DETECÇÃO DE ATAQUES DDOS POR MEIO DE APRENDIZADO DE MÁQUINA

Arthur Lucas Rodrigues
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP Pirituba), Pirituba, SP, Brasil.

Henrique Fernandes Casagrande
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP Pirituba), Pirituba, SP, Brasil.

João Paulo Mascena Gomes
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP Pirituba), Pirituba, SP, Brasil.

Pedro Luiz Agrella Passos
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP Pirituba), Pirituba, SP, Brasil.

Prof. Adriano Jose Ferruzzi
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP Pirituba), Pirituba, SP, Brasil.

Prof. Regivaldo Sousa Ferreira
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP Pirituba), Pirituba, SP, Brasil.

RESUMO: Juntamente com a expansão do mundo cibernético, houve o aumento das ameaças que podem ocorrer no ambiente virtual. Esses ataques podem ter os mais variados objetivos e um dos ataques mais comuns visa a negação de serviços (*Denial of Service* - DoS). Esse tipo de ataque pode ser feito de forma automatizada e

distribuída (*Distributed Denial of Service* - DDoS). Ao mesmo tempo, esses tipos de ataques automatizados podem ser identificados por modelos matemáticos devidamente treinados, pois eles obedecem a um certo padrão de comportamento. Portanto, este trabalho utiliza o aprendizado de máquina com o objetivo de detectar ataques DDoS. Os modelos foram construídos a partir de quatro algoritmos de aprendizado de máquina: o *Naive Bayes*, o SVM (*Support Vector Machine*), a Regressão Logística e o *Random Forest*. Todos foram utilizados com o objetivo de encontrar o algoritmo com o melhor resultado possível. Finalmente, ferramentas de xAI (*Explainable Artificial Intelligence*) foram utilizadas para garantir a explicabilidade e compreensão dos modelos.

PALAVRAS-CHAVE: Aprendizado de máquina. Ataques DDoS e DoS. Explicabilidade.

ABSTRACT: Along with the expansion of the cyber world, there has been an increase in the threats that can occur in the virtual environment. These attacks can have the most varied objectives and one of the most common is denial of service (DoS). This type of attack can be automated and distributed (*Distributed Denial of Service* - DDoS). At the same time, these types of automated attacks can be identified by properly trained mathematical models, as they follow a certain pattern of behavior. This work therefore uses machine learning to detect DDoS attacks. The models were built using four machine learning algorithms: *Naive Bayes*, SVM (*Support Vector Machine*), *Logistic Regression* and *Random Forest*. All were used with the aim of finding the algorithm with the best possible result. Finally, xAI (*Explainable Artificial Intelligence*) tools were used to ensure that the models were explainable and understandable.

KEYWORDS: Machine learning. DDoS and DoS attacks. Explainability.

APRESENTAÇÃO DO TEMA

Ataques cibernéticos crescem à medida que pessoas, empresas e instituições estão tornando seus processos e serviços *online*. Esses ataques podem ter os mais

variados objetivos e um dos ataques mais comuns visa a negação de serviços (*Denial of Service* - DoS). Esse tipo de ataque pode ser feito de forma automatizada e também de forma distribuída (*Distributed Denial of Service* - DDoS) (O QUE SÃO ATAQUES DE DDOS?, 2023). Um exemplo recente é a empresa francesa OVN e o provedor de nomes DYN, os mesmos sofreram diversos ataques de DDoS (PELLOSO *et al.*, 2018). Outro exemplo é relatório de segurança da Microsoft o qual afirma que os ataques DDoS baseados em TCP continuam sendo o vetor de ataque mais comum em serviços web, compreendendo 63% de todo o tráfego de ataque de 2022 (MICROSOFT, 2023)

Esses ataques podem ser classificados em diversos tipos, como ataques baseados em volume, ataques de protocolo e ataques da camada de aplicação. O ataque baseado em volume tem o objetivo de saturar a banda de rede da vítima. O ataque baseado em protocolo visa causar a exaustão de firewalls ou balanceadores de carga. Por sua vez, o ataque na camada de aplicação tem como objetivo acabar com os recursos de uma aplicação web específica. Uma vez entendido como o ataque DDoS funciona, é possível pensar em estratégias de defesa como a detecção dos ataques (SOMANI *et al.*, 2017).

Nesse sentido, o projeto foca precisamente na detecção de possíveis ataques *SYN Flood*. Trata-se de um tipo de ataque DDoS que aproveita do *HandShake* do protocolo TCP, no qual os atacantes inundam o servidor com requisições (*SYN*) (FOX, 2019). O trabalho tem este foco específico com o objetivo de trazer resultados mais precisos.

Para fazer as detecções são utilizadas técnicas de aprendizado de máquina. Segundo Valdati (2020), as técnicas de aprendizado de máquina são indicadas para trabalhar com problemas não determinísticos. Essas técnicas podem ser classificadas em aprendizado supervisionado, não supervisionado e por reforço. O projeto utiliza o aprendizado supervisionado para classificar os ataques. Nesse caso, o algoritmo de aprendizado recebe valores que guiam a criação do modelo matemático destinado a realizar a classificação de um valor de número contínuo. A partir da criação do modelo, torna-se possível classificar novos dados. Dessa forma, o projeto desenvolve

algoritmos de aprendizado de máquina com o objetivo de classificar potenciais ataques DDoS.

PROPOSTA DE PESQUISA E JUSTIFICATIVA

Com o crescimento da hospedagem de serviços e processos online, cada vez mais ataques cibernéticos são criados para indisponibilizar, expor e até mesmo roubar dados desses serviços. Existem casos recentes que comprovam o quanto essas ameaças e ataques impactam a infraestrutura da tecnologia da informação das empresas e instituições, por exemplo, recentemente, houve uma indisponibilidade temporária da plataforma “ChatGPT”. A empresa OpenAI, a criadora dessa plataforma, no dia 08 de novembro de 2023, registrou em seu relatório de incidentes, interrupções periódicas em seus serviços devido a um tráfego que remete a um ataque DDoS (OPENAI STATUS, 2023). Esse caso demonstra como técnicas maliciosas podem interromper os serviços fornecidos na internet. Os motivos de cada ataque são amplos e diversificados, portanto, não é possível evitar qualquer tipo de ameaça sem uma tomada de ação contra elas.

Por outro lado, especialistas em cibersegurança continuamente desenvolvem ferramentas e técnicas com a intenção de criar ou até mesmo fortalecer barreiras que impeçam ameaças e ataques cibernéticos.

Portanto, a proposta desse trabalho se justifica em contribuir com o desenvolvimento de ferramentas de cibersegurança, através da criação de novos modelos de detecção de ataques DDoS.

DESAFIOS

A classificação de um ataque DDoS a partir do aprendizado de máquina enfrenta alguns obstáculos. Esses obstáculos partem principalmente do tratamento dos dados, ou seja, organizar os dados de uma forma que a máquina consiga aprender sem nenhum tipo de interferência em sua interpretação. Essa tarefa se torna difícil, tendo em vista a premissa de um ataque DDoS que é justamente enviar um grande volume de requisições para algum alvo, e essa quantidade alta de informações torna o processo de filtragem dos dados em uma tarefa minuciosa e trabalhosa. Para

chegar ao entendimento da importância dessa etapa, foi necessário realizar um levantamento bibliográfico com o propósito de pesquisar quais são os melhores algoritmos de aprendizado de máquina para fazer a classificação de ataques DDoS e quais variáveis conduzem a uma melhor interpretação da máquina a partir do cenário proposto, além de entender quais dados se tornam indesejáveis para uma categorização eficaz.

Outro desafio enfrentado, é detectar esses ataques DDoS com a expectativa de evitar o falso positivo, ou seja, o aprendizado de máquina também deve ter como o objetivo ter uma boa precisão em relação aos acessos legítimos.

OBJETIVO GERAL

Desenvolver um algoritmo que utilize aprendizado de máquina para identificar possíveis ataques de DoS e de DDoS, mais especificamente, os ataques do tipo *SYN*.

OBJETIVOS ESPECÍFICOS

- Estudar a fundamentação teórica para compreensão do problema;
- Encontrar uma base de dados pública com registros de ataques DDoS;
- Fazer uma análise exploratória para identificar as características do conjunto de dados antes de que ele seja submetido para a criação do modelo e, assim, filtrá-los;
- Desenvolver e testar um modelo de aprendizado de máquina que permite identificar ataques de DDoS;
- Aplicar explicabilidade no modelo que obteve o melhor resultado para melhor compreensão dos resultados.

METODOLOGIA

Com os objetivos e propostas de trabalho definidos, é importante realizar um estudo da metodologia. Como resultado desse estudo, o trabalho foi dividido nas seguintes etapas:

- Levantamento do estado da arte e revisão bibliográfica;
- Busca pela base de dados com registros de ataques DDoS;

- Proposta de estratégia criada a partir da revisão bibliográfica;
- Desenvolver os algoritmos dos modelos de aprendizado de máquina;
- Testar, a partir das métricas de avaliação, os modelos já treinados;
- Aplicar explicabilidade nos modelos para garantir a compreensão das decisões.

DESENVOLVIMENTO DO TRABALHO

Este estudo é pautado em pesquisas teóricas e práticas, feitas por meio de revisão de literatura e levantamento bibliográfico de natureza qualitativa. Ele foi realizado com base em artigos com pesquisas semelhantes a este projeto. Para encontrá-los, foi utilizado a base Período Capes e a ferramenta de pesquisa de artigos científicos Google Acadêmico.

A respeito da seleção dos artigos, foram utilizadas as seguintes palavras-chave portal Periódicos Capes e no Google Acadêmico: DoS, DDoS, *machine learning* e aprendizado de máquina. Sobre os critérios de inclusão, foram utilizados artigos nos idiomas de português e inglês relacionados ao tema, e também os artigos que tiveram proximidade com o contexto deste estudo. Em relação aos critérios de exclusão, foram desconsiderados os artigos que não estiveram nos idiomas determinados e que não trouxessem a temática base para esta pesquisa.

Para a criação do algoritmo, foi necessário utilizar as bases de dados com registros de ataques DoS e DDoS com o objetivo de realizar o treinamento do algoritmo. A base de dados utilizada foi a CIC-DDoS 2019, criada e publicada pelo *Canadian Institute for Cyber Security* (CIC) (DDOS 2019 | DATASETS | RESEARCH | CANADIAN INSTITUTE FOR CYBERSECURITY | UNB, 2019). Essa base fornece um conjunto de dados que inclui registros de vários tipos de ataques DDoS.

O armazenamento da base de dados selecionada foi feito em pastas no *Google Drive*. Para a construção, treinamento e testagem do algoritmo, foi utilizado o *Google Colab*. Essa plataforma foi escolhida, pois fornece uma interface *Jupyter Notebook* gratuita. A interface tem como *backend* uma máquina virtual que trabalha como nó computacional, possuindo 2 núcleos Intel Xeon de 2.20GHz, 12 GB de memória RAM. Além disso, ela dá suporte a bibliotecas como *Scikit Learn* e *Tensor Flow*, utilizada no desenvolvimento de algoritmos de aprendizado de máquina.

Neste projeto são utilizadas as portas 20, 21, 22, 25, 53, 80, 443, 465, 587. Essas portas têm grande importância para ambientes de infraestrutura de tecnologia, pois representam importantes serviços de rede como FTP (*File Transfer Protocol*), SSH (*Secure Shell*), SMTP (*Simple Mail Transfer Protocol*), DNS (*Domain Name System*), HTTP (*Hypertext Transfer Protocol*), HTTPS (*Hypertext Transfer Protocol Secure*) e SMTP com criptografia SSL (*Secure Sockets Layer*) ou TLS (*Transport Layer Security*), respectivamente.

Depois da escolha das portas, foram estudados os tipos de algoritmos de aprendizado de máquina que poderiam contribuir com a detecção dos ataques e classificação de acessos legítimos. De acordo com Ludermir (2021), existem modos no qual você insere esses dados e treina os modelos. Esses modos são classificados em três tipos de aprendizado: aprendizado de máquina supervisionado, aprendizado de máquina não supervisionado e aprendizado por reforço.

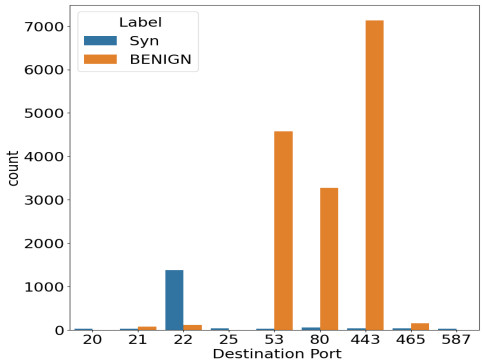
O aprendizado supervisionado tem o objetivo de treinar o modelo com dados rotulados. Essa supervisão ocorre quando há uma inserção de dados que possuem rotulação, ou seja, identificações de cada registro. A principal importância dele é garantir que a máquina entenda como os dados são caracterizados, isso possibilita classificações e hipóteses mais precisas.

Já o aprendizado de máquina não supervisionado tem como objetivo fazer com que a máquina aprenda de maneira independente. Isso ocorre quando há uma inserção de dados que não possuem rotulações ou categorizações. A partir desses dados, a máquina classificará, sem supervisão, esses mesmos dados. Como não há nenhum tipo de rotulação nesses dados, a máquina fará o agrupamento dos dados a partir das semelhanças entre eles.

Por último, há também a existência de um aprendizado de máquina baseado em reforço. Esse aprendizado tem o objetivo de fazer com que a máquina aprenda a partir de reforços, ou seja, a cada hipótese que a máquina faz, pode ser retornada uma recompensa ou uma punição. Se a hipótese for correta é retornada uma recompensa e, se a hipótese for errada, é retornada uma punição. A máquina vai fortalecer seu aprendizado a partir de cada retorno.

Após os estudos, foi escolhido o aprendizado supervisionado para o desenvolvimento do algoritmo, pois a base de dados escolhida possui rótulos que identificam os registros legítimos e de ataque como *benign* e *syn*, respectivamente. A Figura 1 apresenta os registros de cada uma das portas escolhidas.

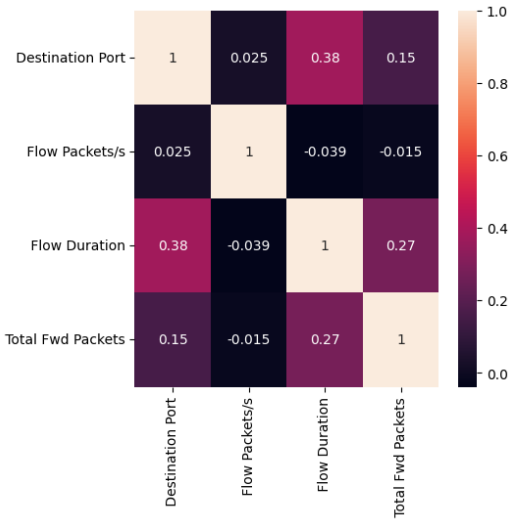
Figura 1 - Registros de ataques e acessos benignos



Fonte: Imagem dos autores (2023).

Durante a criação dos algoritmos, foi necessário a elaboração de três etapas: análise exploratória dos dados, tratamento de dados e, finalmente, o desenvolvimento dos algoritmos de aprendizado de máquina. Para isso, pode-se utilizar de várias técnicas de aprendizado de máquina, sendo que esta pesquisa dá destaque a um tipo de algoritmo: o *Random Forest,* um modelo que trouxe resultados satisfatórios no trabalho de (SILVA *et al*., 2020).

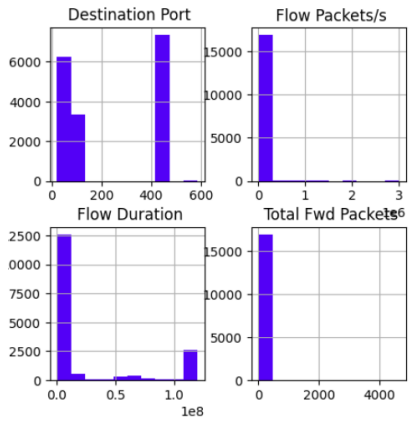
Figura 2 - Gráfico de correlação



Fonte: Imagem dos autores (2023).

A análise efetiva desses dados tornou-se possível após o processo de tratamento dos dados, no qual foi utilizada a biblioteca *Matplotlib* para representar graficamente as análises realizadas. A primeira visualização foi o gráfico de correlação, cujo objetivo principal é identificar as relações entre as variáveis. Nesse tipo de gráfico, os valores variam de -1.0 a 1.0, onde 1.0 indica uma forte correlação positiva, -1.0 indica uma forte correlação negativa, e 0.0 ou valores próximos indicam ausência de correlação.

Figura 3 - Histograma



Fonte: Imagem dos autores (2023).

Analisando o histograma na Figura 3, destaca-se uma significativa correlação entre a variável "*Flow Duration*" e "*Total Fwd Packets*". Essa forte associação sugere que o aumento da duração do fluxo está diretamente relacionado ao aumento no número total de pacotes direcionados para frente. Além disso, a Figura 2 revela uma baixa correlação entre as variáveis "*Destination Port*" e "*Flow Packets/s*". Isso indica que as mudanças na porta de destino não têm uma relação clara com a taxa de pacotes por segundo no fluxo.

Essas observações fornecem *insights* valiosos sobre as interações entre as variáveis estudadas, contribuindo para uma compreensão mais profunda do conjunto de dados. Esse tipo de análise é crucial para identificar padrões, tendências e possíveis *insights* que podem orientar decisões e ações futuras.

Durante esse projeto foi necessário utilizar o *one-hot-encoding*, trata-se de uma técnica de pré-processamento de dados que cria colunas, utilizando como base alguma variável de valores categóricos. Essas colunas mostram a presença ou não daquele determinado valor em um determinado dado. Essa técnica pode ser usada para uma melhor compreensão de algum modelo de aprendizado, já que ela é capaz de transformar os valores em números binários e deixá-los com o mesmo peso no momento da aprendizagem. No caso deste trabalho ele foi utilizado para a conversão da variável “*Destination Port*”.

Figura 4 - Aplicação do *One Hot Encoding*

	0	1	2	3	4	5	6	7	8	Flow Packets/s	Flow Duration	Total Fwd Packets
0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.509937	31376453	4
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	40816.326531	49	2
2	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.206148	38807050	8
3	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	25157.232704	159	2
4	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.300611	33265536	6
...	...	...	...	...	...	...	...	...	...	...	...	...
16971	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	141.008919	28367	1
16972	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	5115.089514	782	1
16973	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	87.062511	45944	2
16974	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	3.880933	5411070	12
16975	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	38961.038961	77	1

16976 rows × 12 columns

Fonte: Imagem dos autores (2023).

Para o treinamento do algoritmo é importante ter um balanceamento entre os dados benignos e malignos. Sabendo dessa informação, foi utilizado o método *over sampling* e *under sampling* (LIMA, 2020). Os dados continham 15327 registros legítimos e 1649 registros malignos, após o balanceamento ambos os casos estão com 1649 registros.

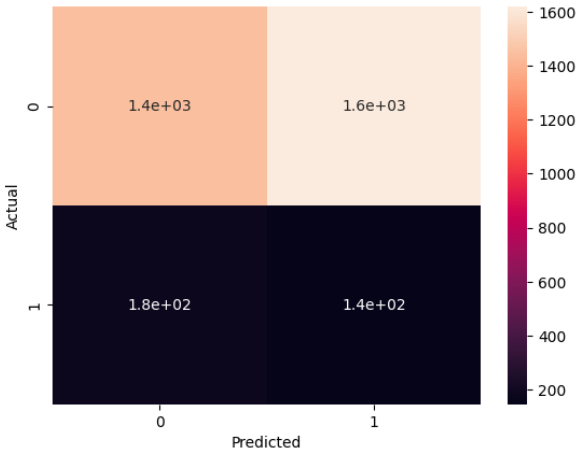
Depois do desenvolvimento do *machine learning*, foi utilizado uma biblioteca chamada SHAP. Ela serve para criação de gráficos sobre explicabilidade, ou seja, ela mostra a importância de cada variável durante a tomada de decisão do modelo.

RESULTADOS E DISCUSSÃO

A partir de um levantamento bibliográfico em busca de modelos de aprendizado de máquina para a classificação de ataques DDoS, foram encontrados quatro modelos: *Naive Bayes* e *Random Forest* de acordo com Figueiredo et al. (2022), Regressão logística de acordo com Clarindo e Silva (2022) e *Support Vector Machine* de acordo com Sarraf (2020).

Ao utilizar os algoritmos de aprendizado selecionados, foi descoberto que o resultado obtido do modelo *Random Forest* obteve o melhor desempenho.

Figura 5 - Matriz de confusão do *Naive Bayes*



Fonte: Imagem dos autores (2023).

Figura 6 - Métricas de avaliação do *Naive Bayes*

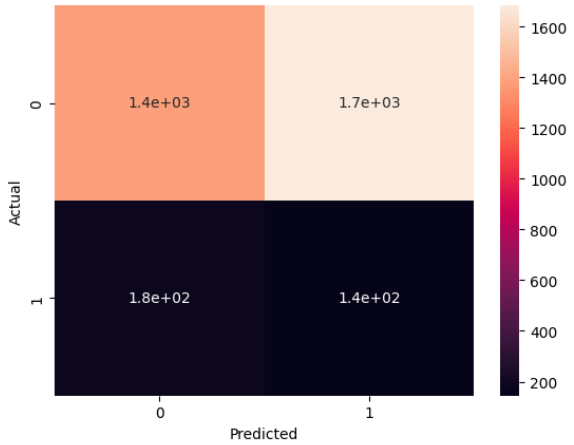
	precision	recall	f1-score
0	0.89	0.47	0.62
1	0.08	0.44	0.14

Fonte: Imagem dos autores (2023).

Inicialmente, o *Naive Bayes* resultou em uma precisão de 89% para acessos benignos, porém em relação a precisão de acessos malignos obteve apenas 8%. Além disso, obteve 47% de acurácia.

No caso da regressão logística, os resultados foram semelhantes com o *Naive Bayes*. Ele apresentou uma precisão de 8% para acessos malignos e 88% para acessos benignos. Resultando assim, uma acurácia de 45%. Conforme apresentado na Figura 7 e Figura 8

Figura 7 - Matriz de confusão da Regressão Logística



Fonte: Imagem dos autores (2023).

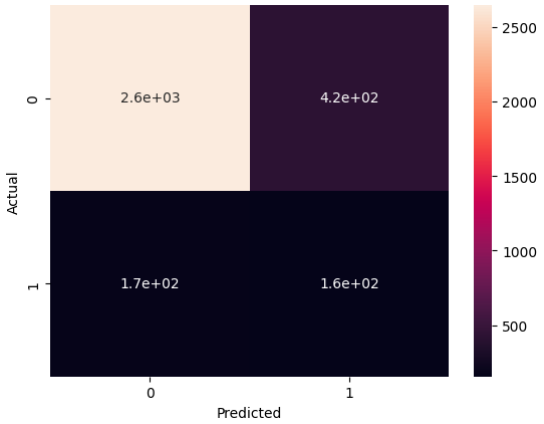
Figura 8 - Métricas de avaliação da Regressão Logística

	precision	recall	f1-score
0	0.88	0.45	0.60
1	0.08	0.44	0.13

Fonte: Imagem dos autores (2023).

O *Support Vector Machine* obteve um resultado de 94% de precisão para acesso legítimos, sendo considerado um bom resultado, porém, acertou apenas 48% dos acessos não legítimos, obtendo uma acurácia de 83%.

Figura 9 - Matriz de confusão do *Support Vector Machine*



Fonte: Imagem dos autores (2023).

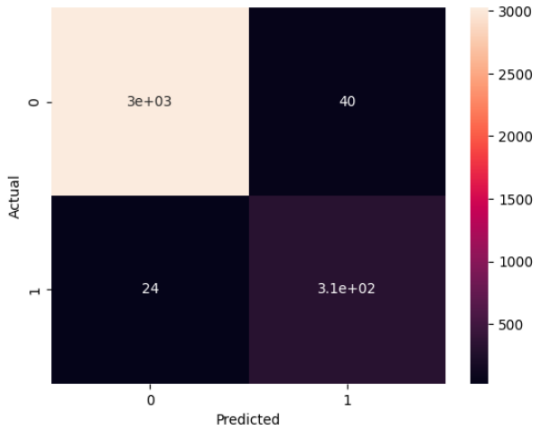
Figura 10 - Métricas de avaliação do *Support Vector Machine*

	precision	recall	f1-score
0	0.94	0.86	0.90
1	0.27	0.48	0.35

Fonte: Imagem dos autores (2023).

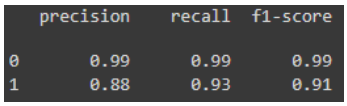
Assim como encontrado nas pesquisas, o *Random Forest* alcançou os melhores resultados. O modelo apresentou 99% de precisão em relação aos acessos legítimos e 88% de precisão para acessos não legítimos, o que mostra um bom resultado, pois o modelo também prioriza acertar quando os acessos são legítimos.

Figura 11 - Matriz de confusão do *Random Forest*



Fonte: Imagem dos autores (2023).

Figura 12 - Métricas de avaliação do *Random Forest*



	precision	recall	f1-score
0	0.99	0.99	0.99
1	0.88	0.93	0.91

Fonte: Imagem dos autores (2023).

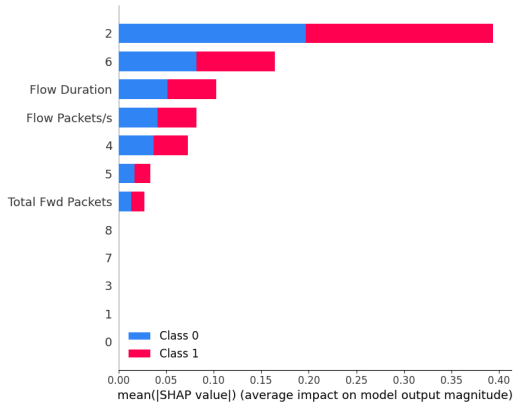
Antes de chegar neste resultado, o algoritmo sofreu *overfitting*. Esse problema foi consertado a partir de ajustes nos hiperparâmetros do aprendizado de máquina.

Inicialmente, vale destacar que o *random forest* é um algoritmo que elabora diversas árvores de decisão. No começo desta criação são selecionadas amostras aleatórias do banco de dados. Para a escolha da variável utilizada no parâmetro de cada nó das árvores, o algoritmo escolhe duas ou mais variáveis entre todas as outras e depois faz um cálculo para determinar qual variável será escolhida no primeiro nó com base nas amostras. No segundo nó, é escolhido uma variável aleatória, exceto a já escolhida no primeiro nó, e assim sucessivamente (Rigatti, 2017).

Sabendo dessas características, o aumento do número de árvores de decisões e o aumento de amostras mínimas que um nó deve conter após se dividir podem evitar o *overfitting*. Para determinar esses aumentos foram utilizados os hiperparâmetros *n_estimators* e o *min_samples_leaf* respectivamente. O primeiro houve um aumento para 120 árvores e o segundo houve um aumento para 200 amostras.

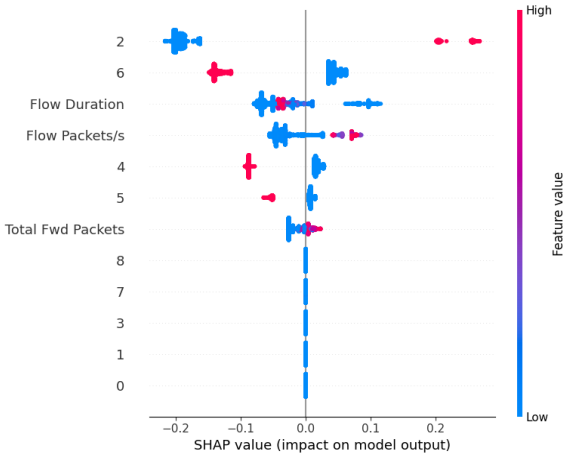
Uma vez terminado o desenvolvimento dos algoritmos, foi utilizada a biblioteca SHAP para explicar os resultados obtidos pelo modelo com maior precisão. A biblioteca SHAP busca os valores utilizados pelo preditor em cada tomada de decisão e cria gráficos com as métricas de importância de cada variável (WELCOME TO THE SHAP DOCUMENTATION — SHAP LATEST DOCUMENTATION, 2018).

Figura 13 - *Summary plot 1*



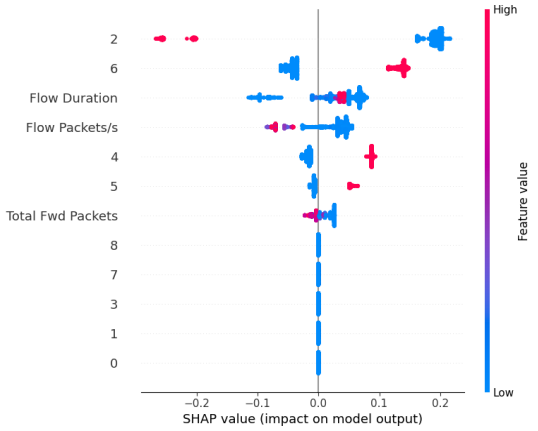
Fonte: Imagem dos autores (2023).

Figura 14 - *Summary plot 2*



Fonte: Imagem dos autores (2023).

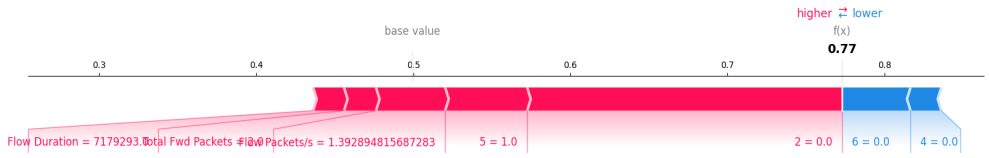
Figura 15 - *Summary plot 3*



Fonte: Imagem dos autores (2023).

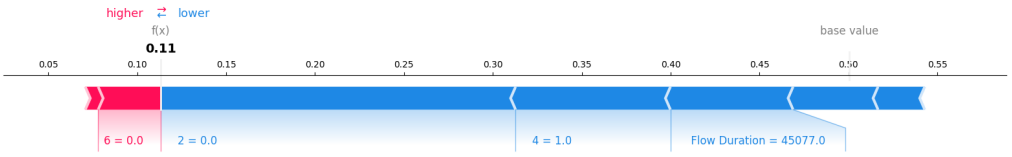
Observando os gráficos da Figura 13, Figura 14 e Figura 15 é possível compreender que as variáveis categóricas das portas de destino 22, 53, 80 e 443 foram de crucial importância, mas principalmente as portas 22 e 443. As portas 22, 443, 53 e 80 estão representadas pelos campos 2, 6, 4 e 5, respectivamente nos gráficos. A relevância deles para a decisão do algoritmo é comprovada na Figura 1 que mostra a relação da porta com a quantidade de acessos legítimos e não legítimos antes do balanceamento dos dados. Importante ressaltar que o balanceamento só afetou o número de acessos legítimos, isso não acarreta grandes mudanças para o treinamento neste caso.

Figura 16 - *Plot Force 1*



Fonte: Imagem dos autores (2023).

Figura 17 - *Plot force 2*

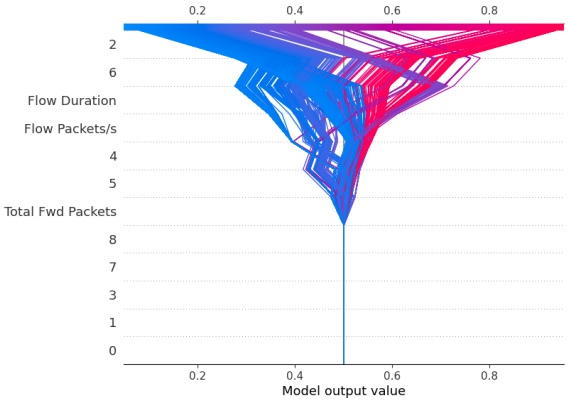


Fonte: Imagem dos autores (2023).

Os gráficos do tipo *Plot Force* apresentados na Figura 16 e Figura 17, permitem identificar como as variáveis "*Flow Duration*", "*Flow Packets/s*" e "*Total Fwd Packets*" contribuíram para a previsão do modelo em cada observação específica. A seleção dessas variáveis apresenta coerência, já que o ataque DDoS é baseado no fluxo de pacotes.

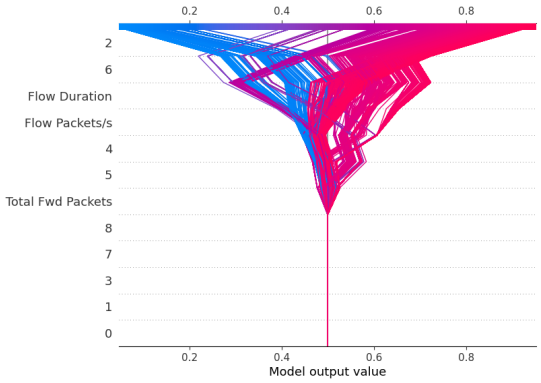
Os gráficos de decisão SHAP mostram como modelos complexos chegam às suas previsões, nele, pode ser observado o valor de importância e o caminho percorrido em cada tomada de decisão (DECISION PLOT — SHAP LATEST DOCUMENTATION, 2018). A Figura 18 apresenta os caminhos escolhidos para a classificação de acessos legítimos, enquanto a Figura 19 apresenta os caminhos escolhidos para a classificação dos ataques. Em ambos os casos as portas 22 e 443 são de grande importância assim como o "*Flow Duration*" e o "*Flow Packets/s*".

Figura 18 - *Decision plot 1*



Fonte: Imagem dos autores (2023).

Figura 19 - *Decision plot 2*



Fonte: Imagem dos autores (2023).

Através desses resultados, o trabalho demonstra como é possível utilizar o aprendizado de máquina supervisionado para fazer a detecção de ataques em serviços de rede. Claramente, os resultados obtidos nesses treinamentos refletem os registros da base de dados utilizada e pode ser diferente se aplicada em outros ambientes com outros comportamentos.

CONCLUSÕES

O projeto explorou técnicas de aprendizado de máquina para detecção de ataques DDoS. Utilizando conjuntos de dados abrangentes e diversificados, foi possível desenvolver um modelo capaz de identificar padrões e comportamentos distintivos associados a ataques DDoS. Embora os modelos de Regressão Logística, *naive bayes* e *support vector machine* não tenham produzido resultados satisfatórios, o resultado ainda sugere a viabilidade da classificação de ataques. Em contraste, o modelo de *Random Forest* obteve resultados excepcionais na classificação, destacando a eficácia dessa abordagem na categorização do tráfego de rede.

Essa abordagem proativa não apenas fortalece as defesas existentes, mas também fornece uma camada adicional de segurança necessária para proteger ativos digitais críticos. A pesquisa também destaca a necessidade contínua de inovação e adaptação na área de segurança cibernética, utilizando as mais recentes tecnologias para manter a resiliência diante das ameaças emergentes.

Finalmente, o trabalho contribui de maneira significativa com a evolução de ferramentas de defesa cibernética, com o objetivo de identificar ataques que permitam contribuir com a integridade e a disponibilidade das infraestruturas de tecnologia.

AGRADECIMENTOS

Os autores agradecem a comunidade acadêmica do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo do campus Pirituba, aos professores do curso técnico de redes de computadores integrado ao ensino médio e ao Programa Institucional de Bolsas de Iniciação Científica. (PIBIFSP - Ensino Técnico - 41/2022) pelos recursos disponibilizados e pelo suporte a esta pesquisa.

REFERÊNCIAS

2022 in review: DDoS attack trends and insights. Microsoft, 2023. Disponível em: <https://www.microsoft.com/en-us/security/blog/2023/02/21/2022-in-review-ddos-attack-trends-and-insights/>. Acesso em: 23 de Fevereiro de 2023.

CLARINDO, Anderson Beloni; SILVA, Keterly Geovana Gouveia. Detectando ataques DDoS com inteligência artificial. Centro Universitário Sagrado Coração - UNISAGRADO, 2022. Disponível em: <https://repositorio.unisagrado.edu.br/jspui/handle/handle/1387>. Acesso em: 16 ago. 2023.

DDoS 2019 | Datasets | Research | Canadian Institute for Cybersecurity | UNB, 2019. Disponível em: <https://www.unb.ca/cic/datasets/ddos-2019.html>. Acesso em: 13 ago. 2023.

decision plot — SHAP latest documentation. 2018. Disponível em: <https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/decision_plot.html>. Acesso em: 16 nov. 2023.

FIGUEIREDO, Bruno Ianoni; FERREIRA, Frederico Reid Sulahian; SIMANTOB,

Marco Gabriel de Melo. Estudo e investigação de técnicas de IA para detecção de ataques DDoS. Universidade Presbiteriana Mackenzie, 2022. Disponível em: <https://dspace.mackenzie.br/handle/10899/31220>. Acesso em: 16 ago. 2023.

FOX, Eduardo Farias Brinds-Ley. Detecção de ataques syn-flooding em redes definidas por software. Dissertação. 2019. Disponível em: <https://repositorio.ufpb.br>. Acesso em: 13 nov. 2023.

LIMA, Jefferson Luiz Pessoa. Adversarial oversampling : um método para balanceamento baseado em Redes Neurais Adversárias. masterThesis. 2020. Disponível em: <https://repositorio.ufpe.br/handle/123456789/39224>. Acesso em: 16 nov. 2023.

LUDERMIR, Teresa Bernarda. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. Estudos Avançados, SciELO Brasil, v. 35, p. 85–94, 2021.

OpenAI Status. 2023. Disponível em: <https://status.openai.com/>. Acesso em: 10 nov. 2023.

O que são ataques de DDoS? 2023. Disponível em: <https://www.kaspersky.com.br/resource-center/threats/ddos-attacks>. Acesso em: 13 ago. 2023.

PELLOSO, Mateus; VERGÜTZ, Andressa; SANTOS, Aldri; NOGUEIRA, Michele. Um Sistema Autoadaptável para Predição de Ataques DDoS Fundado na Teoria da Metaestabilidade. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS (SBRC), 36. , 2018, Campos do Jordão. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2018 . p. 726-739. ISSN 2177-9384. DOI: https://doi.org/10.5753/sbrc.2018.2454. Acesso em: 13 ago. 2023.

SILVA, Henrique Cesar Ferreira; PIETRO, Luca Baron; DARIO, Luís Gustavo

Beccheri; MORAES, Eduardo Alves; HERNANDES JR., Paulo R. Galego; BAREA, Emerson Rogério Alves. Aprendizado de Máquina Aplicado na Classificação de Alertas de Ataques de DoS em Sistemas de Detecção de Intrusão. In: WORKSHOP DE TRABALHOS DE INICIAÇÃO CIENTÍFICA E DE GRADUAÇÃO - SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS (SBRC), 38. , 2020, Rio de Janeiro. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2020 . p. 241-248. ISSN 2177-9384. DOI: https://doi.org/10.5753/sbrc_estendido.2020.12425. Acesso em: 13 ago. 2023.

RIGATTI, Steven J. Random Forest. Journal of Insurance Medicine, v. 47, n. 1, p. 31–39, 2017. Acesso em: 17 nov. 2023.

SARRAF, Saman. Analysis and Detection of DDoS Attacks Using Machine Learning Techniques. American Scientific Research Journal for Engineering, Technology, and Sciences, v. 66, p. 95–104, 2020. Acesso em: 17 nov. 2023.

SOMANI, Gaurav; GAUR, Manoj Singh; SANGHI, Dheeraj; CONTI, Mauro; BUYYA, Rajkumar. DDoS attacks in cloud computing: Issues, taxonomy, and future directions. Computer Communications, v. 107, p. 30–48, 2017. Acesso em: 17 nov. 2023.

VALDATI, Aline de Brittos. Inteligência artificial - IA. 1. ed. São Paulo: Contentus, 2020. *E-book*. Disponível em: https://plataforma.bvirtual.com.br. Acesso em: 16 ago. 2023.

Welcome to the SHAP documentation — SHAP latest documentation. 2018. Disponível em: <https://shap.readthedocs.io/en/latest/index.html>. Acesso em: 16 nov. 2023.