Laboratório de Redes

Experimentação e Aprendizado de Máquina em Redes de Computadores

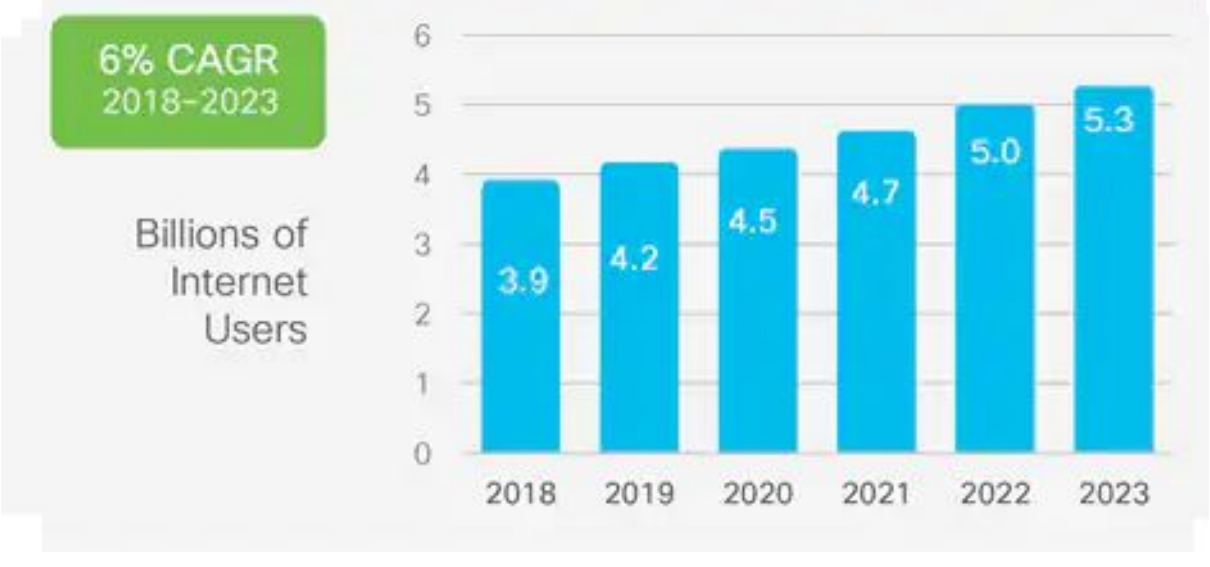
Alunos: Wagner Porto Ferreira
Willen Borges Coelho
Vitor Fontana Zanotelli

Prof.: Rodolfo da Silva Villaça

UFES
UNIVERSIDADE FEDERAL
DO ESPÍRITO SANTO

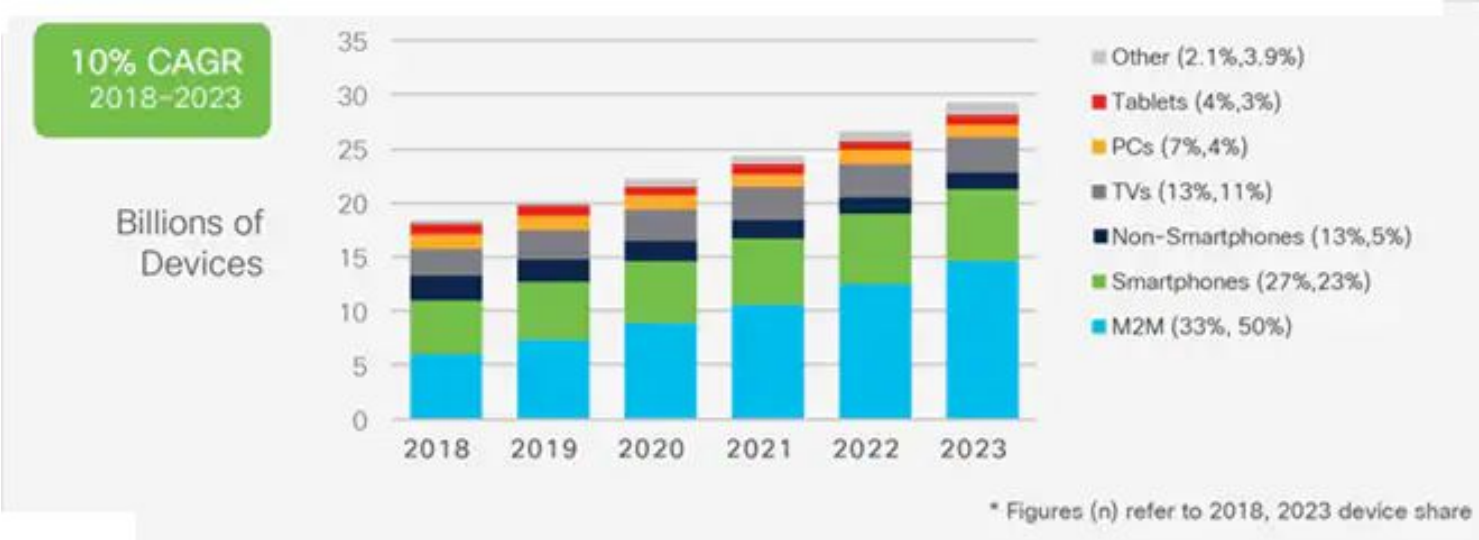
Previsões mostram que o número de usuários de internet continuará crescendo nos próximos anos

- ⅔ da população mundial terá acesso à internet até 2023;
- São 5.3 bilhões de usuários, um aumento de 35% em relação a 2018



O número de dispositivos conectados em redes IP também

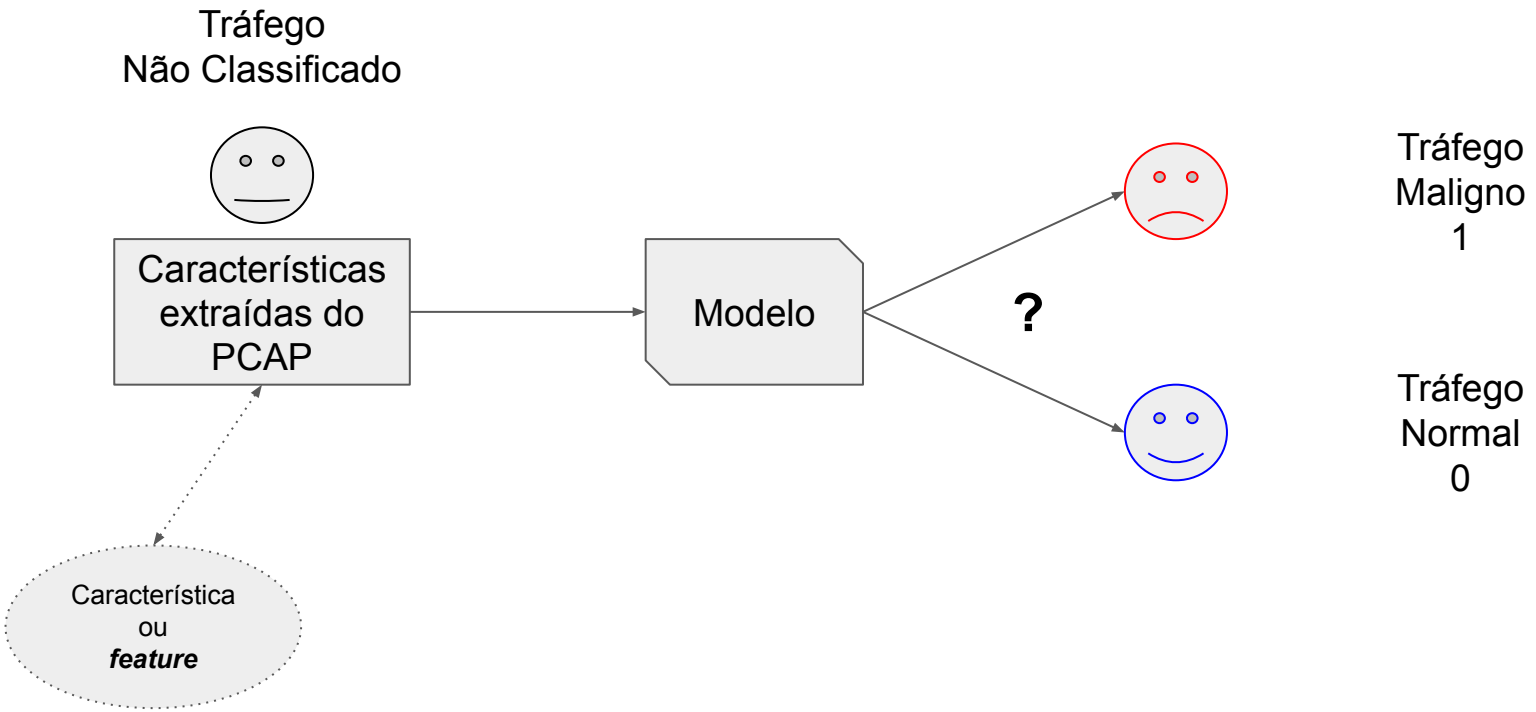
- Até 2023, o número de dispositivos conectados será maior que três vezes a população mundial (~ 3.6 dispositivos por usuário);
- Um aumento de 50% em relação ao valor de 2018 (~ 2.4 dispositivos por usuário).



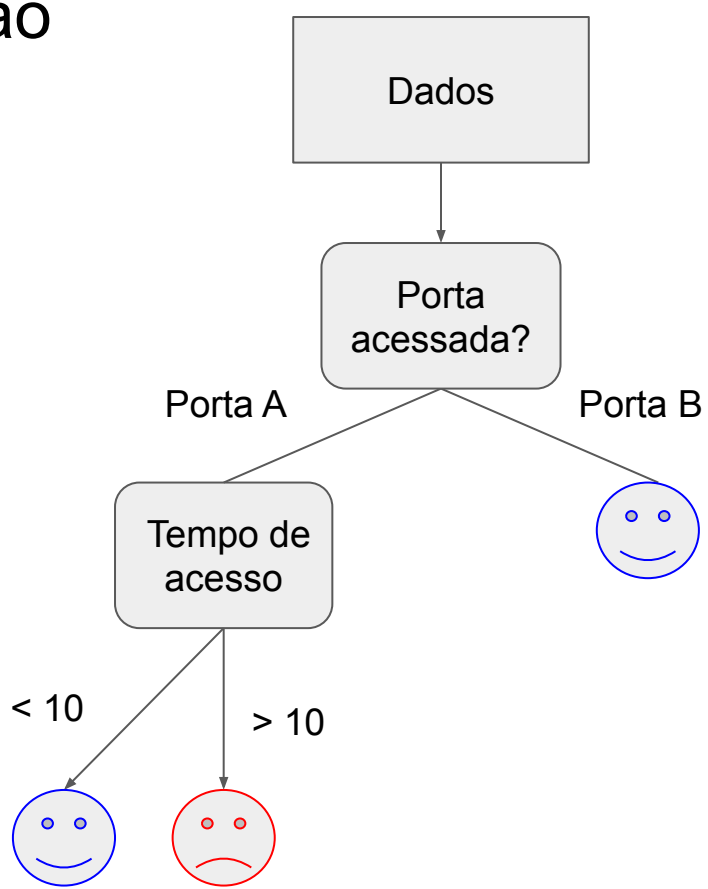
Os crimes cibernéticos evoluíram ao longo do tempo

PANDEY, K., KUMAR, V., PUNIA, D. Facing the Reality Of Cyber Threats In The Power Sector, Wipro Ltd. , 11 2013.

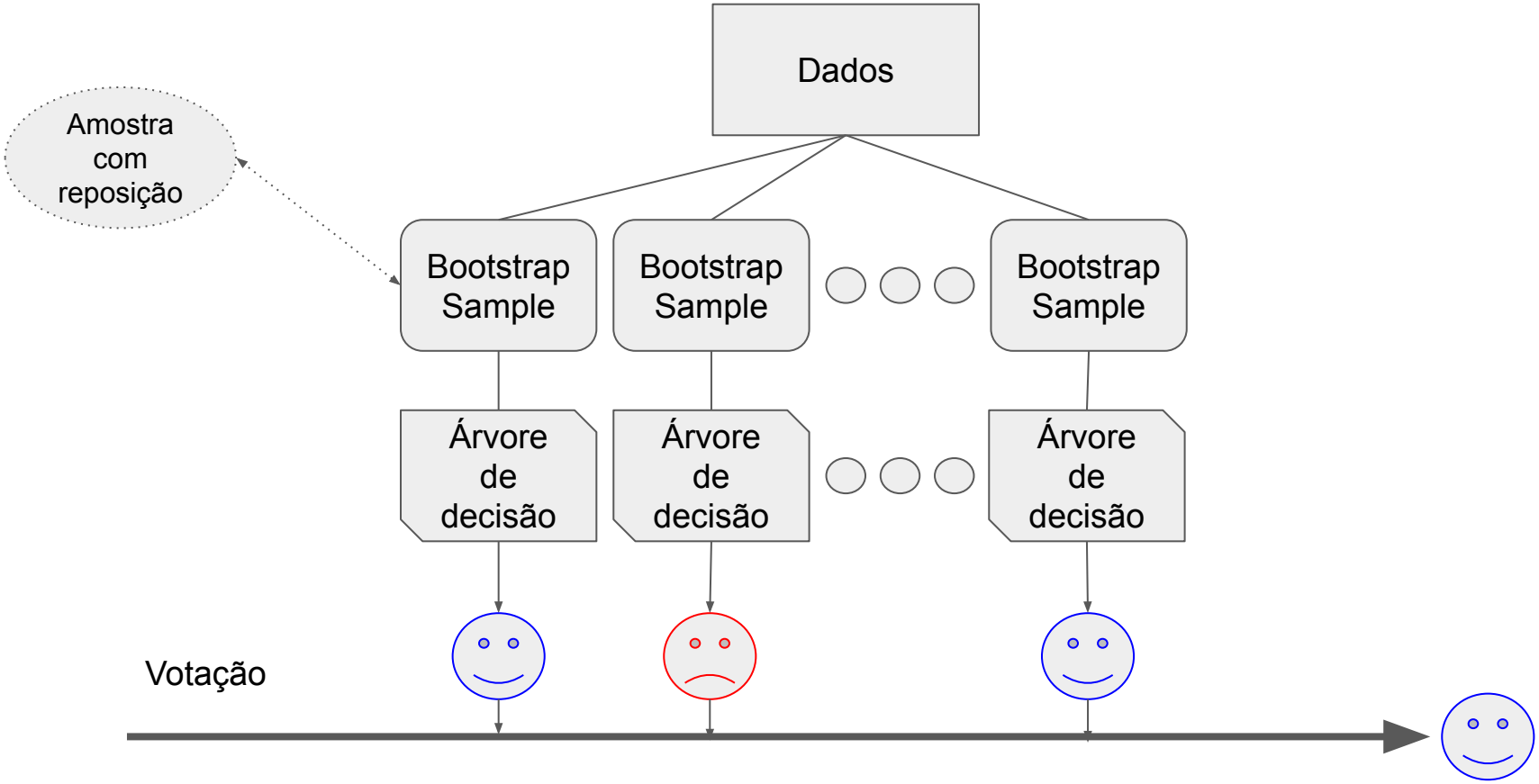
Objetivo: Predição de tráfego maligno (*botnets*)



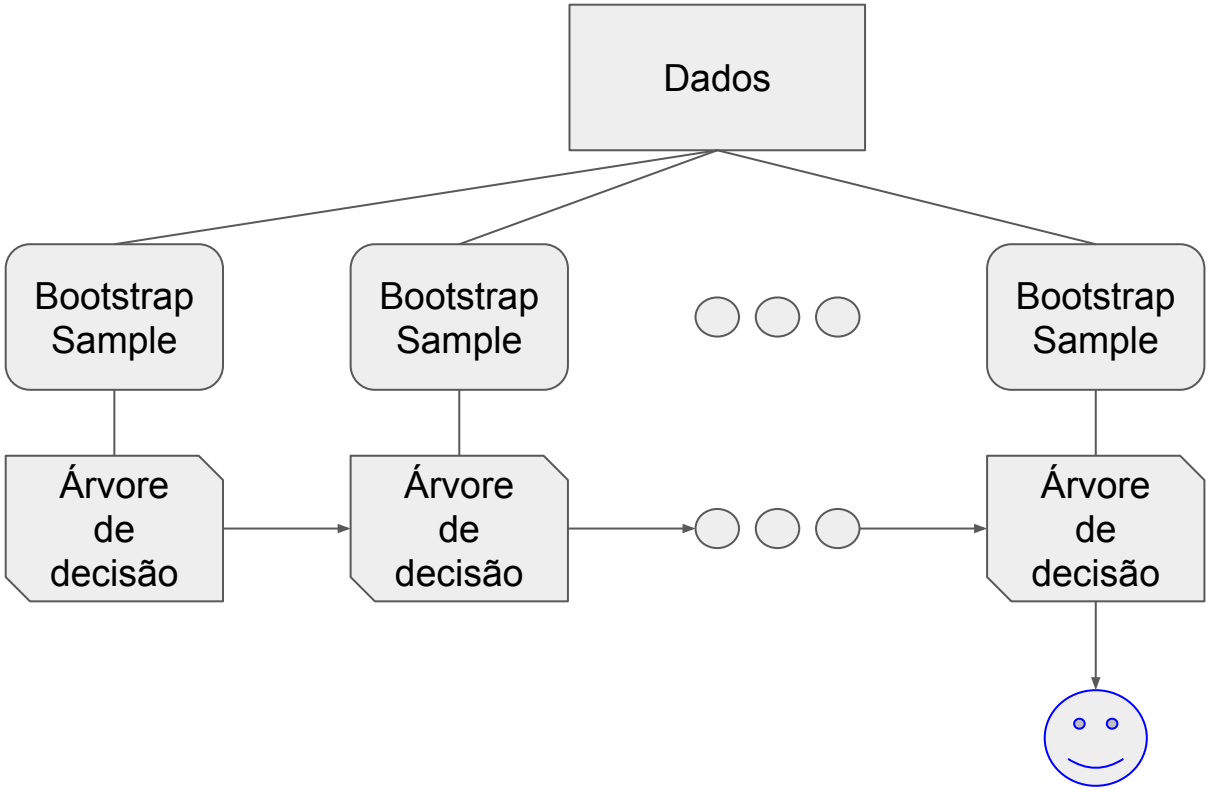
Árvore de decisão



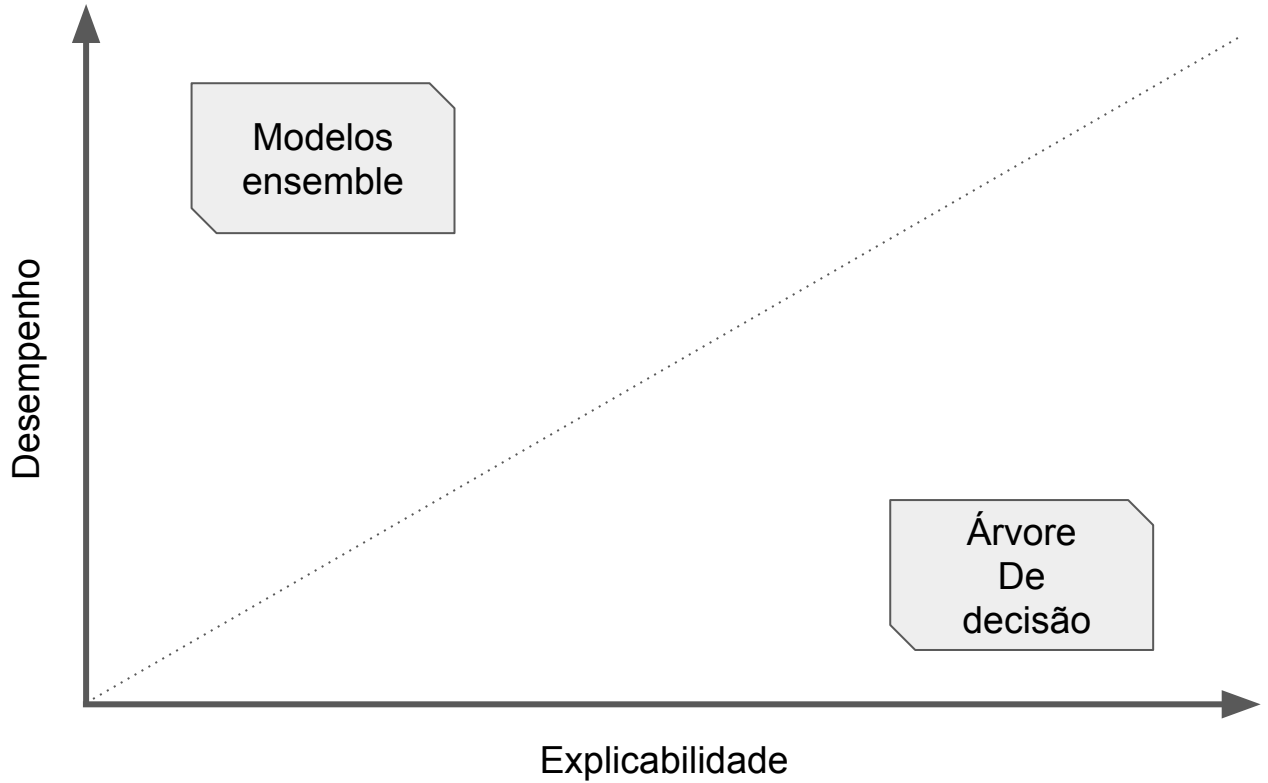
Floresta aleatória (Bagging Ensamble)



XGBoost (eXtreme Gradient Boosting - Boosting Ensamble)

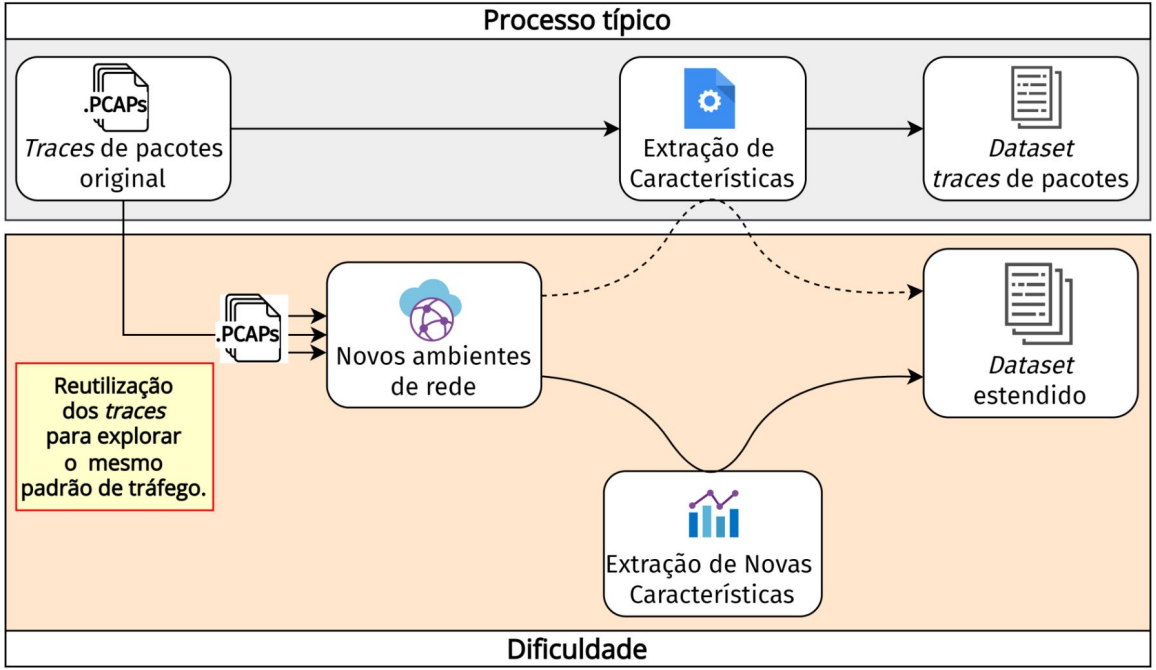


Capacidade de predição x explicabilidade



Datasets de trace

Figura 1 – Dificuldade em reproduzir *datasets* de *traces* de pacotes.



Fonte: Produzido pelo autor, 2022.

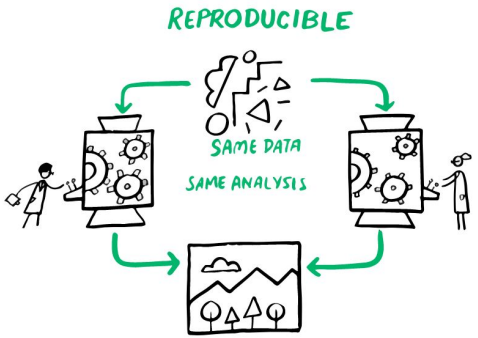
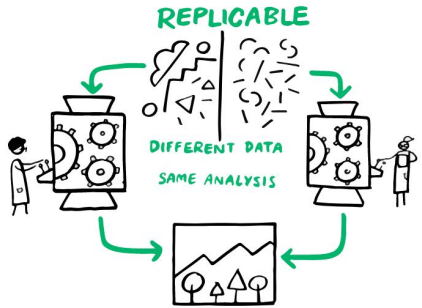
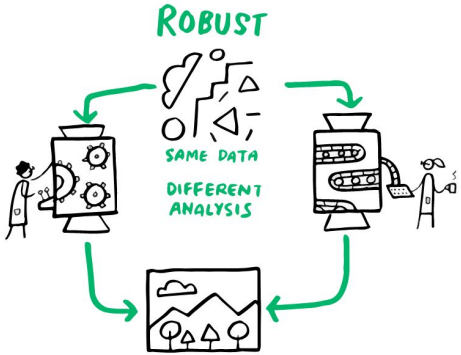
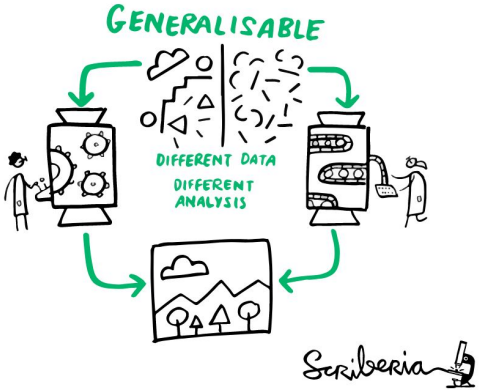
Elementos para a pesquisa aberta

- Dados abertos
- Código fonte aberto
- Hardware aberto
- Acesso aberto
- Cadernos abertos

The Turing Way: A Handbook for Reproducible Data Science

Dado

Análise

	Igual	**Diferente**
Igual		
Diferente		

The Turing Way: A Handbook for Reproducible Data Science

CTU-13

- Conjunto de *datasets* de tráfego de botnet capturado pela Universidade CTU em 2011.
- O objetivo é construir base de dados com uma grande captura de tráfego real contendo fluxo de botnet misturado com tráfego normal e de fundo.
- https://www.stratosphereips.org/datasets-ctu13
- "An empirical comparison of botnet detection methods" Sebastian Garcia, Martin Grill, Jan Stiborek and Alejandro Zunino. Computers and Security Journal, Elsevier. 2014. Vol 45, pp 100-123. http://dx.doi.org/10.1016/j.cose.2014.05.011

Lista de datasets (Cenários)

Table 2 – Characteristics of the botnet scenarios. (CF: ClickFraud, PS: Port Scan, FF: FastFlux, US: Compiled and controlled by us.)

Id	IRC	SPAM	CF	PS	DDoS	FF	P2P	US	HTTP	Note
1	√	√	√							
2	√	√	√							
3	√			√				√		
4	√				√			√		UDP and ICMP DDoS.
5		√		√					√	Scan web proxies.
6				√						Proprietary C&C. RDP.
7									√	Chinese hosts.
8				√						Proprietary C&C. Net-BIOS, STUN.
9	√	√	√	√						
10	√				√			√		UDP DDoS.
11	√				√			√		ICMP DDoS.
12							√			Synchronization.
13		√		√					√	Captcha. Web mail.

Table 2. Characteristics of botnet scenarios

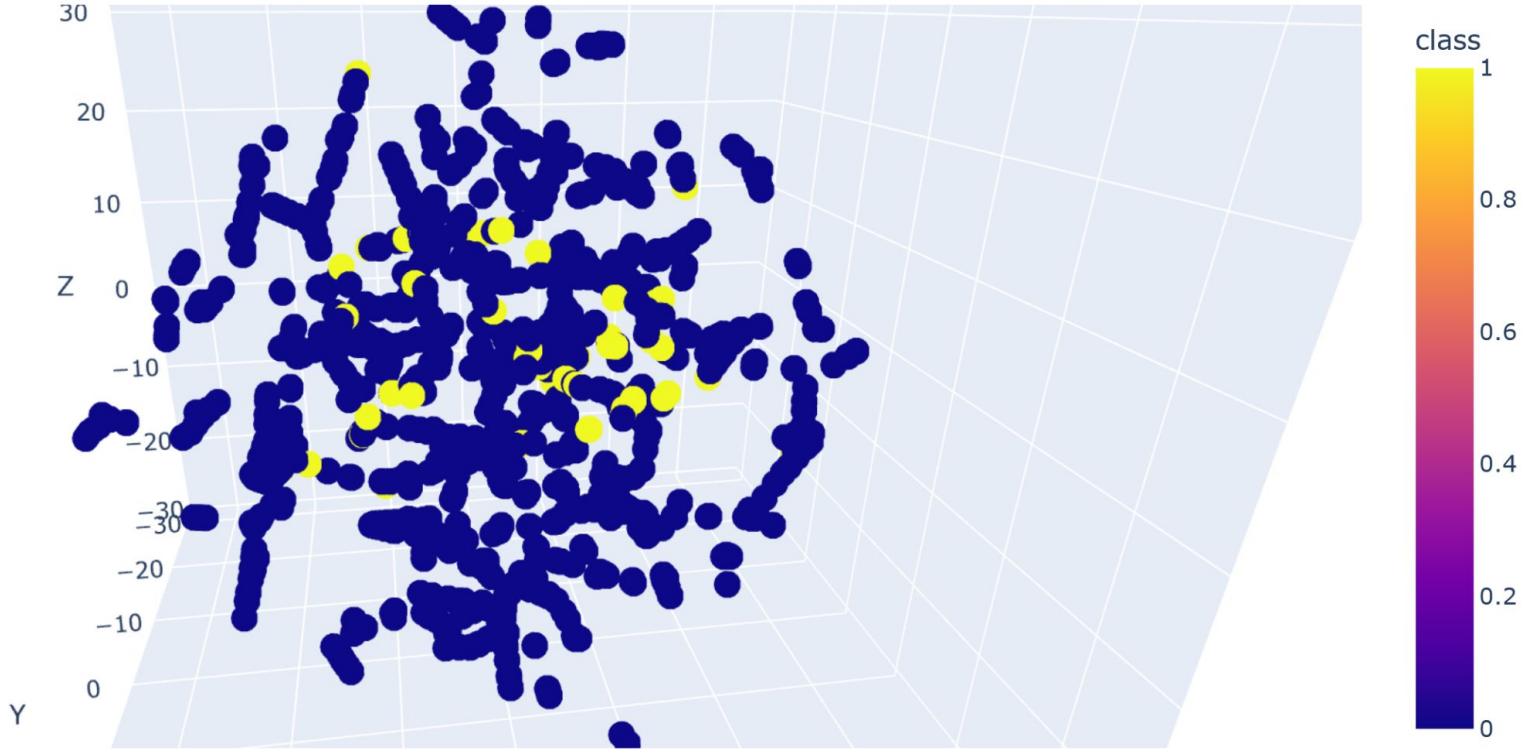
CTU-Malware-Capture-Botnet-48 (Id 7)

- Nome: Sogou
- Duração: 0 horas, 21 minutos e 0 segundos
- Binário utilizado: sogou_explorer_silent_1.4.0.418_2136.exe
- Host infectado
  - IP: 147.32.84.165
  - OS: Windows XP
  - English version Name: SARUMAN
  - Label: Botnet
- Número de features: 32
- Número de observações: 114077
- Alvo:
  - Tráfego normal 114014 (~ 0.99%)
  - Tráfego botnet 63 (~ 0.01%)

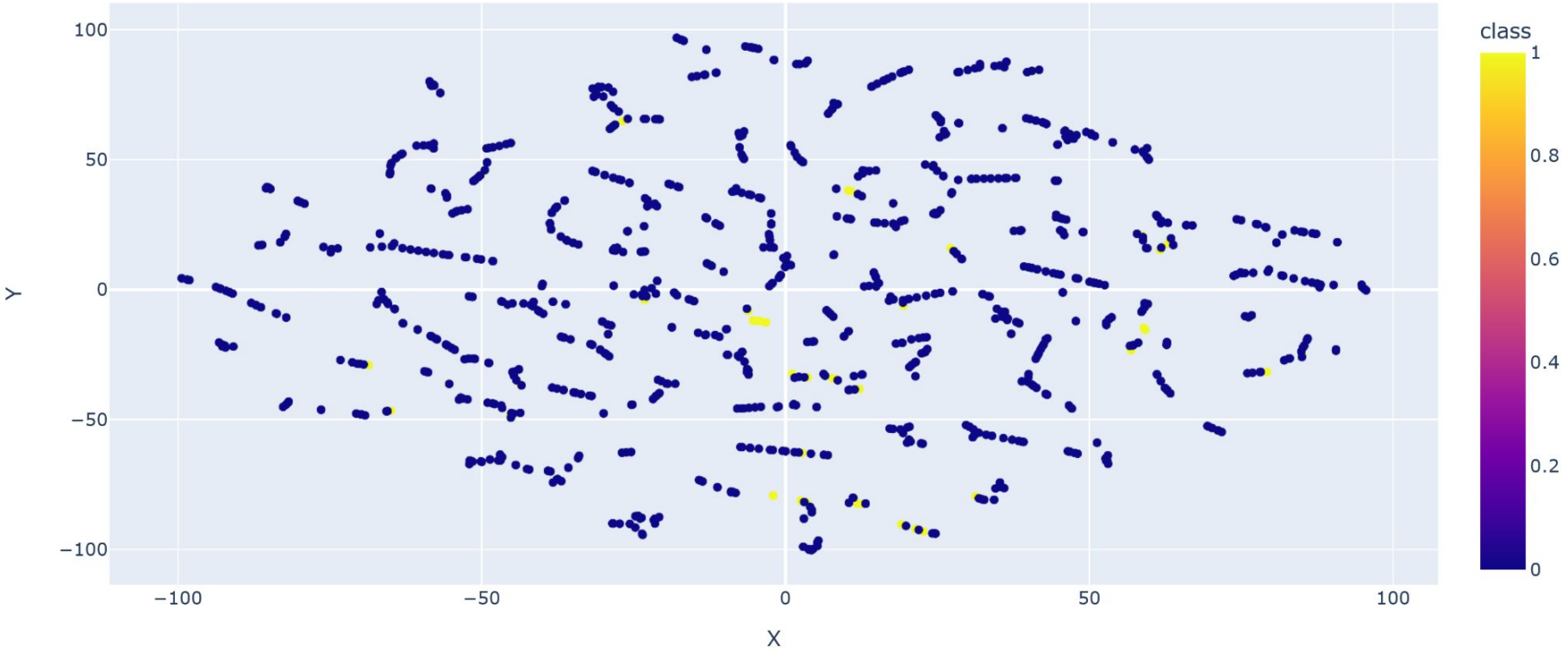
Features Selecionadas (24)

- Proto
- State
- sTos
- dTos
- sHops
- dHops
- sTtl
- dTtl
- TcpRtt
- SynAck
- AckDat
- SrcPkts
- DstPkts
- SrcBytes
- DstBytes
- SAppBytes
- DAppBytes
- Dur
- TotPkts
- TotBytes
- TotAppByte
- Rate
- SrcRate
- DstRate

Visualização 3D (t-SNE)



Visualização 2D (t-SNE)

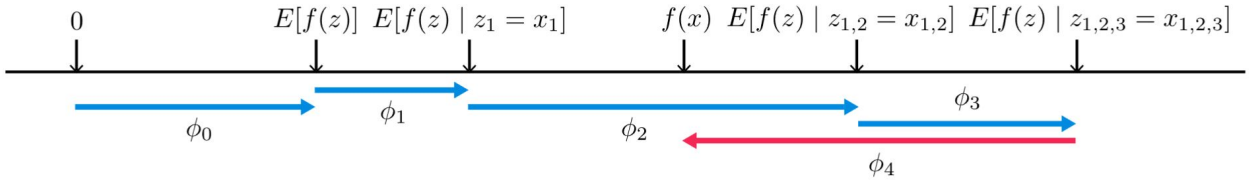


Resultados

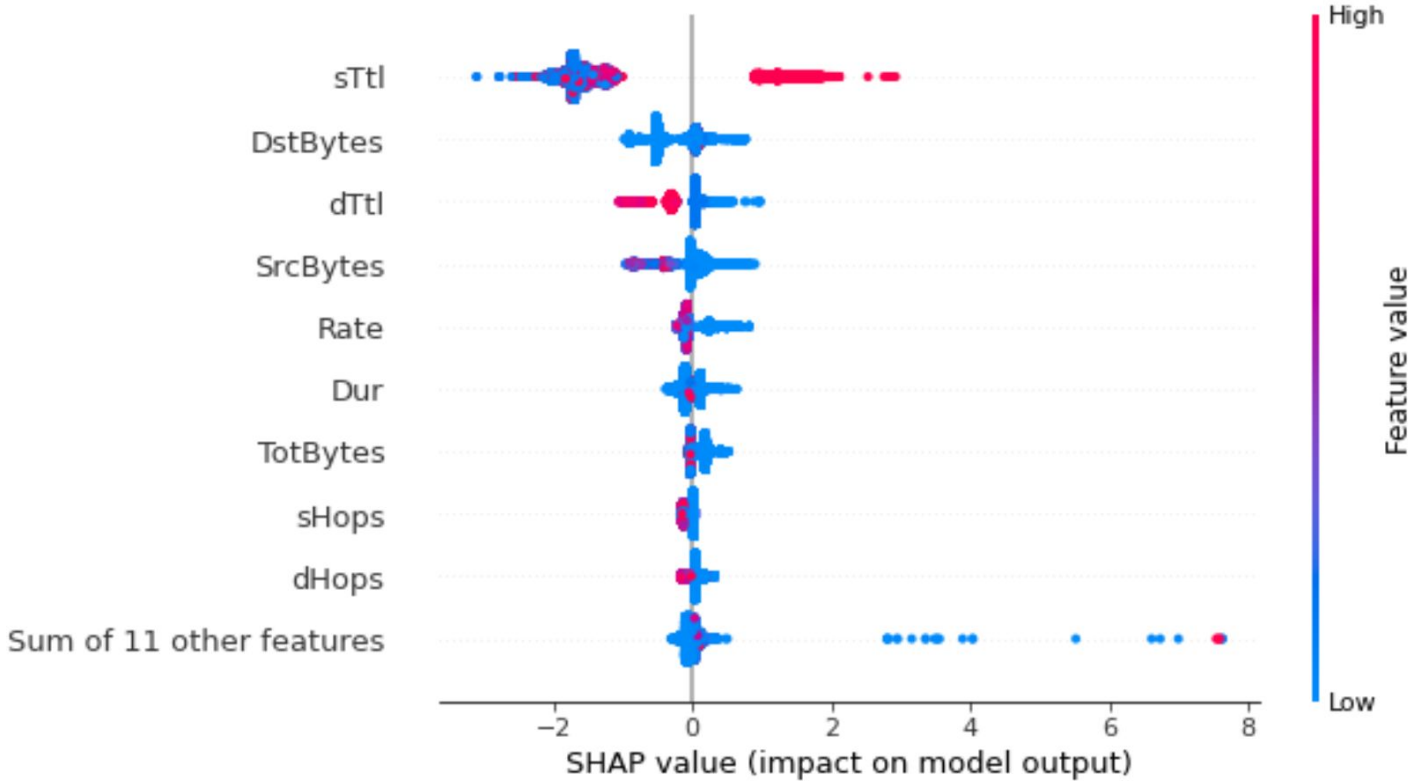
	precision	recall	f1-score	support
0	1.00	1.00	1.00	21274
1	1.00	0.60	0.75	15
accuracy			1.00	21289
macro avg	1.00	0.80	0.87	21289
weighted avg	1.00	1.00	1.00	21289

Interpretando modelos de aprendizado de máquinas, SHapley Additive exPlanations

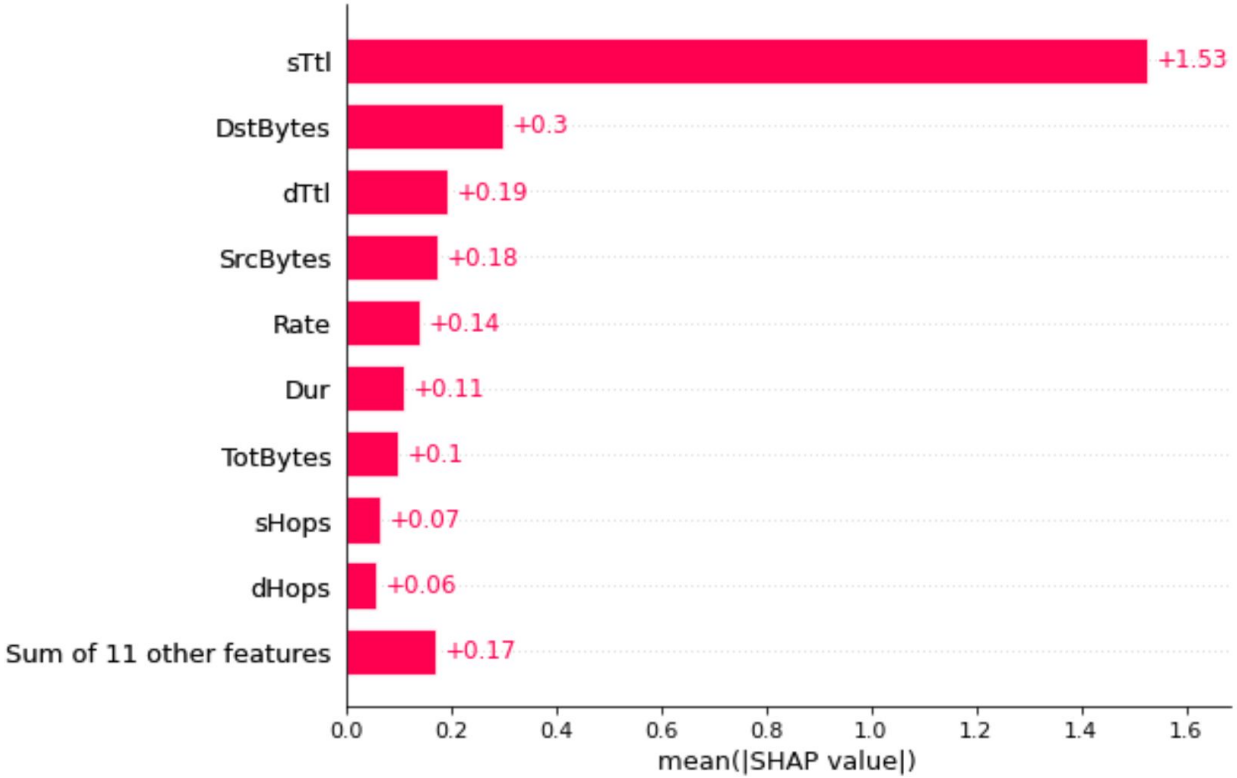
SHAP



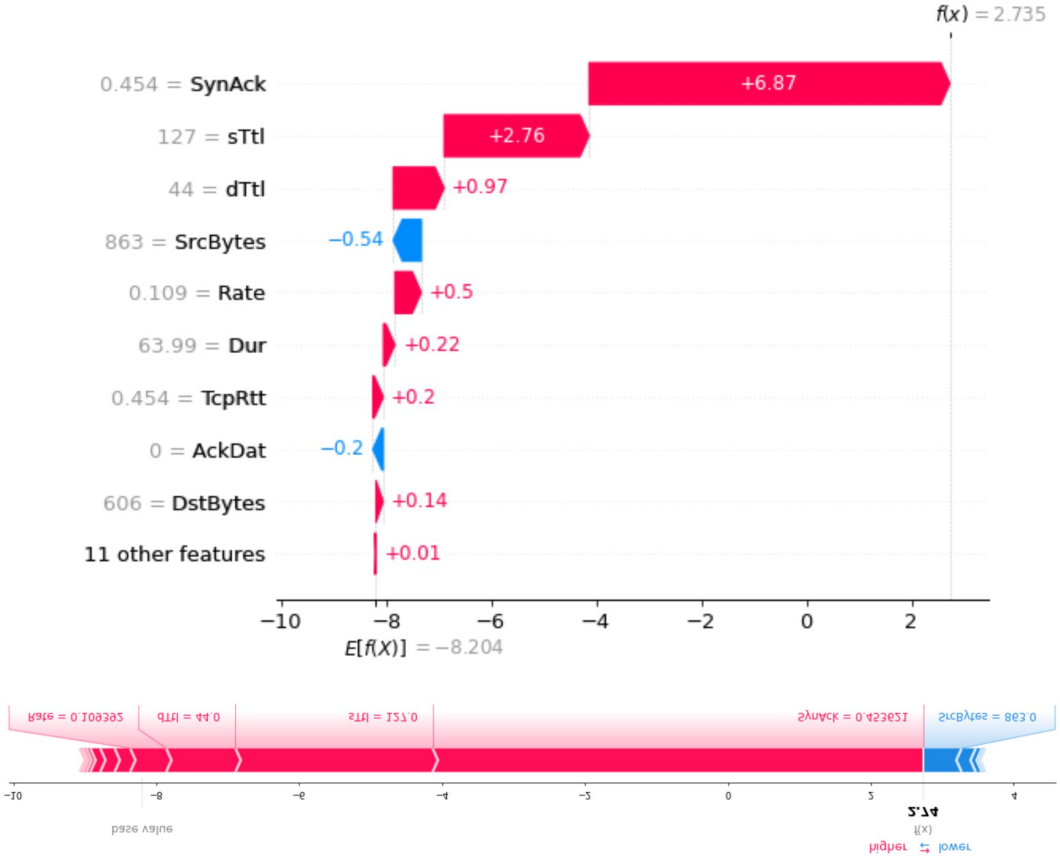
Análise dos resultados



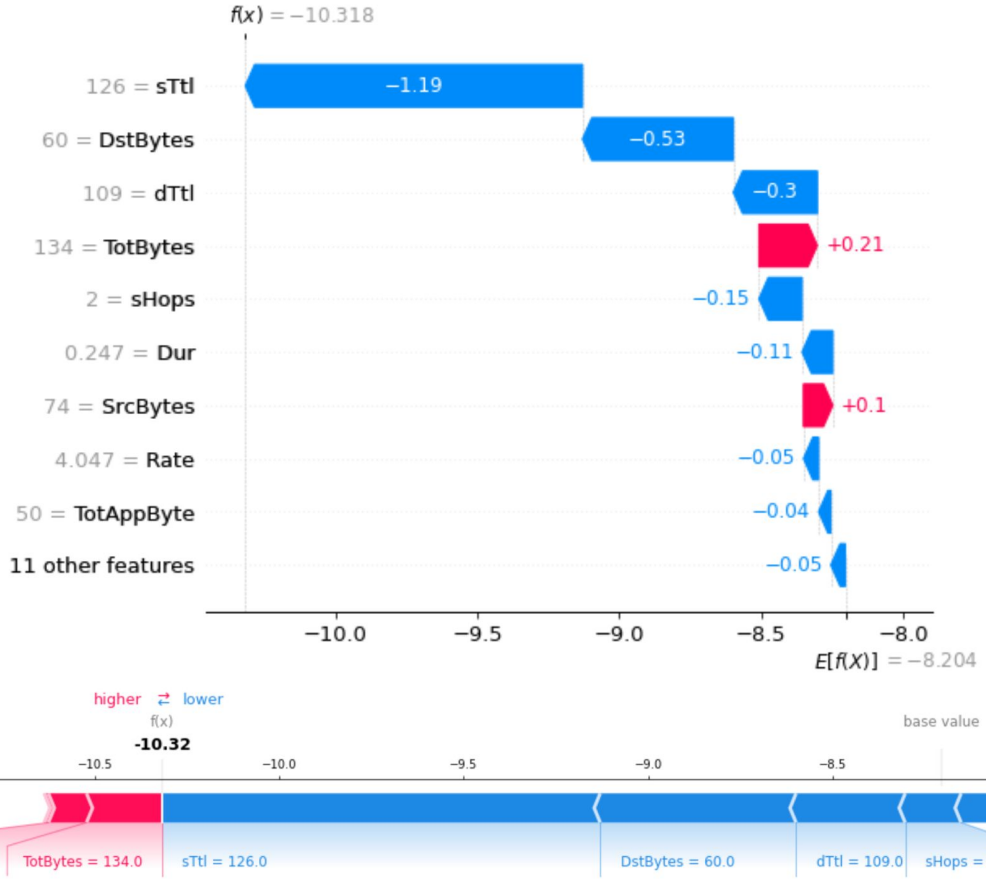
Análise dos resultados



Um exemplo de observação y = 1



Um exemplo para y = 0



1a tentativa de reprodução

- Aprendizado Profundo para a Predição de Ataques de Negação de Serviço Distribuído
- https://sol.sbc.org.br/index.php/sbrc/article/view/21191/21016
- Utilizar os PCAPs oriundos dos do portal Stratosphere Lab
- Injeção do tráfego utilizando o TCPdump

1a tentativa de reprodução - Resultados

- Não foi possível reproduzir o experimento do artigo.
  - Causas
    - Falta de informações sobre o experimento tanto no artigo quanto no repositório do dataset gerado.
    - Problema de sincronização e retransmissão.

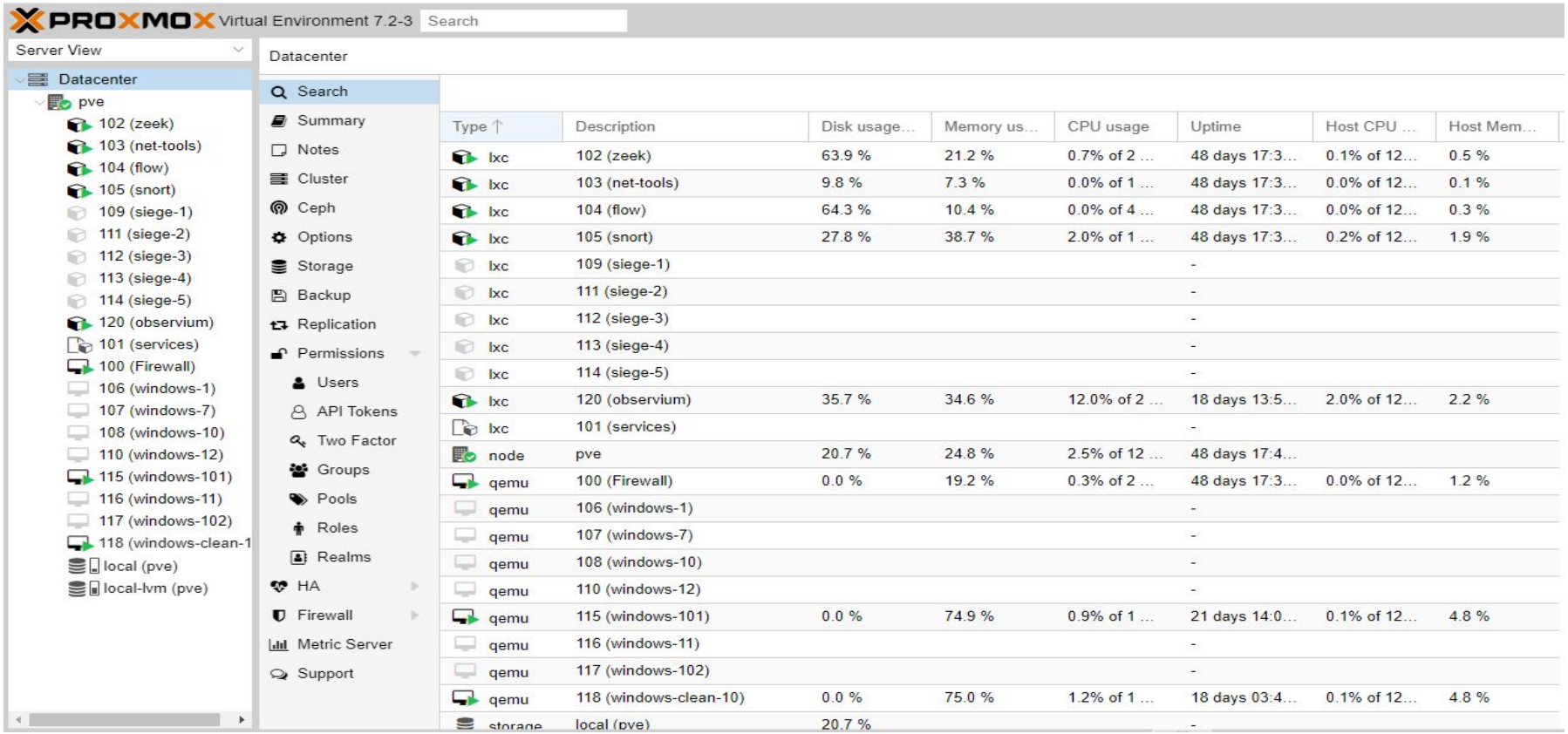
2a tentativa

- Reproduzir o cenário proposto por Sebastian Garcia, Martin Grill, Jan Stiborek and Alejandro Zunino no artigo.
  - "An empirical comparison of botnet detection methods" Sebastian Garcia, Martin Grill, Jan Stiborek and Alejandro Zunino. Computers and Security Journal, Elsevier. 2014. Vol 45, pp 100-123. http://dx.doi.org/10.1016/j.cose.2014.05.011
- Criação de um novo dataset com as mesma features
  - A coleta durou 13 horas e 06 minutos
  - Produzindo um total de 2.212.874 de fluxos
    - Normal: 2.205.550 (99,67%)
    - Botnet: 7.324 (0,33%)

Cenário do Ambiente - Laboratório LabNerds

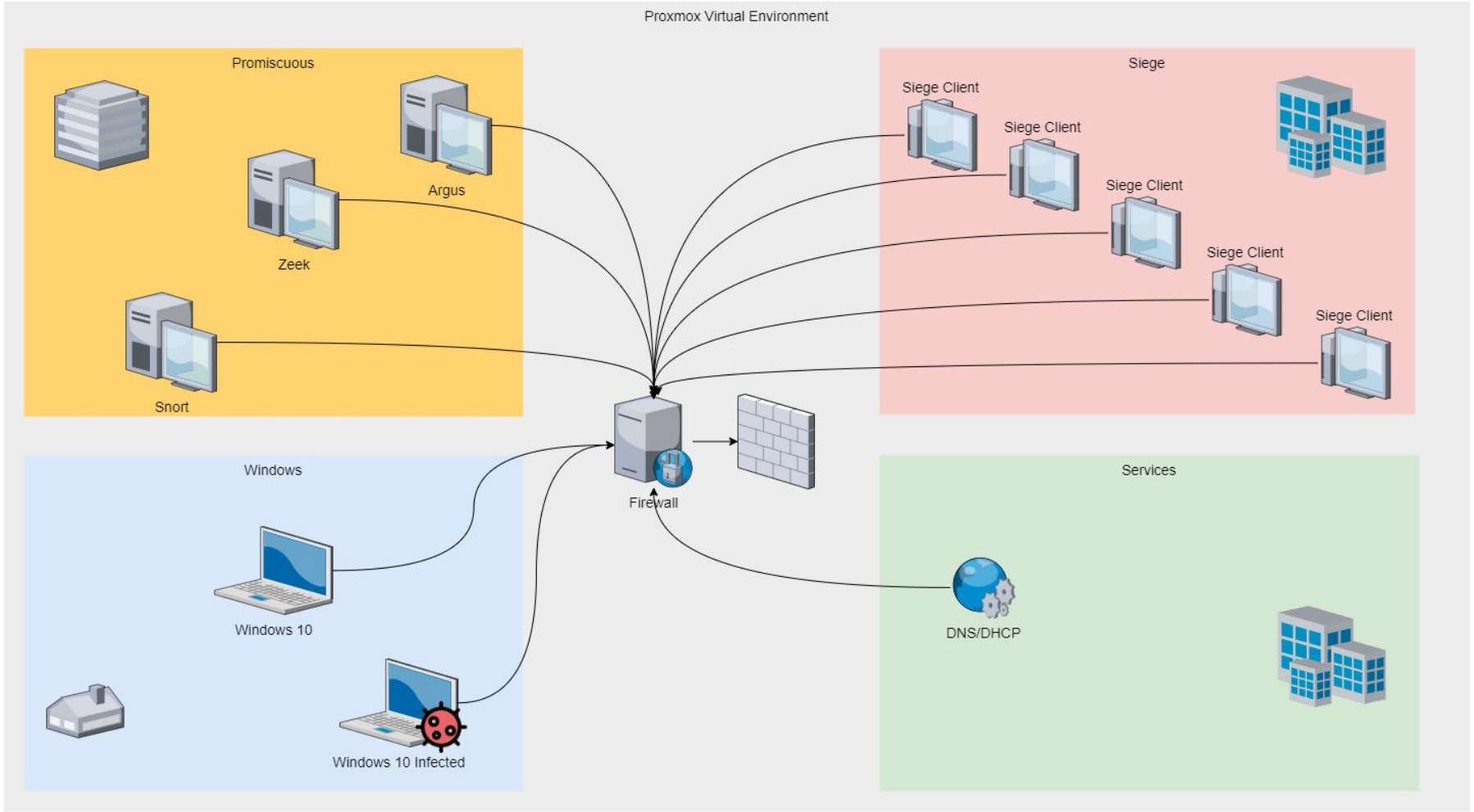
- Dell PowerEdge T430
  - Processador Intel Xeon E5-2620
  - Memória RAM 32GB
  - Armazenamento de 1TB
- Proxmox Virtual Environment
  - KVM hypervisor
  - Linux Containers (LXC)

Proxmox VE


PROXMOX Virtual Environment 7.2-3
Search
Server View
Datacenter
pve
102 (zeek)
103 (net-tools)
104 (flow)
105 (snort)
109 (siege-1)
111 (siege-2)
112 (siege-3)
113 (siege-4)
114 (siege-5)
120 (observium)
101 (services)
100 (Firewall)
106 (windows-1)
107 (windows-7)
108 (windows-10)
110 (windows-12)
115 (windows-101)
116 (windows-11)
117 (windows-102)
118 (windows-clean-1
local (pve)
local-lvm (pve)
Datacenter
Search
Summary
Notes
Cluster
Ceph
Options
Storage
Backup
Replication
Permissions
Users
API Tokens
Two Factor
Groups
Pools
Roles
Realms
HA
Firewall
Metric Server
Support
Type
Description
Disk usage...
Memory us...
CPU usage
Uptime
Host CPU ...
Host Mem...
lxc 102 (zeek) 63.9 % 21.2 % 0.7% of 2 ... 48 days 17:3... 0.1% of 12... 0.5 %
lxc 103 (net-tools) 9.8 % 7.3 % 0.0% of 1 ... 48 days 17:3... 0.0% of 12... 0.1 %
lxc 104 (flow) 64.3 % 10.4 % 0.0% of 4 ... 48 days 17:3... 0.0% of 12... 0.3 %
lxc 105 (snort) 27.8 % 38.7 % 2.0% of 1 ... 48 days 17:3... 0.2% of 12... 1.9 %
lxc 109 (siege-1) -
lxc 111 (siege-2) -
lxc 112 (siege-3) -
lxc 113 (siege-4) -
lxc 114 (siege-5) -
lxc 120 (observium) 35.7 % 34.6 % 12.0% of 2 ... 18 days 13:5... 2.0% of 12... 2.2 %
lxc 101 (services) -
node pve 20.7 % 24.8 % 2.5% of 12 ... 48 days 17:4...
qemu 100 (Firewall) 0.0 % 19.2 % 0.3% of 2 ... 48 days 17:3... 0.0% of 12... 1.2 %
qemu 106 (windows-1) -
qemu 107 (windows-7) -
qemu 108 (windows-10) -
qemu 110 (windows-12) -
qemu 115 (windows-101) 0.0 % 74.9 % 0.9% of 1 ... 21 days 14:0... 0.1% of 12... 4.8 %
qemu 116 (windows-11) -
qemu 117 (windows-102) -
qemu 118 (windows-clean-10) 0.0 % 75.0 % 1.2% of 1 ... 18 days 03:4... 0.1% of 12... 4.8 %
storage local (pve) 20.7 % -

Cenário Virtual

- No experimento foram utilizadas 12 máquinas virtuais.
  - Firewall - SO Rocky Linux 9
  - Zeek - Rocky Linux 9
  - DNS/DHCP - Rocky Linux 9
  - Snort - Rocky Linux 9
  - Argus - Rocky Linux 9
  - 5 Maquinas com Siege - Rocky Linux 9
  - Windows 10
  - Windows 10 - infectada com a botnet Sogou


Proxmox Virtual Environment
Promiscuous
Argus
Zeek
Snort
Siege
Siege Client
Siege Client
Siege Client
Siege Client
Siege Client
Firewall
Windows
Windows 10
Windows 10 Infected
Services
DNS/DHCP

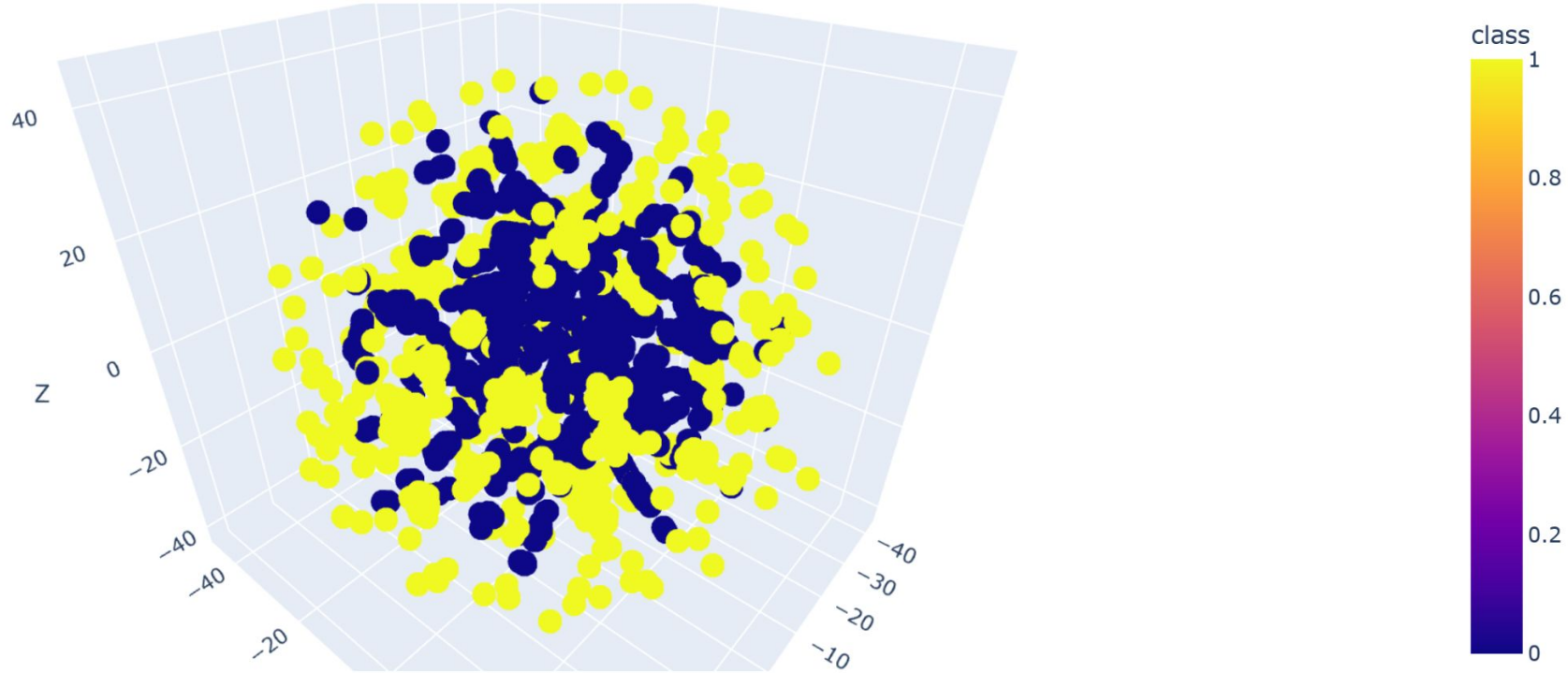
argus

PROXMOX

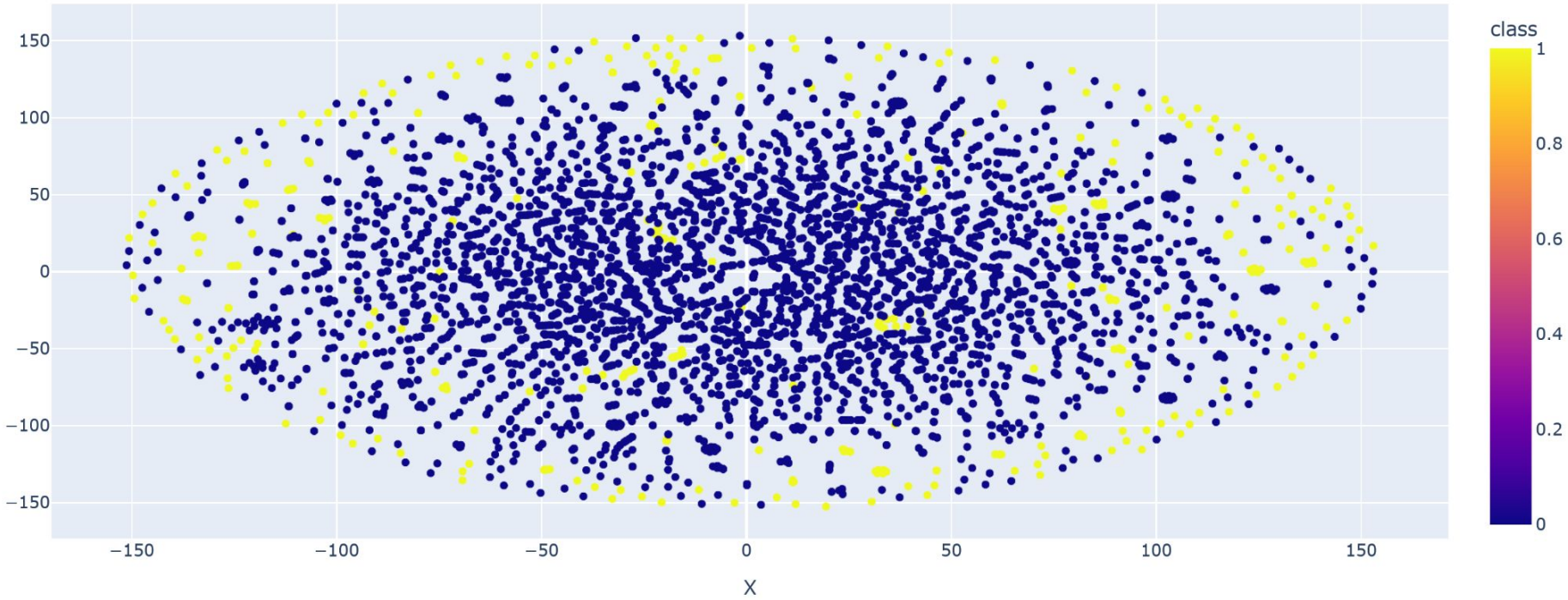
Descrição das aplicações.

- Firewall para garantir a segurança e a segmentação do ambiente
- Zeek para monitorar todo o tráfego do ambiente
- DNS e DHCP para prover infraestrutura e o funcionamento da rede
- Argus para gerar os fluxos de redes
- Snort para analisar o tráfego da botnet com assinaturas já existentes
- Siege para simular o tráfego normal

Visualização 3D (t-SNE)



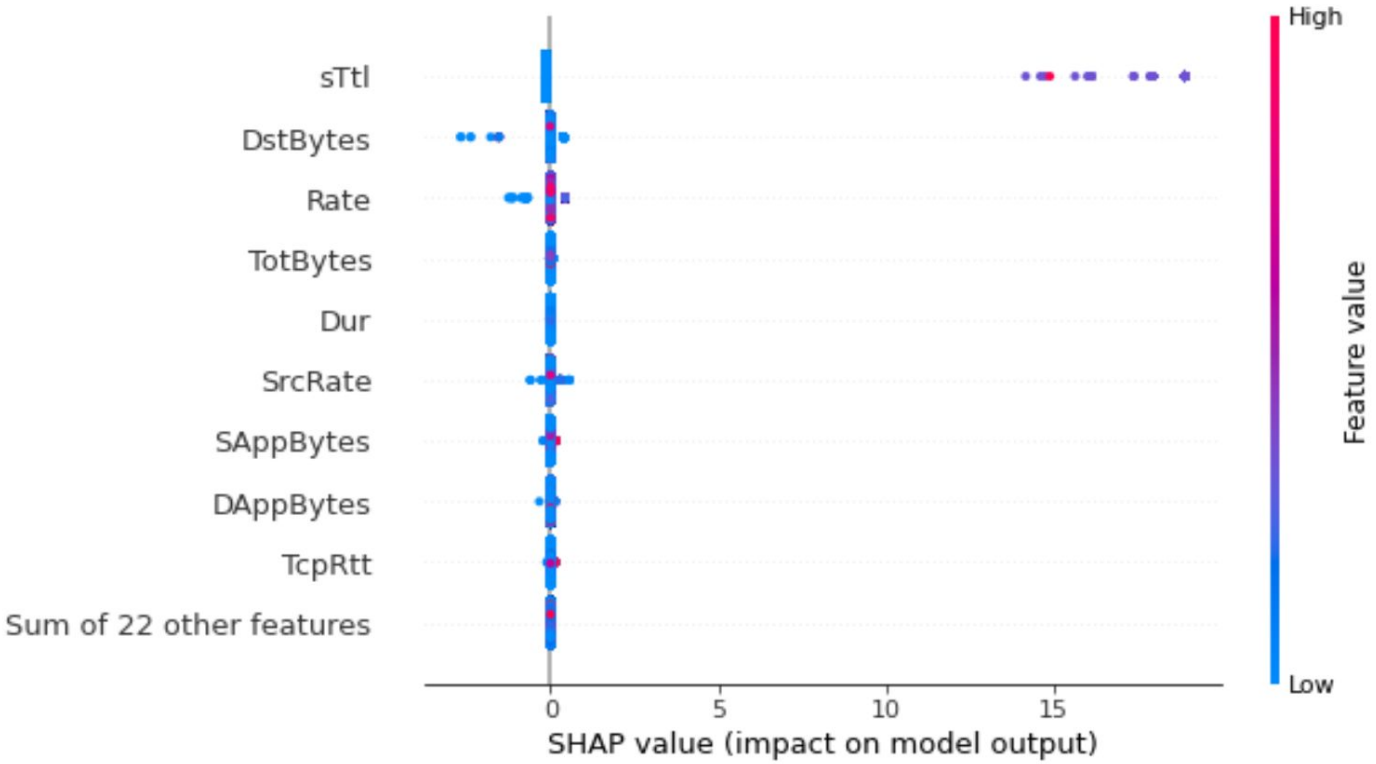
Visualização 2D (t-SNE)



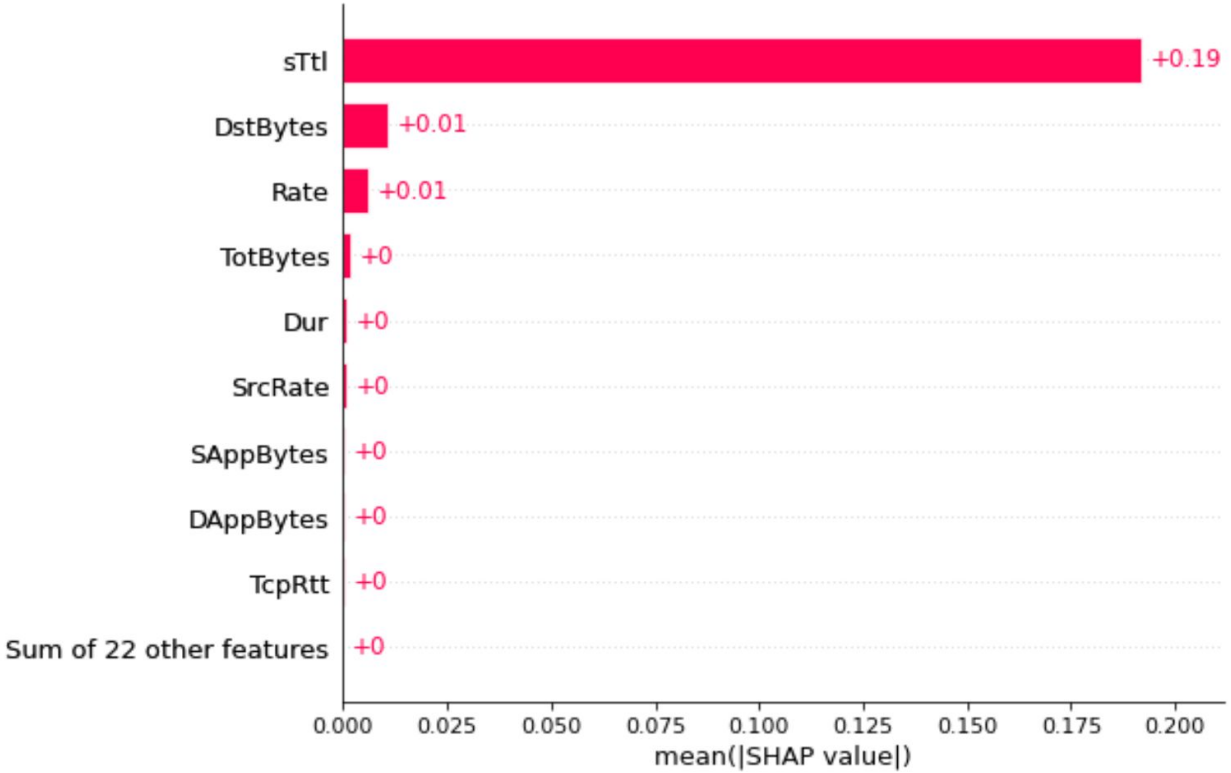
Resultados

	precision	recall	f1-score	support
0	1.00	1.00	1.00	440370
1	1.00	1.00	1.00	1396
accuracy			1.00	441766
macro avg	1.00	1.00	1.00	441766
weighted avg	1.00	1.00	1.00	441766

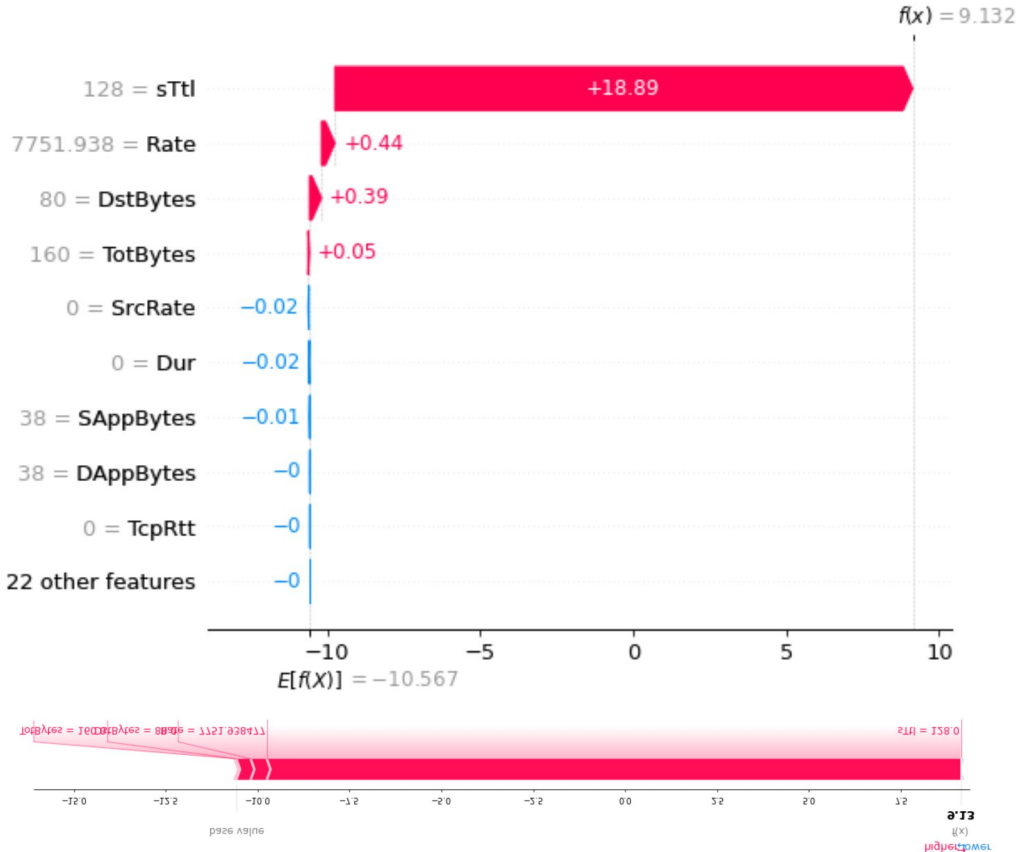
Análise dos resultados



Análise dos resultados



Um exemplo de observação y = 1



Um exemplo para y = 0