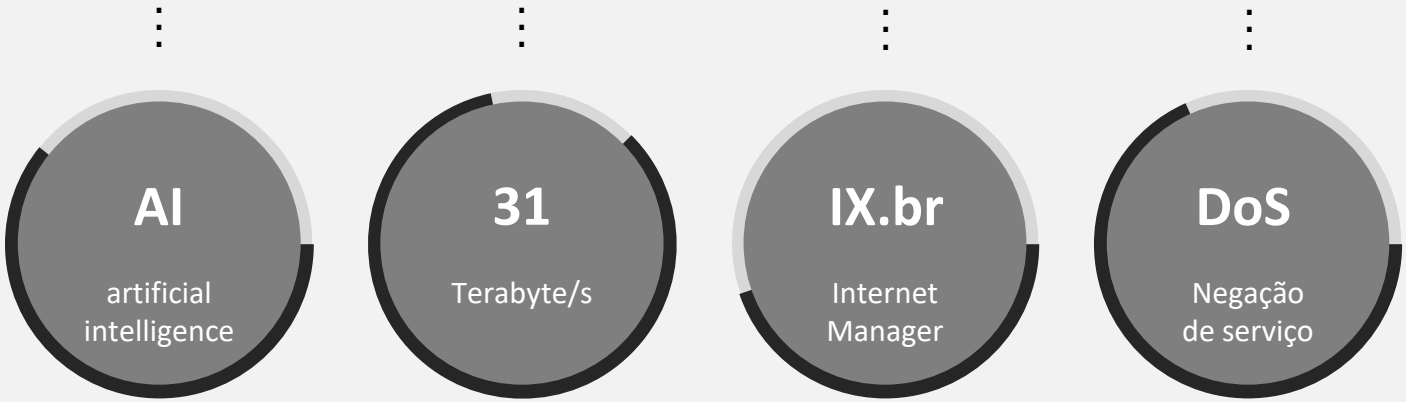
DAVI JORGE LEITE SANTOS

Aplicação do algoritmo Random Forest para classificação de tráfego de rede benigno em meio a ataques DDoS UDP

INTRODUÇÃO



Para analise

Trafego gerado

Reistro nacional

Quem gerou

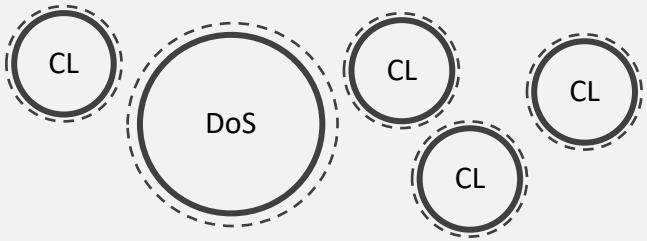
A negação de serviço é uma questão séria, capaz de causar transtornos aos usuários e grandes prejuízos financeiros às empresas. Por essa razão, a identificação desses eventos é crucial.

PROBLEMA MOTIVADOR

Problema?

- O objetivo é contribuir para a utilização da inteligência artificial na análise de tráfego para prever ataques DoS (negação de serviço).
- Foi realizada uma investigação sobre a viabilidade do algoritmo Random Forest.
- A questão é se é viável empregar um algoritmo Random Forest para classificar tráfego benigno quando confrontado com um ataque DDoS (Distributed Denial of Service) coletado através do protocolo Netflow.

OBJETIVOS E JUSTIFICATIVA



OBJETIVOS

Analisar a utilização da aprendizado de máquinas no âmbito da cibersegurança

E Investigar um ataque DDoS do tipo inundação de UDP utilizando uma base de dados coletada pelo protocolo Netflow para classificar o tráfego benigno.

OBJETIVOS ESPECIFICOS

Analisar a viabilidade da utilização da inteligência artificial para tomadas de decisões no âmbito da cibersegurança;

Realizar análise exploratória do tráfego de rede UDP e verificar características do ataque DoS e DDoS, para metrificá-las para utilização do Random Forest;

Classificar o tráfego dos usuários benignos em meio a um ataque DoS, separando-os do próprio ataque

JUSTIFICATIVA

O estudo é importante para aumentar o nível da identificação de tráfego benigno na área de cibersegurança, onde o foco seja a classificação do DDoS UDP e que possam ainda manter as informações de seus clientes ativas

REFERENCIAL TEÓRICO

DDOS E DOS

Um ataque DoS (Denial of Service) é direcionado a um único dispositivo ou recurso, visando sobrecarregá-lo com tráfego, impedindo seu acesso legítimo. Já um DDoS (Distributed Denial of Service) envolve múltiplos dispositivos coordenados para atacar simultáneamente, aumentando consideravelmente o volume de tráfego. (SILVA; SALLES, 2015)

Métricas para detecção de ataques DDoS

- Ataques de negação de serviços pretende é fazer que uma rede ou algum tipo de serviço, tenham um mau funcionamento e até mesmo a parada dele.

Artificial Intelligence in Cyber Security

- Indústrias e empresas de setores privados já adotam e utilizam programas que tenham a IA.

PREDICTION OF CYBER ATTACKS USING MACHINE LEARNING TECHNIQUE

- A utilização de um modelo de análise para prevenção, podem assim auxiliar na prevenção redes hostis.

Machine Learning based DDos Detection Through

- A importância de análise (visando métodos de identificação) e a análise de dados feita pelo netflow por uma IA, pode beneficiar as pesquisas nessa área

METODOLOGIA

METODOLOGIA

Análise Exploratória: Compreensão dos dados e suas correlações para escolha da métrica.

Treinamento do Algoritmo: Utilização do Random Forest após a seleção da métrica.

Avaliação da Acurácia: Verificação da precisão do algoritmo e sua capacidade de identificar ataques DDoS.

Python3 e Jupyter-Lab: Desenvolvimento da aplicação. Scikit-learn, Matplotlib, Seaborn, GraphViz, Pandas e NumPy: Ferramentas para manipulação e visualização dos dado.

Bibliotecas

Base de Dados

CIC-DDoS2019: Conjunto de dados com informações de tráfego benigno e DDoS, escolhendo o DrDoS-UDP para o estudo.

Divisão da base em 75% para treinamento e 25% para testes visando validar o algoritmo.

Testes

RESULTADOS OBTIDOS

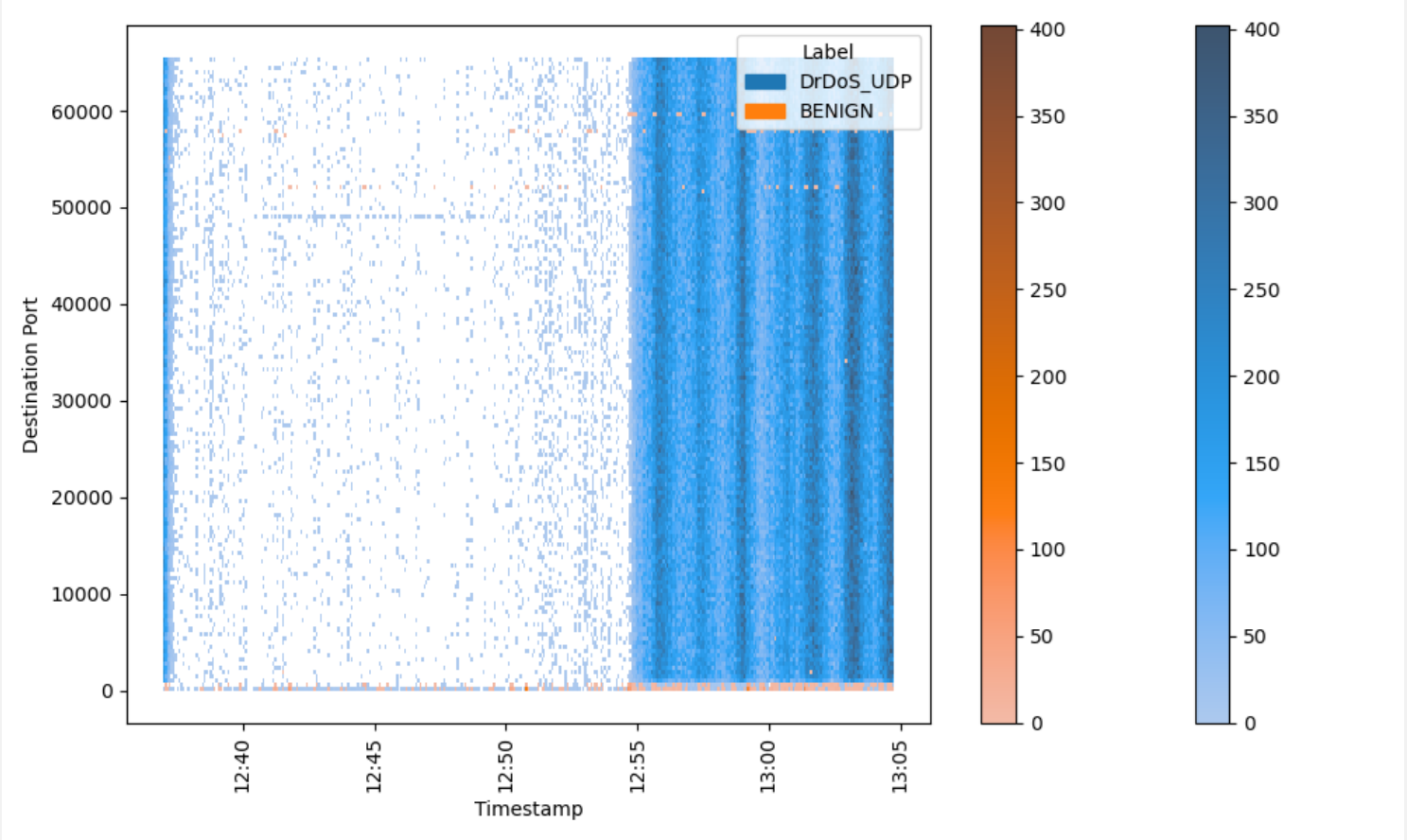
RESULTADOS OBTIDOS

Inicialmente, foi feito uma análise exploratória da base de dados, onde teve um foco de entender o comportamento dos dados.

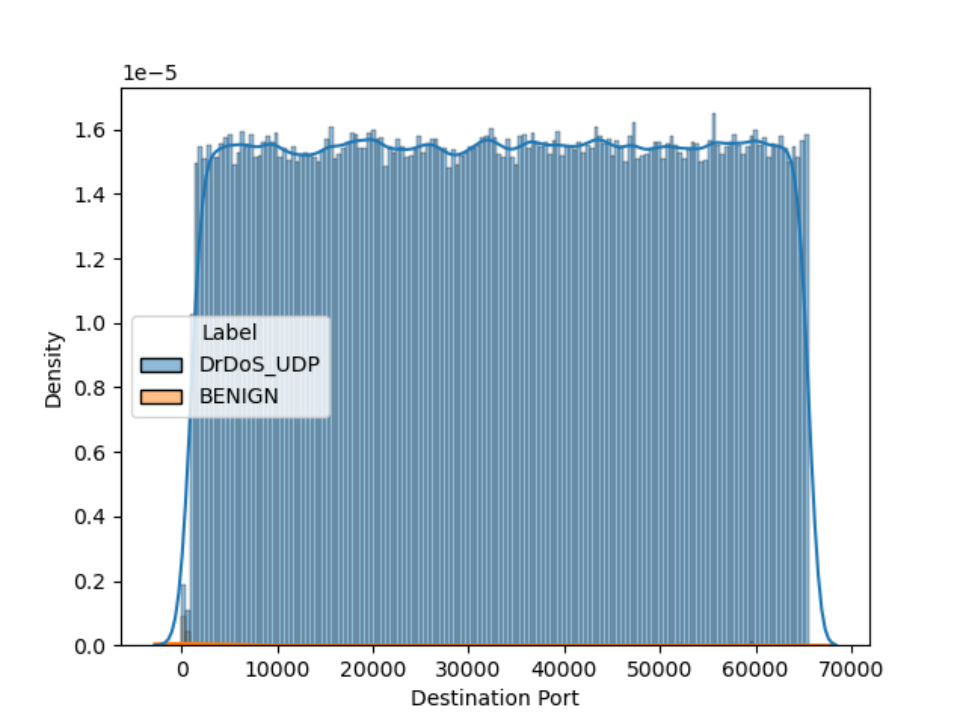
RESULTADOS OBTIDOS

A base de dados é composta por 3.136.802 linhas de registros de tráfego DDoS e benigno. A distribuição de informações dentro dessa base é a seguinte: 2157 informações benignas e 3.134.645 dados de DDoS.

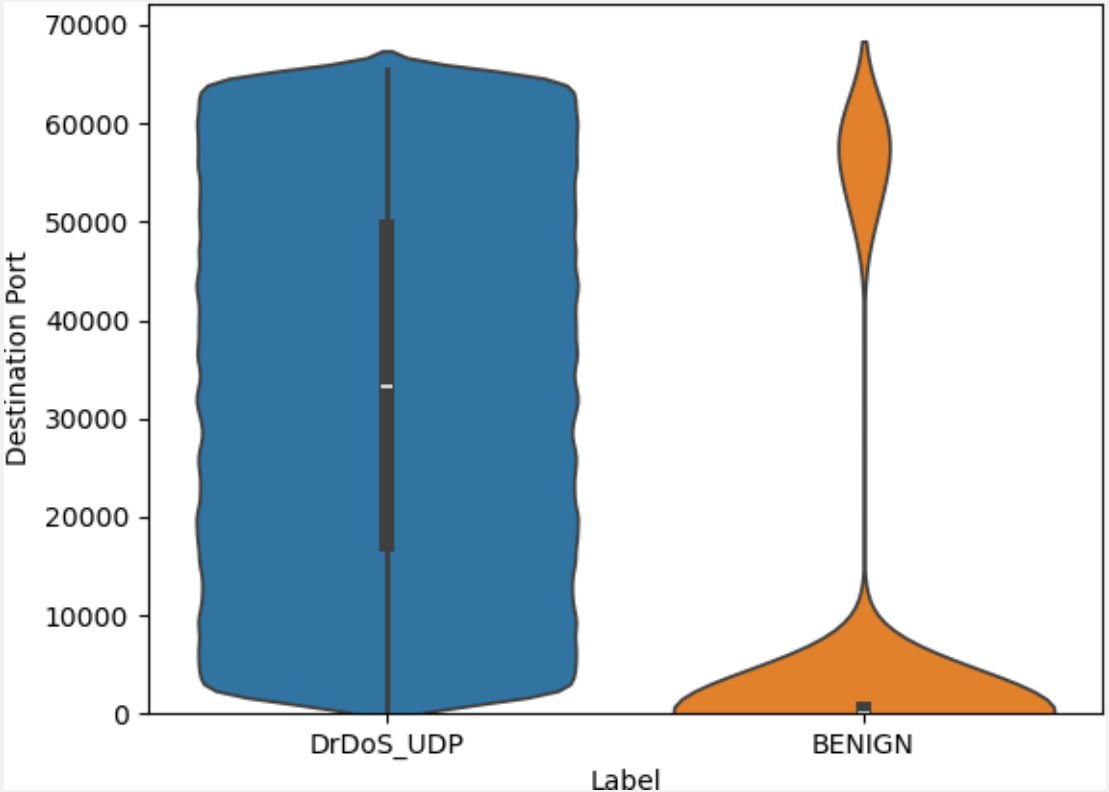
RESULTADOS OBTIDOS



RESULTADOS OBTIDOS

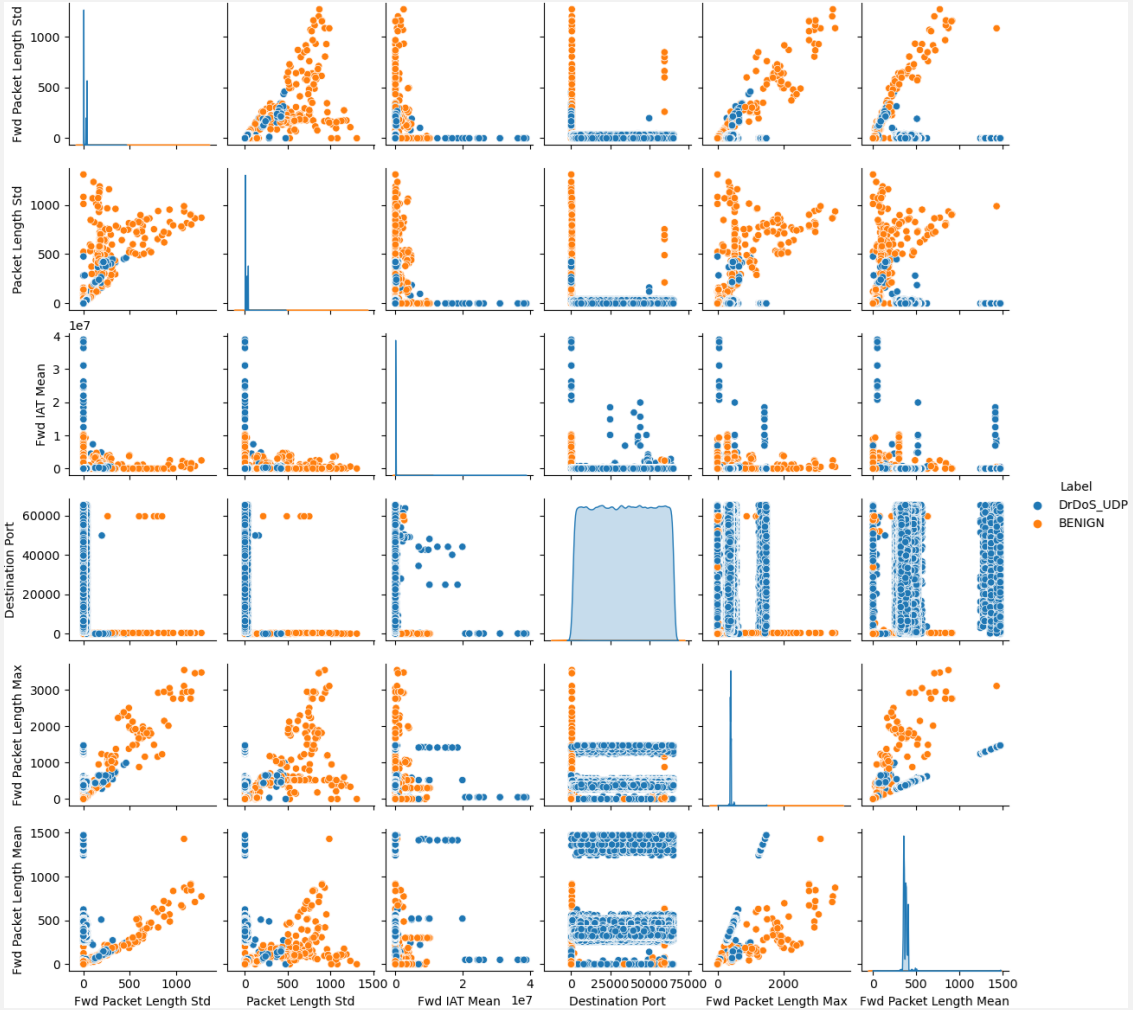


RESULTADOS OBTIDOS

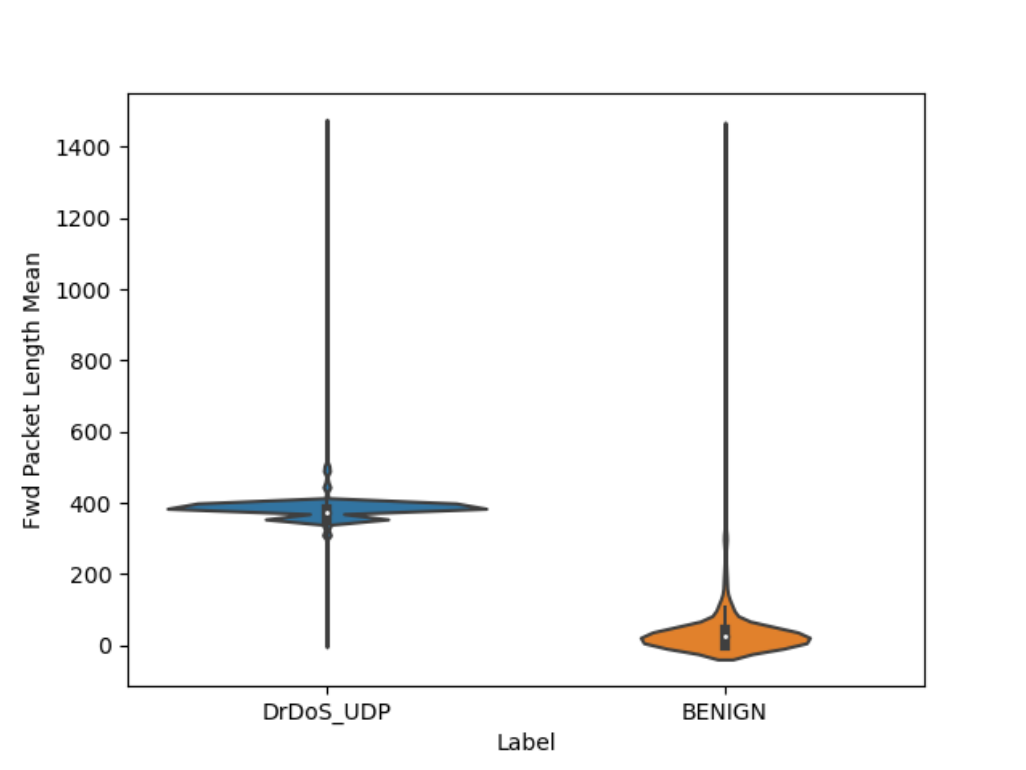


DAVI JORGE LEITE SANTOS

RESULTADOS OBTIDOS



RESULTADOS OBTIDOS



TREINAMENTO DOS ALGORITMOS

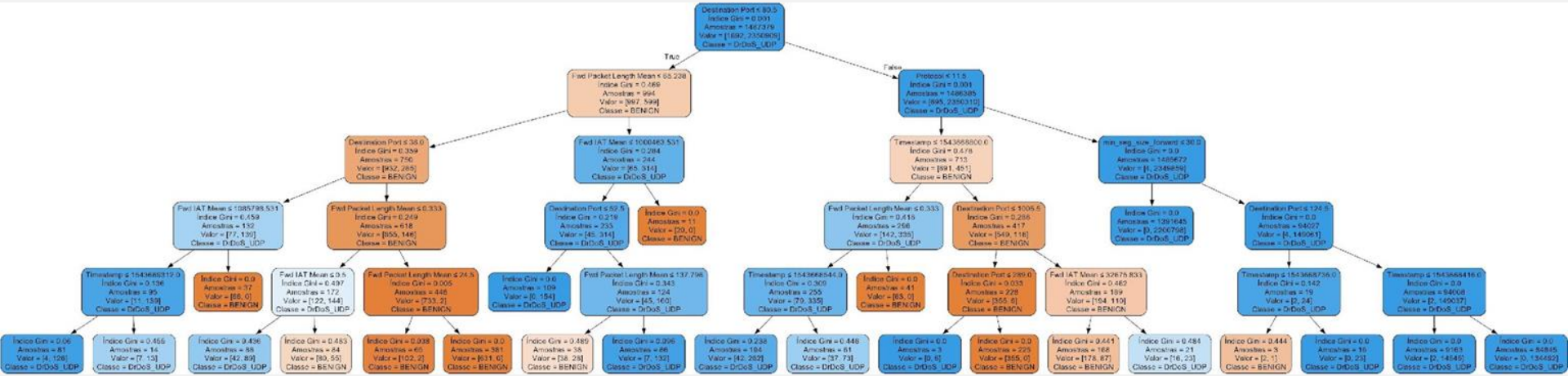
SOBRE OS TREINAMENTOS

Utilizou-se a biblioteca Scikit-learn para o treinamento dos algoritmos Random Forest na base de dados, após o pré-processamento. Três experimentos foram conduzidos.

EXPERIMENTO 1: IDENTIFICAÇÃO NA BASE COMPLETA

O primeiro experimento empregou a base de dados completa. Com a primeira parametrização, foi observado overfitting, gerando até 19 níveis na árvore, com nós-folha de apenas 2 amostras. Optou-se por 100 árvores de até 5 níveis, obtendo bons resultados na classificação de DDoS e dados benignos.

EXPERIMENTO 1: IDENTIFICAÇÃO NA BASE COMPLETA



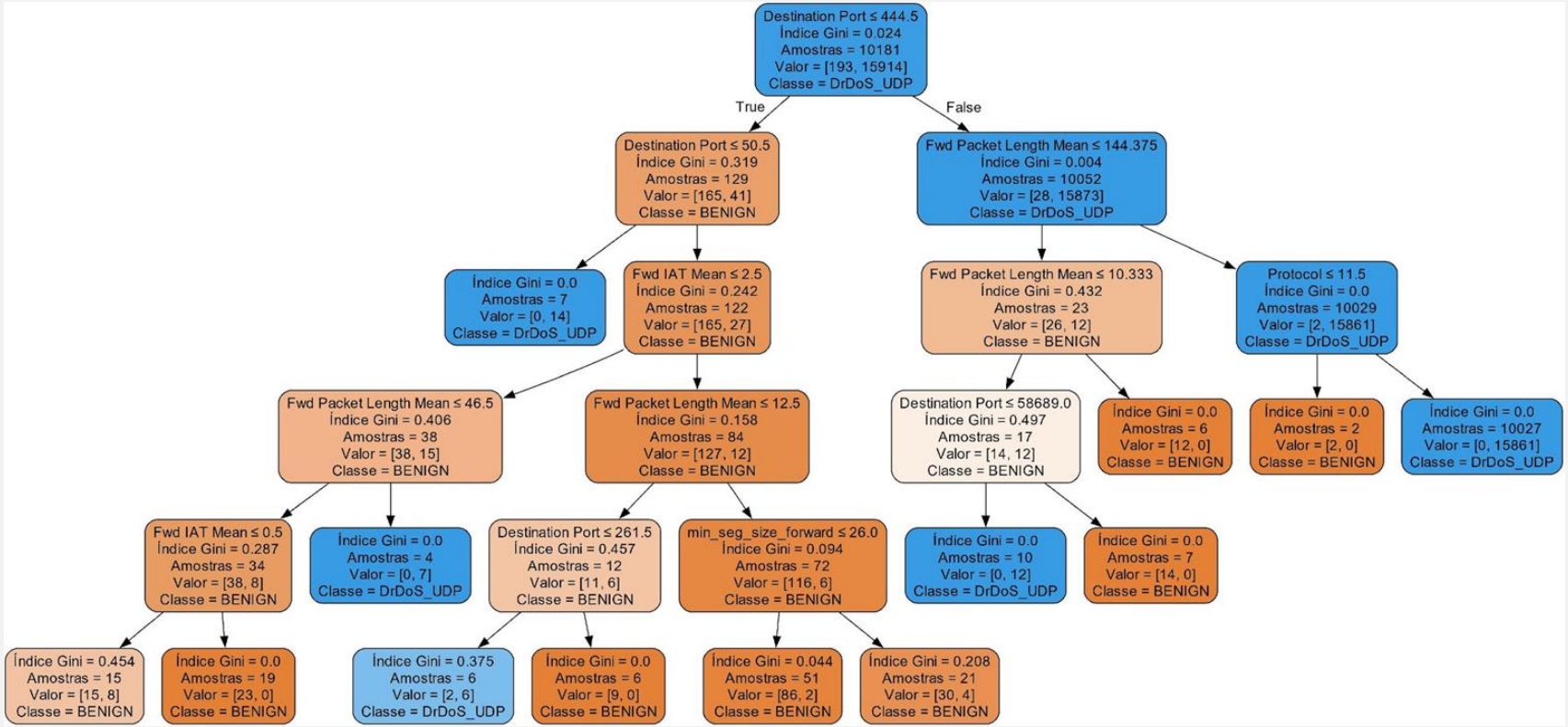
EXPERIMENTO 1: IDENTIFICAÇÃO NA BASE COMPLETA

Floresta Aleatória	Precisão	Amostras
BENIGN	99%	531
DrDoS_UDP	100%	783.670
Macro AVG	99%	784.201
Weighted AVG	100%	784.201

EXPERIMENTO 2: IDENTIFICAÇÃO NO INÍCIO DOS ATAQUES

Filtrando o início dos ataques (entre 12:54 e 12:55), o treinamento alcançou 100% de precisão na classificação de DDoS e dados benignos. Apesar da escassez de dados de DDoS, a identificação eficiente desses dados foi possível, especialmente dos dados benignos, ressaltando a importância de manter o fluxo normal das informações.

EXPERIMENTO 2: IDENTIFICAÇÃO NO INÍCIO DOS ATAQUES



EXPERIMENTO 2: IDENTIFICAÇÃO NO INÍCIO DOS ATAQUES

Floresta Aleatória	Precisão	Amostras
BENIGN	100%	62
DrDoS_UDP	100%	5.307
Macro AVG	100%	5.369
Weighted AVG	100%	5.369

EXPERIMENTO 3: ANÁLISE DA GENERALIZAÇÃO

Este teste avaliou a capacidade de generalização do treinamento anterior em toda a base de dados. Apesar da mistura de dados benignos e DDoS, obteve-se uma classificação de 100% para DDoS e 83% para dados benignos. Isso reflete a realidade, indicando um fluxo menor de dados benignos comparado aos ataques de negação de serviço.

EXPERIMENTO 3: ANÁLISE DA GENERALIZAÇÃO

Floresta Aleatória	Precisão	Amostras
BENIGN	83%	531
DrDoS_UDP	100%	783.670
Macro AVG	91%	784.201
Weighted AVG	100%	784.201

CONCLUSÃO

CONCLUSÃO

Este estudo analisou o comportamento do DDoS em uma base de dados específica, identificando tráfego DDoS e informações benignas com sucesso usando o algoritmo Random Forest.

CONCLUSÃO

Contribui para análises de segurança ao possibilitar a identificação eficaz de ataques DDoS e tráfego comum.

CONCLUSÃO

Alcançou os objetivos ao analisar detalhadamente os dados, obtendo uma taxa de classificação de usuários benignos de 83%.

CONCLUSÃO

Destaca-se a importância dessa classificação para manter a integridade das operações durante ataques DDoS, assegurando a continuidade do tráfego normal para usuários não vinculados aos ataques.

THANKS

Questions?

"A verdade é aquilo que resiste ao teste da experiência."

Retirado do livro "As leis da ciência e as leis da ética",1950 - Albert Einstein

THANKS
Redes sociais

LinkedIn - Davi J Leite Santos

Instagram - davij9