CENTRO UNIVERSITÁRIO DE BRASÍLIA
FACULDADE DE TECNOLOGIA E CIÊNCIAS SOCIAIS APLICADAS

RESUMO EXECUTIVO

ENGENHARIA REVERSA DE MALWARE

Membros do Projeto

22153474 Augusto Oliveira Silva

Orientador

Prof. MSc. Valdemir dos Santos Silva

BRASÍLIA, junho de 2025

AGRADECIMENTOS

A Deus, por ter me dado a coragem, a força e o discernimento para perseverar, e à Virgem Maria, Mãe Imaculada, por me confortar e me acalmar diante das preocupações diárias. Aos meus pais, Anthony e Rose Meire que, pelo exemplo, ensinamentos e correções me ajudaram a me manter firme nos momentos de fraqueza e me mostraram o caminho a ser seguido. Obrigado por todo o amor e carinho. Aos meus irmãos, Daniel e Gustavo, pelo companheirismo e parceria. À Zélia, minha segunda mãe, Elizabeth, minha cunhada, Daniel, Santiago e Estêvão, meus sobrinhos, Pedro e Maria, meus padrinhos de batismo. Muito obrigado por toda a ajuda, oração e alegria que me proporcionam. Aos meus avós, Silva, Rosa, Dalva e Webert, por toda a ajuda e palavras de motivação. À Ana Luíza, minha namorada que, nos momentos de angústia, me escutou, me acolheu e me ajudou a continuar de pé e, nas ansiedades, me ajudou a viver o momento presente. Aos irmãos da Comunidade 9 do Caminho Neocatecumenal da Paróquia Nossa Senhora da Assunção por me escutarem nas minhas dificuldades e rezarem por mim. Aos meus professores e colegas de turma pelos ensinamentos passados e disponibilidade ao longo de toda a minha formação acadêmica.

RESUMO

O presente trabalho teve como objetivo aplicar modelos de Machine Learning juntamente com a Engenharia Reversa para a identificação de padrões de tráfego de redes, buscando detectar malwares e ataques cibernéticos. Para isso, foram utilizados os seguintes conjuntos de dados CIC-IDS2017 e CSE-CIC-IDS2018. Esses modelos preditivos, com características gerais de conjuntos são amplamente utilizados e reconhecidos como ótimas fontes para estudos na área da cibersegurança. O pré-processamento incluiu o tratamento de valores nulos, duplicados, colunas e a normalização de variáveis. Para a modelagem preditiva foram testados diferentes algoritmos, com ênfase em modelos interpretáveis como Random Forest e XGBoost, com foco tanto em classificações binárias quanto multiclasse, obtendo-se resultados satisfatórios em termos de acurácia e desempenho. Além da criação dos modelos citados, foi desenvolvida uma interface interativa que permite ao usuário testar e realizar previsões automatizadas, facilitando assim a análise de riscos associados a possíveis arquivos maliciosos. Por fim, através dos resultados que foram alcançados esta pesquisa possibilitou o desenvolvimento de soluções mais confiáveis e seguras, promovendo o aumento da utilização da inteligência artificial na detecção de malwares e melhorando a proteção e segurança contra ataques em redes modernas.

Palavras-chave: Engenharia Reversa. Malware. Machine Learning. Cibersegurança.

SUMÁRIO

1. PROBLEMA/OPORTUNIDADE

A crescente sofisticação das ameaças cibernéticas tem alavancado a necessidade de técnicas avançadas para a detecção de ataques em redes de comunicação. Com o crescimento exponencial do volume de dados e maior complexidade dos padrões de tráfego, os sistemas tradicionais de segurança têm enfrentado desafios nunca vistos. Estes avanços tecnológicos ampliam significativamente a superfície de ataque, tornando redes e dispositivos Internet das Coisas (IoT) mais suscetíveis a sofrerem com atividades maliciosas, como ataques DDoS (Distributed Denial of Service — Ataque de Negação de Serviço Distribuído), injeções de código malicioso e exploração de vulnerabilidades.

Neste caso, o uso da Inteligência Artificial (IA) e Machine Learning (ML) pode ser uma boa solução para possíveis riscos e ameaças virtuais. Os modelos de aprendizagem de máquina têm a habilidade de analisar grandes quantidades de tráfego na rede, achar padrões e estruturas que podem indicar ataques. Mas, devido à complexidade dos modelos de ML, um dos grandes problemas é a interpretabilidade, ou seja, a análise e o entendimento das decisões tomadas pelo próprio algoritmo. Isso se deve ao fato de que não é fácil entender como e porque esses algoritmos tomam suas decisões. Essa falta de clareza pode trazer efeitos negativos no processo de implantação dessas técnicas, principalmente em locais corporativos, onde compreender o que levou o modelo a tomar uma decisão é a chave para confiar ou não nos respectivos algoritmos.

À vista disso, o presente projeto visa responder a pergunta “Como detectar e se prevenir de malwares e qual o seu impacto em sistemas de cibersegurança?” através do desenvolvimento de modelos de aprendizado de máquina de fácil entendimento para análise de tráfego de rede e identificação de malwares. Com esse intuito foram utilizados os conjuntos de dados CIC-IDS2017 e CSE-CID-IDS2018, pois ambos contêm registros de tráfego de rede. O principal objetivo deste trabalho foi desenvolver um sistema de machine learning que, através da engenharia reversa, seja capaz de analisar, identificar e prever se um arquivo é ou não malicioso.

Ao final desta pesquisa espera-se que os resultados obtidos possam colaborar para o avanço de soluções na área da cibersegurança, permitindo entender, de forma mais clara, qual o impacto causado por um arquivo malicioso.

2. BENEFÍCIOS DA SOLUÇÃO

A solução proposta no projeto combina engenharia reversa e machine learning para reforçar a segurança de redes contra ameaças digitais. Na esfera tecnológica, destaca-se a automação na detecção de ameaças e a adaptabilidade a diferentes tipos de tráfego. No campo educacional, a aplicação serve como ferramenta didática para demonstrar a análise de malwares e facilita a abordagem interdisciplinar entre IA, redes e segurança. Já na esfera social, contribui para a redução de riscos e danos causados por ataques cibernéticos, além de promover o acesso à cibersegurança por meio de uma interface intuitiva, acessível até para usuários com pouco conhecimento técnico.

3. PÚBLICO-ALVO

O sistema desenvolvido tem como potenciais usuários profissionais de segurança da informação, profissionais da área da tecnologia. Além disso, a aplicação pode ser implementada dentro de organizações de pequeno e médio porte, uma vez que ela proporciona uma camada extra de segurança e também dá suporte ao monitoramento e diagnóstico rápido de padrões de tráfego. Por fim, o sistema pode ser acessado e utilizado por todos os usuários que tenham interesse técnico em aprofundar os conhecimentos voltados para o desenvolvimento de um algoritmo de aprendizado de máquina.

4. PROTÓTIPO VISUAL

A interface do sistema foi feita utilizando o Streamlit, um framework de código aberto destinado à criação de aplicações na área de ciência de dados e aprendizado de máquina. O objetivo principal foi desenvolver uma aplicação web leve e de fácil navegação e que, ao mesmo tempo, fornecesse ao usuário a interação com os modelos de machine learning, sem a necessidade de conhecimento técnico profundo sobre o assunto. Todas as etapas necessárias de instalação e execução local da aplicação foram descritas em um repositório do GitHub. Para acessá-lo, clique no link a seguir: https://github.com/gut0oliveira/Data-Science-Capstone

4.1. Funcionalidades

A aplicação adota um layout e paleta de cores pensada na experiência visual do usuário final. O seu

layout é configurado em tela ampla e com a barra lateral inicialmente recolhida, otimizando o uso do espaço principal da aplicação. Através disso, algumas de suas principais funcionalidades são:

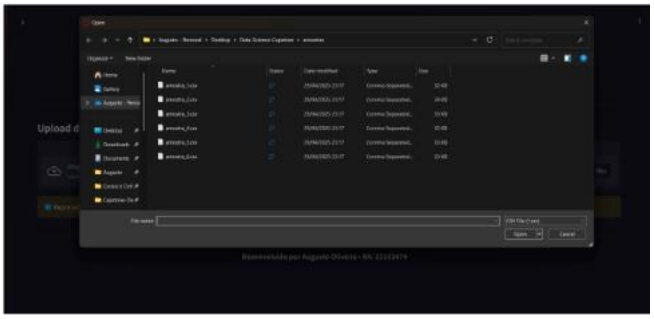
- Upload de arquivos CSV com dados de tráfego de rede;
- Detecção automática do tipo de classificação: Binária ou Multiclasse;
- Escolha de modelos de ML previamente treinados;
- Visualização dos resultados com gráficos;
- Barra lateral com breves informações sobre o projeto.

4.2. Abertura da aplicação

O usuário acessa o sistema via navegador, sendo recebido com o título e uma breve descrição da funcionalidade.

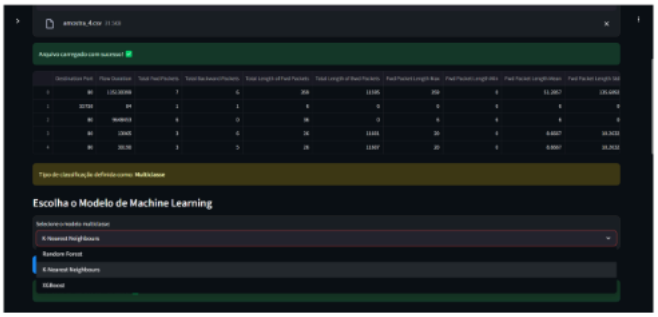
4.3. Upload de arquivo CSV

O usuário seleciona uma das amostras aleatórias de arquivo CSV disponibilizadas na instalação local do projeto com tráfego de rede para análise.



4.4. Seleção do modelo

O sistema detecta automaticamente se a amostra selecionada poderá ser classificada como binário ou multiclasse. Após isso, o usuário escolhe o modelo desejado.

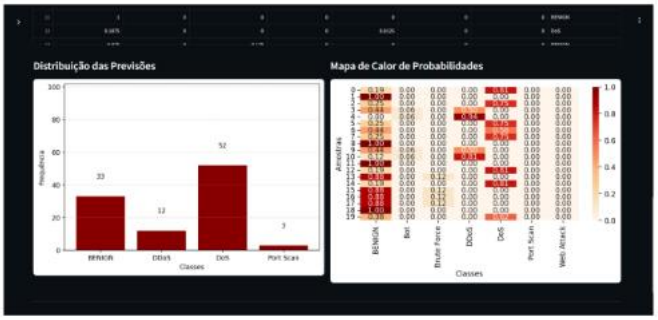


4.5. Visualização dos resultados

Após o processamento, são exibidos os resultados em formato de tabela, como a exibição de um conjunto de dados, visualizado na etapa de no resultado da etapa de upload do arquivo CSV.

4.6. Interpretação e decisão

O usuário pode interpretar os resultados e decidir medidas de segurança com base na classificação obtida dos dados e nos gráficos gerados.



5. CONSIDERAÇÕES FINAIS

Este projeto de pesquisa teve como proposta investigar e colocar em prática o uso da engenharia reversa associada a modelos de machine learning para a identificação e prevenção de malwares nos tráfegos de rede, tendo como principal foco a análise do impacto e da aplicabilidade dessas tecnologias dentro da área da cibersegurança. Através da criação de algoritmos e experimentos baseados nos conjuntos de dados CIC-IDS2017 e CSE-CIC-IDS2018, conseguiu-se realizar a previsão e verificar a eficácia, tanto da classificação binária — benigno ou malicioso — quanto da classificação multiclasse — ataques específicos.

Ao retomar a pergunta de pesquisa “Como detectar e se prevenir de malwares e qual o seu impacto em sistemas de cibersegurança?”, pode-se concluir que a mesma foi respondida ao longo do estudo com evidências práticas. Provou-se que, com modelos estruturados, como o Support Vector Machine (SVM) e o XGBoost, explanados nos tópicos anteriores, a identificação com elevada acurácia do método, é possível, desde que feita da maneira mais adequada para o contexto em que está inserida. Ademais, após a análise das portas de destino mais vulneráveis e dos tipos de ataques mais frequentes, é perceptível a importância de reforçar as ações preventivas específicas e estratégicas para a respectiva variante de ataque.

No que diz respeito aos objetivos gerais da pesquisa de desenvolver um sistema de machine learning que, através da engenharia reversa, seja capaz de analisar, identificar e prever se um arquivo é ou não malicioso, os mesmos foram alcançados através da criação dos algoritmos e de uma interface interativa que permite ao usuário, de forma automatizada, ter uma previsibilidade do risco de segurança que um arquivo pode ou não oferecer. Quantos aos objetivos específicos, por meio da literatura específica, da prática e da utilização de modelos eficazes, testados e avaliados por métricas amplamente reconhecidas no mercado e na área da ciência de dados, é constatável que os respectivos objetivos foram atingidos com êxito. Por fim, este projeto apresenta importantes contribuições para o cenário atual da cibersegurança no mundo. Mostrando que, o emprego de técnicas transparentes na análise de um malware, facilita e abre caminhos para futuras investigações, como o uso de Deep Learning — Aprendizado Profundo — em ambientes com tráfego de rede elevado, como também, a análise de arquivos maliciosos em tempo real.

Portanto, a junção de engenharia reversa e machine learning não é somente viável, mas também, altamente propícia como solução estratégica de defesa cibernética frente ao rápido avanço das ameaças digitais.

REFERÊNCIAS

AMAZON WEB SERVICES. ***Validação cruzada*.** Disponível em: https://docs.aws.amazon.com/pt_br/machine-learning/latest/dg/cross-validation.html. Acesso em: 13 maio 2025.

CRONAPP. **Engenharia reversa**. Disponível em: https://blog.cronapp.io/engenharia-reversa/. Acesso em: 28 fev. 2025.

CRONAPP. **Engenharia reversa de software**. Disponível em: https://blog.cronapp.io/engenharia-reversa-de-software/#Como_a_engenharia_reversa_de_softw are_funciona_na_pratica. Acesso em: 28 fev. 2025.

DEVMedia. **Trabalhando com engenharia reversa – Revista Engenharia de Software Magazine 59.** Disponível em: https://www.devmedia.com.br/trabalhando-com-engenharia-reversa-revista-engenharia-de-softw are-magazine-59/28203. Acesso em: 28 fev. 2025.

IBM. **A história do malware**. Disponível em: https://www.ibm.com/br-pt/think/topics/malware-history. Acesso em: 10 mar. 2025.

IBM. **O que é malware?**. Disponível em: https://www.ibm.com/br-pt/topics/malware. Acesso em: 10 mar. 2025.

IBM. **O que são SVMs?**. Disponível em: https://www.ibm.com/br-pt/think/topics/support-vector-machine. Acesso em: 1 abr. 2025.

MICROSOFT. **O que é malware?.** Disponível em: https://www.microsoft.com/pt-br/security/business/security-101/what-is-malware. Acesso em: 10 mar. 2025.

OTTONI, Otton. **Engenharia reversa**. Disponível em: http://www2.ic.uff.br/~otton/graduacao/informatical/apresentacoes/eng_reversa.pdf. Acesso em: 28 fev. 2025.

UNIVERSITY OF NEW BRUNSWICK. **Canadian Institute for Cybersecurity – About**. Disponível em: https://www.unb.ca/cic/about/index.html. Acesso em: 15 mar. 2025.

WISHBOX. **Engenharia reversa: o que é, como funciona e onde aplicar.** Disponível em: https://www.wishbox.net.br/blog/engenharia-reversa/. Acesso em: 2 mar. 2025.