, devido as métricas obtidas, as quais dependem dos números de Falsos positivos e negativos, e seu melhor desempenho em relação aos falsos negativos. O número de acertos do modelo não é alto o suficiente para a utilização em um caso real, já que o número de conexões realizadas em uma rede corporativa costuma estar na casa dos milhões, senão bilhões, diariamente.

Esse número é compreensível, já que o objetivo deste trabalho é a classificação genérica de ataques a partir do tráfego de redes, sem a definição de regras pré-definidas conhecidas, referentes a cada ataque, como é feito no trabalho de Ambedkar e Babu (2015).

6 CONCLUSÃO

A classificação de dados de rede entre tráfego saudável e anomalias é crucial para manter ambientes corporativos em pleno funcionamento e evitar vazamentos de dados que podem comprometer clientes e até mesmo a própria organização.

Devido ao grande volume de dados gerados em pouco tempo de utilização da rede, é necessária a utilização de técnicas de inteligência computacional para fazer análises efetivas do que foi coletado.

Ao utilizar técnicas de *Machine Learning* para fazer essa análise, foi possível obter resultados positivos e com porcentagens de acerto acima dos 75%. A obtenção desse resultado é devido a utilização de modelos aditivos, os quais permitiram trabalhar em cima dos dados até não ser possível mais melhorar sua precisão.

Embora os resultados sejam positivos, a porcentagem de acertos obtida ainda não é adequada para o uso em uma situação real, embora o número de falsos negativos tenha sido reduzido pelo algoritmo *Xgboost.* Para a utilização em casos reais seria necessária a implementação de outras técnicas de inteligência artificial para melhorar o número de falsos negativos do modelo, a fim de evitar que ataques passem despercebidos.

Os próximos passos para o melhoramento dos resultados seriam aplicações de técnicas mais refinadas de aprendizado de máquina combinadas com um processamento adequado dos dados para o novo método selecionado.

REFERÊNCIAS

ALMSEIDIN, Mohammad et al. Evaluation of Machine Learning Algorithms for Intrusion Detection System. 2020. 6 f. TCC (Graduação) - Curso de Tecnologia da Informação, Information Technology Department, University Of Miskolc, Miskolc, 2020.

AMBEDKAR, CH.; BABU, V. Kishore. Detection of Probe Attacks Using Machine Learning Techniques. **International Journal Of Research Studies In Computer Science And Engineering (IJRSCSE)**, Ongole, Andhra Pradesh, India, v. 2, n. 3, p. 25-29, mar. 2015.

BROADBAND SEARCH (org.). **Key Internet Statistics to Know in 2021 (Including Mobile)**. 2021. Disponível em: https://www.broadbandsearch.net/blog/internet-statistics. Acesso em: 30 abr. 2021.

BROWNLEE, Jason. **What is a Confusion Matrix in Machine Learning**. 2016. Disponível em: https://machinelearningmastery.com/confusion-matrix-machine-learning/. Acesso em: 17 out. 2021.

CLINCY, Victor; SHAHRIAR, Hossain. Web Application Firewall: network security models and configuration. In: 2018 IEEE 42ND ANNUAL COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE (COMPSAC), 42., 2018, Kennesaw. **2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC).** [S.L.]: Ieee, 2018. p. 835-836.

CHOUDHARY, Sarika; KESSWANI, Nishtha. Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 Datasets using Deep Learning in IoT. **Procedia Computer Science**, [S.L.], v. 167, n. 177, p. 1561-1573, abr. 2020. Elsevier BV. http://dx.doi.org/10.1016/j.procs.2020.03.367.

COPPOLA, Daniela. **Retail e-commerce sales worldwide from 2014 to 2024**. 2020. Disponível em: https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/. Acesso em: 30 abr. 2021.

COUNTRYMETERS (org.). **Relógio da população da África**. 2021. Disponível em: https://countrymeters.info/pt/Africa. Acesso em: 30 abr. 2021.

COUNTRYMETERS (org.). **Relógio da população da América do Norte**. 2021. Disponível em: https://countrymeters.info/pt/North_America. Acesso em: 30 abr. 2021.

CYNET. **Incident Response**. 2021. Disponível em: https://www.cynet.com/incident-response/. Acesso em: 04 maio 2021.

DEBAR, H.; BECKER, M.; SIBONI, D. A neural network component for an intrusion detection system. In: 1992 IEEE COMPUTER SOCIETY SYMPOSIUM ON RESEARCH IN SECURITY AND PRIVACY, 1992., 1992, Oakland, Ca, Usa. Proceedings 1992 IEEE Computer Society Symposium on Research in Security and Privacy. [S.L.]: Ieee Comput. Soc. Press, 1992. p. 240-250.

DEMBLA, Gaurav Dembla Gaurav. **Intuition behind Log-loss score**. 2020. Disponível em: https://towardsdatascience.com/intuition-behind-log-loss-score-4e0c9979680a. Acesso em: 17 out. 2021.

DENG, Li. Deep Learning: methods and applications. **Foundations And Trends® In Signal Processing**, [S.L.], v. 7, n. 3-4, p. 197-387, 30 jun. 2014. Now Publishers. http://dx.doi.org/10.1561/2000000039.

DIETTERICH, Thomas G. Ensemble Methods in Machine Learning. **Multiple Classifier Systems**, [S.L.], v. 1857, n. 1, p. 1-15, dez. 2000. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/3-540-45014-9_1.

EQUIPE TOTVS. **Deep Learning**: conheça o conceito e suas aplicações. Conheça o conceito e suas aplicações. 2020. Disponível em: https://www.totvs.com/blog/inovacoes/deep-learning/. Acesso em: 12 set. 2021.

FRIEDMAN, Jerome H. Stochastic gradient boosting. **Computational Statistics & Data Analysis**, [S.L.], v. 38, n. 4, p. 367-378, fev. 2002. Elsevier BV. http://dx.doi.org/10.1016/s0167-9473(01)00065-2.

GOMES, Elias Amadeu de Souza. **APLICABILIDADE DE ALGORITMOS DE APRENDIZADO DE MÁQUINAPARA DETECÇÃO DE INTRUSÃO E ANÁLISEDE ANOMALIAS DE REDE**. 2019. 58 f. Dissertação (Mestrado) - Curso de Especialização em Informática, Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Brasília, 2019.

HUILGOL, Purva. **Accuracy vs. F1-Score**. 2019. Disponível em: https://medium.com/analytics-vidhya/accuracy-vs-f1-score-6258237beca2. Acesso em: 20 out. 2021.

IBM. **Machine Learning**. 2020. Disponível em: https://www.ibm.com/cloud/learn/machine-learning. Acesso em: 13 set. 2021.

JAVAID, Ahmad y; SUN, Weiqing; NIYAZ, Quamar; ALAM, Mansoor. A Deep Learning Approach for Network Intrusion Detection System. In: 9TH EAI INTERNATIONAL CONFERENCE ON BIO-INSPIRED INFORMATION AND COMMUNICATIONS TECHNOLOGIES, 9., 2016, New York. **Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS).** [S.L.]: Acm, 2016. p. 21-26

JEONG, Jaehoon (Paul); HONG, Dongjin; KIM, Jinyoug; HYUN, Daeyoung. **SDN-based Network Security Functions for Effective DDoS Attack Mitigation**. 2017. 6 f. Dissertação (Mestrado) - Curso de Master’s Degree Program For Information Security, Department Of Computer Science & Engineering, Sungkyunkwan University, Sungkyunkwan, 2017.

LALONDE, Carole; BOIRAL, Olivier. Managing risks through ISO 31000: a critical analysis. **Risk Management**, [S.L.], v. 14, n. 4, p. 272-300, nov. 2012. Springer Science and Business Media LLC. http://dx.doi.org/10.1057/rm.2012.9.

LEOBONS, Rodrigo Maestrelli. **Detecção de Intrusos**. 2021. Disponível em: https://www.gta.ufrj.br/grad/07_2/rodrigo_leobons/deteccao.html. Acesso em: 12 set. 2021.

MARBLESTONE, Adam H.; WAYNE, Greg; KORDING, Konrad P.. Toward an Integration of Deep Learning and Neuroscience. **Frontiers In Computational Neuroscience**, [S.L.], v. 10, n. 94, p. 1-41, 14 set. 2016. Frontiers Media SA. http://dx.doi.org/10.3389/fncom.2016.00094.

MIDWEST ARTIFICIAL INTELLIGENCE AND COGNITIVE SCIENCE CONFERENCE, 2011, Cincinnati, Ohio. **Proceedings of the Twentysecond Midwest Artificial Intelligence and Cognitive Science Conference.** Madison, Wisconsin: Omnipress, 2011.

MLITZ, Kimberly. **Current and planned usage of public cloud platform services running applications worldwide from 2020 to 2021**. 2021. Disponível em: https://www.statista.com/statistics/511467/worldwide-survey-public-coud-services-running-application/. Acesso em: 30 abr. 2021.

OPITZ, David; MACLIN, Richard. Popular Ensemble Methods: an empirical study. **Journal Of Artificial Intelligence Research**, El Segundo, Ca, v. 11, n. 4, p. 169-198, ago. 1999.

ORACLE. **O que é Big Data?** Disponível em: https://www.oracle.com/br/big-data/what-is-big-data/. Acesso em: 12 set. 2021

SALESFORCE. **Inteligência Artificial**: o que é? Disponível em: https://www.salesforce.com/br/products/einstein/ai-deep-dive/. Acesso em: 12 set. 2021.

SANCHES, Mariana. **Covid-19**: anthony fauci diz que brasil deve considerar seriamente fazer lockdown. Anthony Fauci diz que Brasil deve considerar seriamente fazer lockdown. 2021. Disponível em: https://www.bbc.com/portuguese/internacional-56655828. Acesso em: 04 maio 2021.

SCIKIT-LEARN. **Confusion Matrix**. 2021. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html. Acesso em: 20 out. 2021.

SHAIKH, Rahil. **Feature Selection Techniques in Machine Learning with Python**. 2018. Disponível em: https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e. Acesso em: 17 out. 2021.

SINGH, Harshdeep. **Understanding Gradient Boosting Machines**. 2018. Disponível em: https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab. Acesso em: 05 out. 2021.

SUGIYAMA, Shigeki. **Human Behavior and Another Kind in Consciousness**: emerging research and opportunities. Japão: Igi Global, 2019.

TAMIR, Dr. Michael. **What Is Machine Learning?** 2020. Disponível em: https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/. Acesso em: 13 set. 2021.

TAVALLAEE, Mahbod; BAGHERI, Ebrahim; LU, Wei; GHORBANI, Ali A.. A detailed analysis of the KDD CUP 99 data set. In: 2009 IEEE SYMPOSIUM ON COMPUTATIONAL INTELLIGENCE FOR SECURITY AND DEFENSE APPLICATIONS (CISDA), 2., 2009, Ottawa. **2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications.** [S.L.]: Ieee, 2009. p. 1-6.

TODOROV, Georgi. **65 Artificial Intelligence Statistics for 2021 and Beyond**. 2021. Disponível em: https://www.semrush.com/blog/artificial-intelligence-stats/. Acesso em: 11 set. 2021.

UFRJ. **Diferenças entre IDS e IPS**. Disponível em: https://www.gta.ufrj.br/grad/16_2/2016IDS/conceituacao.html. Acesso em: 12 set. 2021.

UNSUPERVISED AND SEMI-SUPERVISED LEARNING. Conway, Arkansas, Usa: Springer, 2020.

UOL (São Paulo) (org.). **Outra globalização é possível depois da covid, segundo economistas**. 2021. Disponível em: https://noticias.uol.com.br/ultimas-noticias/afp/2021/03/22/outra-globalizacao-e-possivel-depois-da-covid-segundo-economistas.htm. Acesso em: 04 maio 2021.

VOVK, Vladimir. The Fundamental Nature of the Log Loss Function. **Fields Of Logic And Computation II**, [S.L.], v. 9300, p. 307-318, set. 2015. Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-23534-9_20.

WOOD, Thomas. **F-Score**. Disponível em: https://deepai.org/machine-learning-glossary-and-terms/f-score. Acesso em: 18 out. 2021.

XU, Zhixiang; HUANG, Gao; WEINBERGER, Kilian Q.; ZHENG, Alice X.. Gradient boosted feature selection. **Proceedings Of The 20Th Acm Sigkdd International Conference On Knowledge Discovery And Data Mining**, [S.L.], v. 20, n. 11, p. 522-531, 24 ago. 2014. ACM. http://dx.doi.org/10.1145/2623330.2623635.

ZIEN, Alexander; KRÄMER, Nicole; SONNENBURG, Sören; RÄTSCH, Gunnar. The Feature Importance Ranking Measure. **Machine Learning And Knowledge Discovery In Databases**, [S.L.], v. 5782, n. 46, p. 694-709, ago. 2009. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-04174-7_45.

ZIZU. **Dartmouth Workshop**: the birthplace of ai. The Birthplace Of AI. 2018. Disponível em: https://medium.com/rla-academy/dartmouth-workshop-the-birthplace-of-ai-34c533afe992. Acesso em: 11 set. 2021.