Uma Arquitetura baseada em Inteligência Artificial Explicável (XAI) para Sistemas de Detecção de Intrusões em Smart Grids

Camilla B. Quincozes, Henrique C. Oliveira, **Silvio E. Quincozes**, Rodrigo S. Miani, Vagner E. Quincozes

Qual é o papel da Inteligência Artificial Explicável (XAI) na Detecção de Intrusões?

Introdução

- **Subestações digitais: elemento central em Smart Grids**
  - Transmissão e distribuição de energia elétrica até os consumidores.
- **Comunicação em subestações:**
  - A norma IEC-61850 integra a **rede elétrica**, **computação** e **comunicação**.
    - **Propõe protocolos que facilitam a comunicação entre IEDs.**

Introdução

- **Subestações digitais: elemento central em Smart Grids**
  - Transmissão e distribuição de energia elétrica até os consumidores.
- **Comunicação em subestações:**
  - A norma IEC-61850 integra a **rede elétrica**, **computação** e **comunicação**.
    - **Propõe protocolos que facilitam a comunicação entre IEDs.**

> ⚠ **Consequência**
>
> A rede elétrica ficou vulnerável a ataques cibernéticos!

Protocolo GOOSE

- **GOOSE (Generic Object Oriented Substation Event)**
  - É um protocolo definido pelo padrão IEC-61850 para a troca de mensagens sobre eventos que acontecem na subestação.
    - **Abertura e fechamento de disjuntores para interromper ou restabelecer a transmissão de energia.**

      ⚠️ **Alvo em potencial para ataques!**

Motivação

“As empresas de água e energia gastaram 400% mais, em 2023, para se recuperarem após ataques cibernéticos.”

Motivação

“As empresas de água e energia gastaram 400% mais, em 2023, para se recuperarem após ataques cibernéticos.”

Incêndio em transformador na subestação Monacillo, em San Juan (Porto Rico)

Motivação

“As empresas de água e energia gastaram 400% mais, em 2023, para se recuperarem após ataques cibernéticos.”

Aconteceu logo após um ataque DDoS e afetou 500 mil pessoas.

Introdução

- **Como lidar com ataques cibernéticos?**
  - **Medidas de proteção (preventivas):**
    - **em cenários tradicionais, é comum o uso de mecanismos de autenticação, firewalls, criptografia, etc.**

Introdução

- **Como lidar com ataques cibernéticos?**
  - **Medidas de proteção (preventivas):**
    - em cenários tradicionais, é comum o uso de mecanismos de autenticação, firewalls, criptografia, etc.
    - **Esses mesmos mecanismos não aplicáveis em subestações digitais!**
      - ✓ Requisitos de tempo real, atacante interno, etc.

Introdução

- **Como lidar com ataques cibernéticos?**
  - **Medidas de proteção (preventivas):**
    - em cenários tradicionais, é comum o uso de mecanismos de autenticação, firewalls, criptografia, etc.
    - **Esses mesmos mecanismos não aplicáveis em subestações digitais!**
      - ✓ **Requisitos de tempo real, atacante interno, etc.**

A detecção de intrusões se torna fundamental!

Problema de Pesquisa

- **Como detectar intrusões?**
  - **Há diferentes abordagens de detecção:** baseada em assinaturas, anomalias, baseada em regras de domínio...
    - **O estado da arte é o uso de Inteligência Artificial (IA)**
      - ⚠️ **Porém, na maioria dos casos, não compreendemos adequadamente as decisões tomadas pela IA!**

Proposta

- **Arquitetura “X-IDS”**
  - Arquitetura **Explicável** para Sistemas de Detecção de Intrusões
- **Investigação da Explicabilidade**
  - Random Forest
  - XGBoost
  - Decision Tree
- **Proposta e avaliação de novas features**
  - Estudo do seu impacto no X-IDS!

Trabalhos Relacionados

Referência	Domínio	Ataques
[Wang et al. 2020]	Redes Corporativas	Tradicionais
[Sivamohan et al. 2023]	Sistemas Ciberfísicos	Tradicionais
[Kuzlu et al. 2020]	Geração Fotovoltaica	Não se aplica
[Munir et al. 2023]	Recursos Energéticos	Tradicionais
[Zolanvari et al. 2021]	IIoT	Tradicionais
[Dresch et al. 2024]	Redes Intra-veiculares	Especializados
Este trabalho	**Subestações Elétricas**	**Especializados**

Proposta


Geração de Assinaturas de Ataque
Entradas
Interface Web
Entradas
Gerador de Tráfego
Conjunto de Dados
Usuário
Explicabilidade
Shap
Detecção de Intrusões
XGBoost
Decision Tree
Resultados

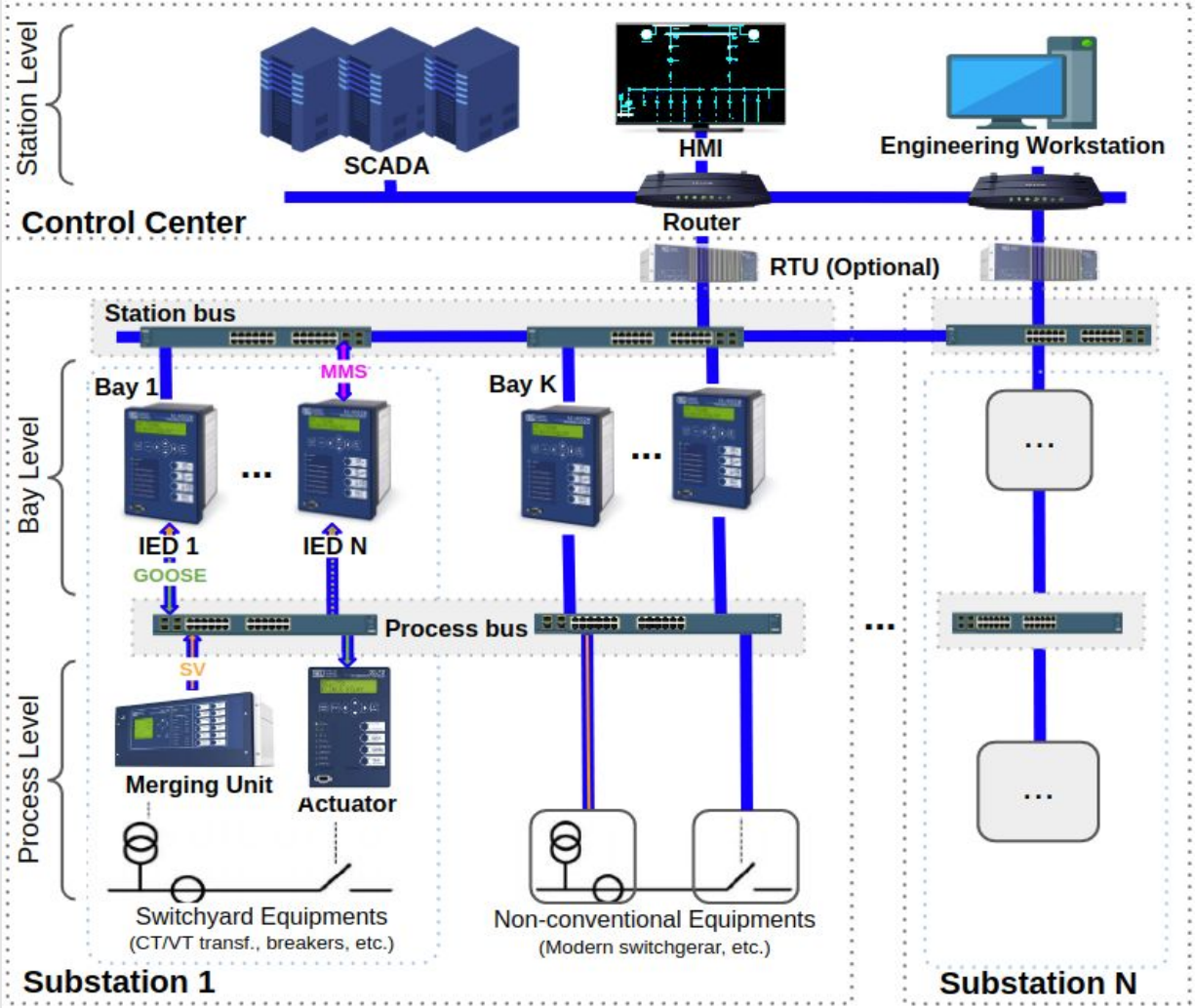
Cenário

Subestações Elétricas Digitais

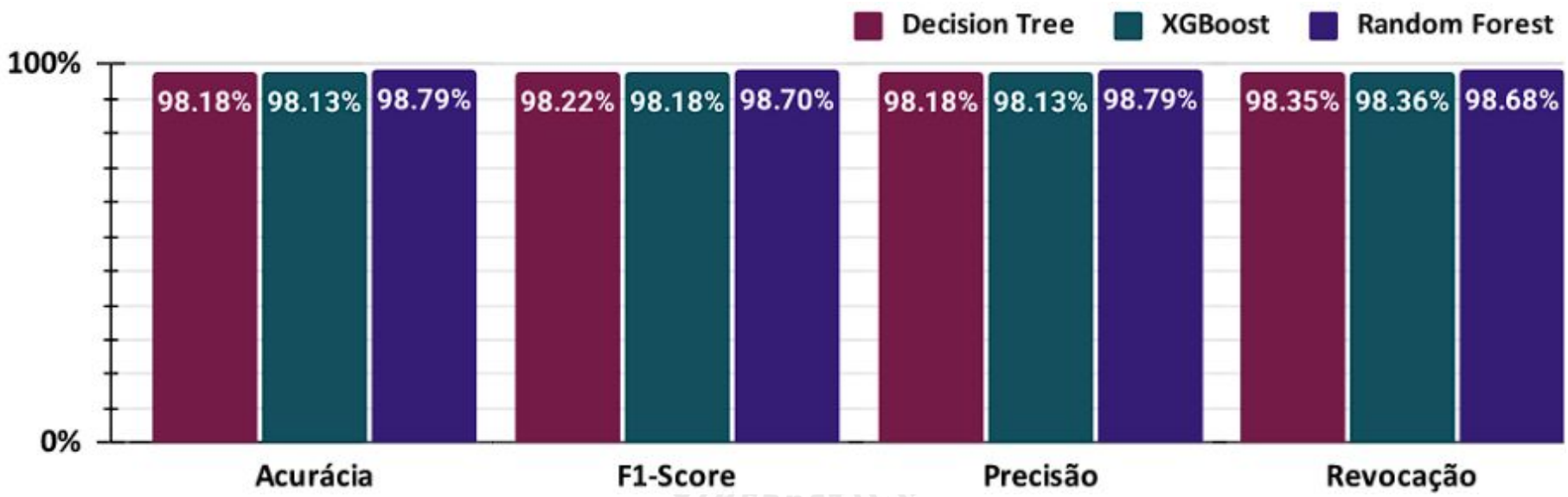
ERENO

Gerador de Datasets

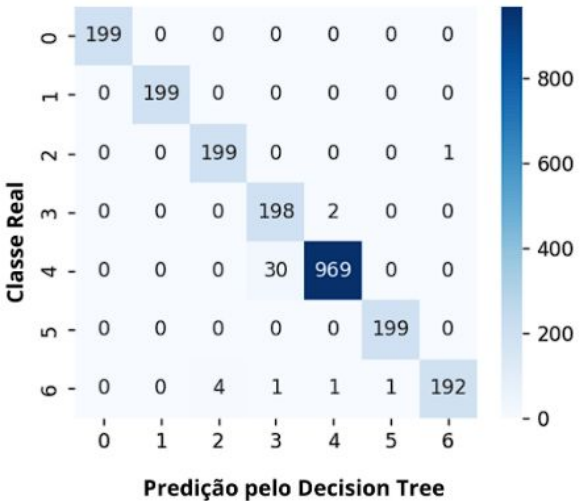
Sete classes de ataques ao protocolo GOOSE de subestações digitais.

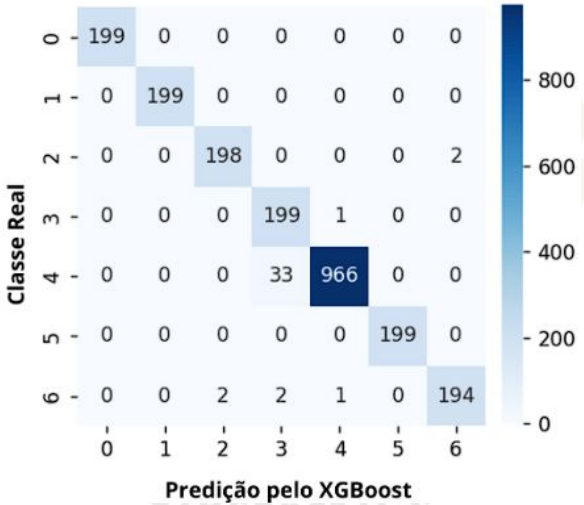


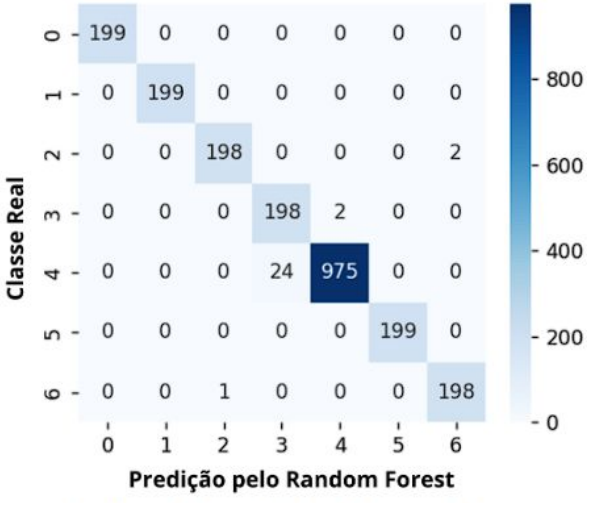
Resultados

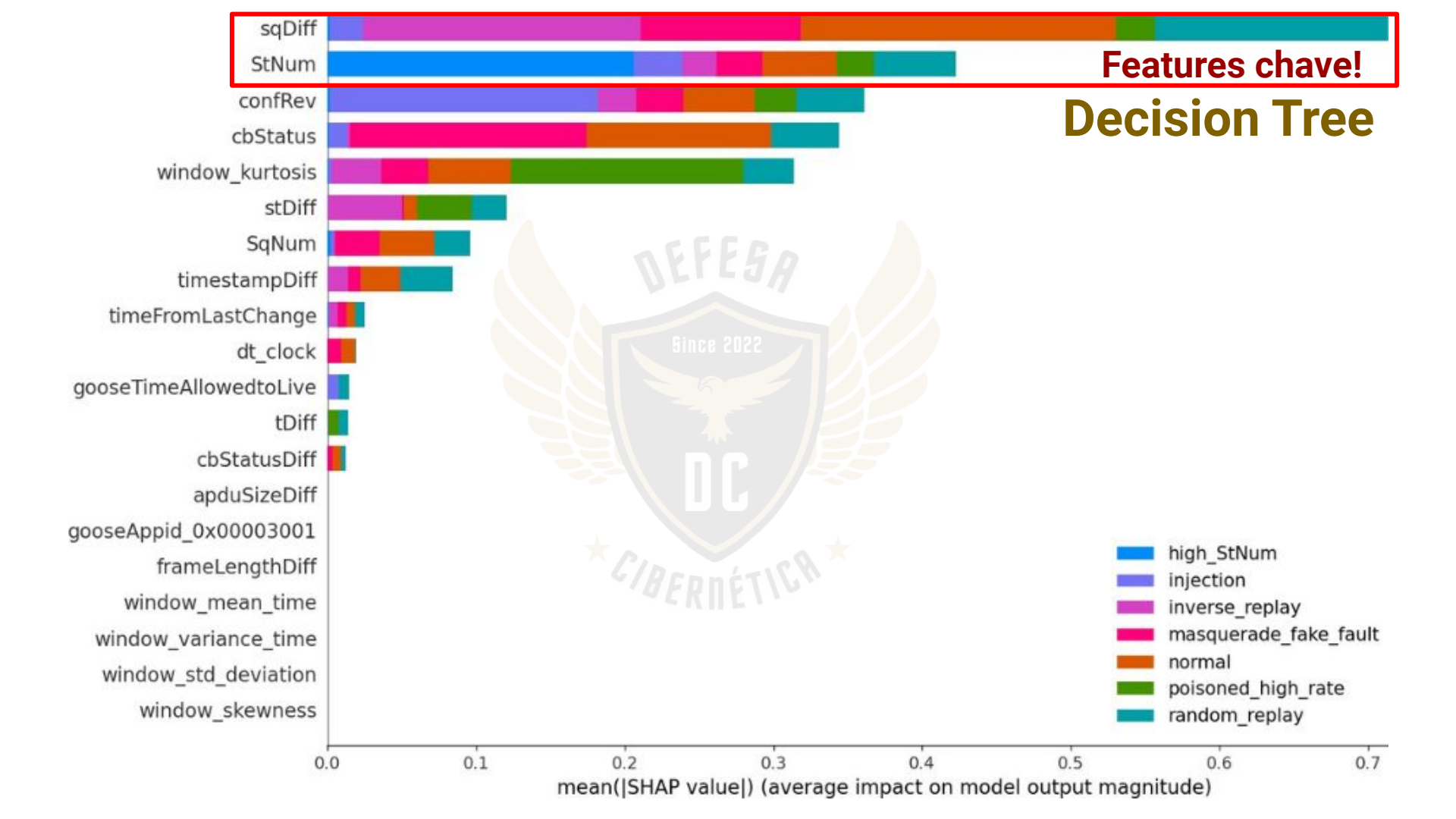

Decision Tree
XGBoost
Random Forest
100%
0%
98.18%
98.13%
98.79%
98.22%
98.18%
98.70%
98.18%
98.13%
98.79%
98.35%
98.36%
98.68%
Acurácia
F1-Score
Precisão
Revocação

Resultados

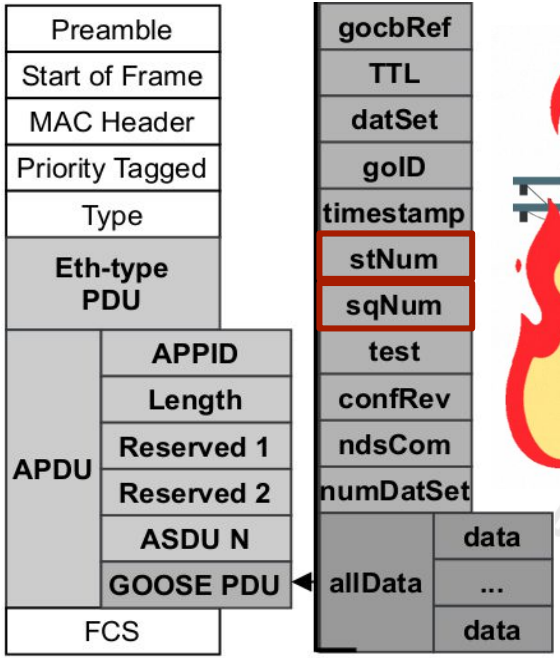

Classe Real
Predição pelo Decision Tree
0 1 2 3 4 5 6
199 0 0 0 0 0 0
0 199 0 0 0 0 0
0 0 199 0 0 0 1
0 0 0 198 2 0 0
0 0 0 30 969 0 0
0 0 0 0 0 199 0
0 0 4 1 1 1 192
800 600 400 200 0

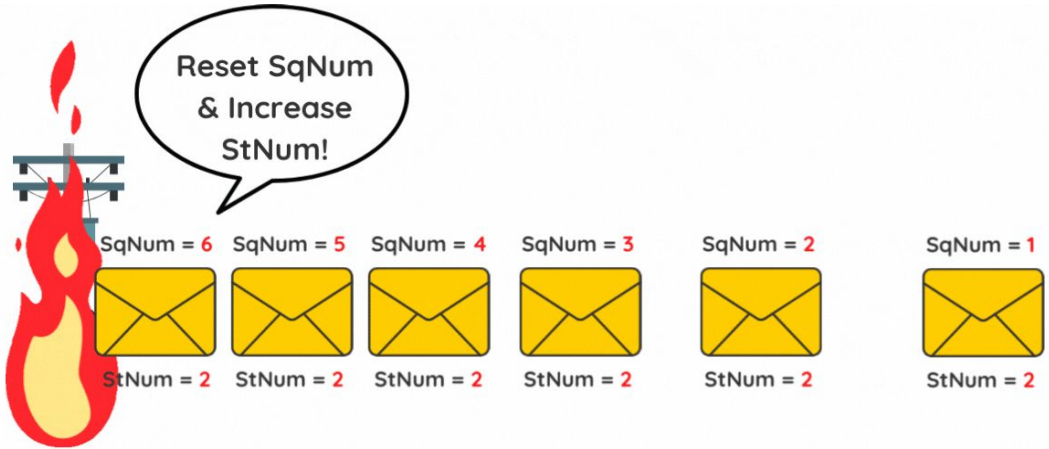

Classe Real
Predição pelo XGBoost
0 1 2 3 4 5 6
199 0 0 0 0 0 0
0 199 0 0 0 0 0
0 0 198 0 0 0 2
0 0 0 199 1 0 0
0 0 0 33 966 0 0
0 0 0 0 0 199 0
0 0 2 2 1 0 194
800 600 400 200 0


Classe Real
Predição pelo Random Forest
0 1 2 3 4 5 6
199 0 0 0 0 0 0
0 199 0 0 0 0 0
0 0 198 0 0 0 2
0 0 0 198 2 0 0
0 0 0 24 975 0 0
0 0 0 0 0 199 0
0 0 1 0 0 0 198
800 600 400 200 0

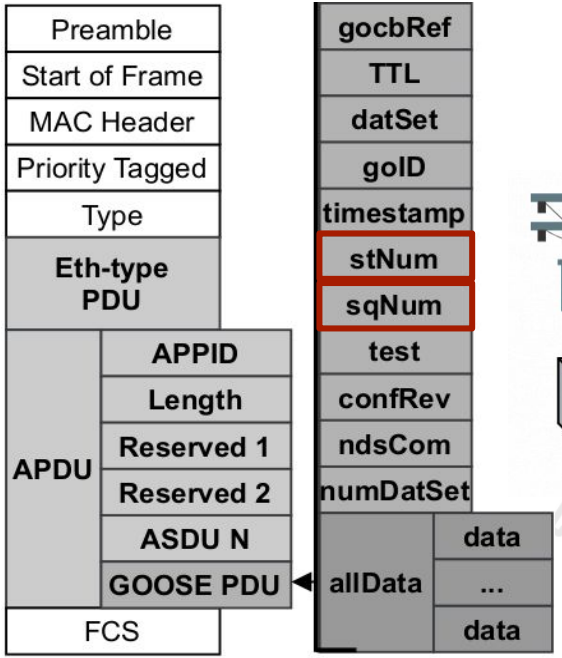

sqDiff
StNum
confRev
cbStatus
window_kurtosis
stDiff
SqNum
timestampDiff
timeFromLastChange
dt_clock
gooseTimeAllowedtoLive
tDiff
cbStatusDiff
apduSizeDiff
gooseAppid_0x00003001
frameLengthDiff
window_mean_time
window_variance_time
window_std_deviation
window_skewness
Features chave!
Decision Tree
high_StNum
injection
inverse_replay
masquerade_fake_fault
normal
poisoned_high_rate
random_replay
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
mean(|SHAP value|) (average impact on model output magnitude)

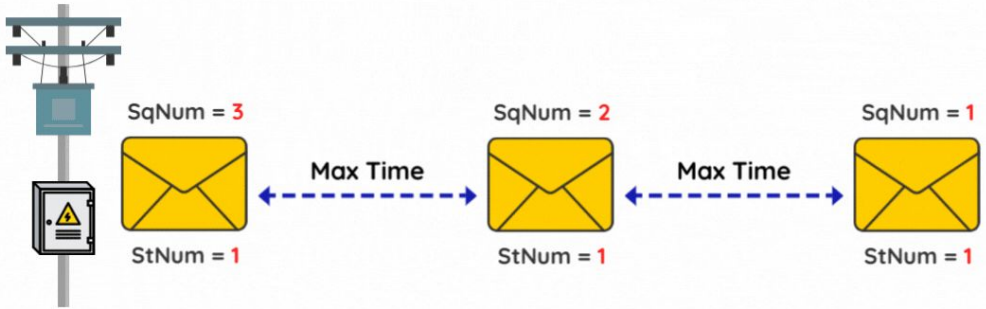
Resultados

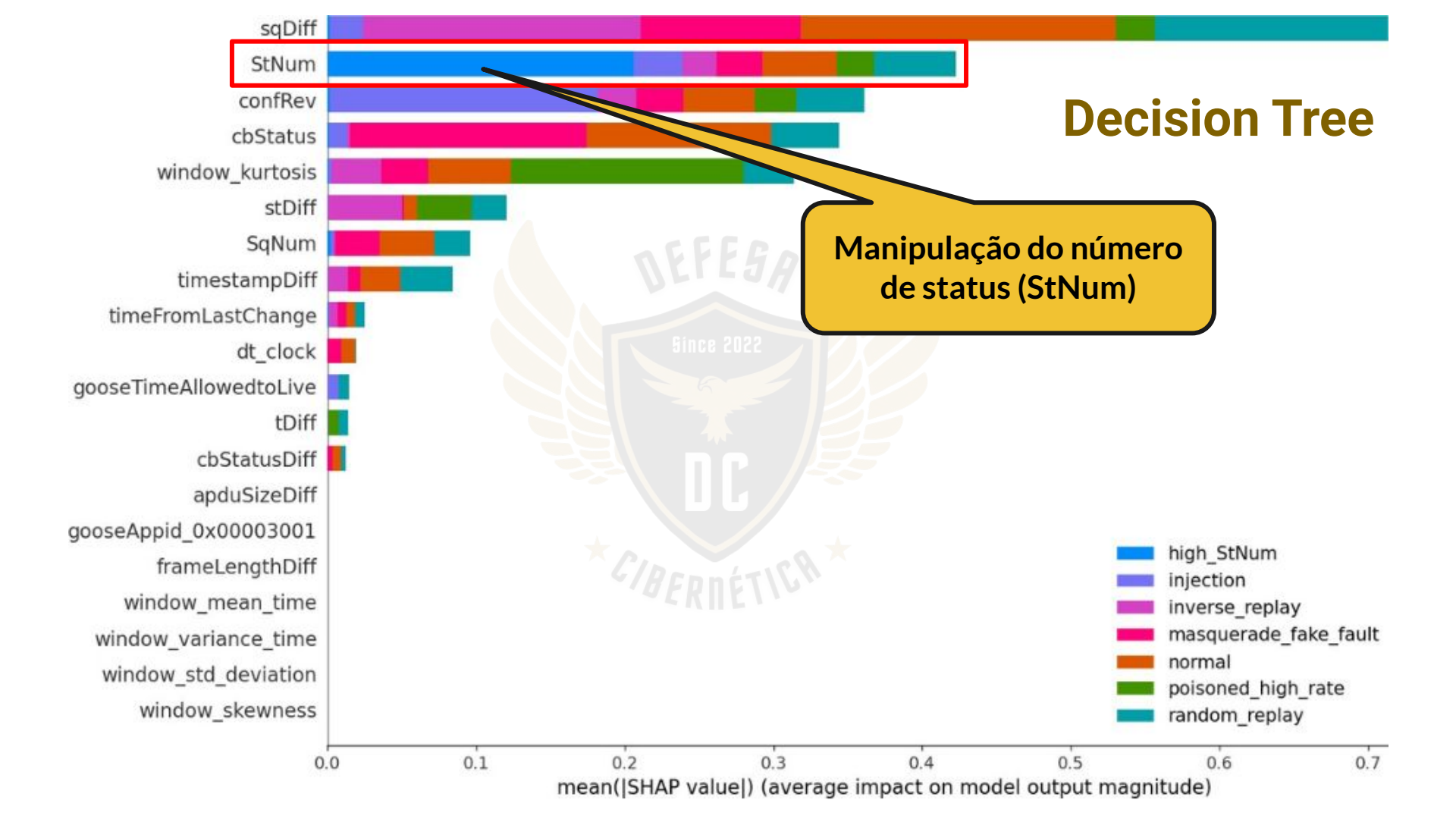

Preamble
Start of Frame
MAC Header
Priority Tagged
Type
Eth-type PDU
APDU
APPID
Length
Reserved 1
Reserved 2
ASDU N
GOOSE PDU
FCS
gocbRef
TTL
datSet
goID
timestamp
stNum
sqNum
test
confRev
ndsCom
numDatSet
allData
data
...
data

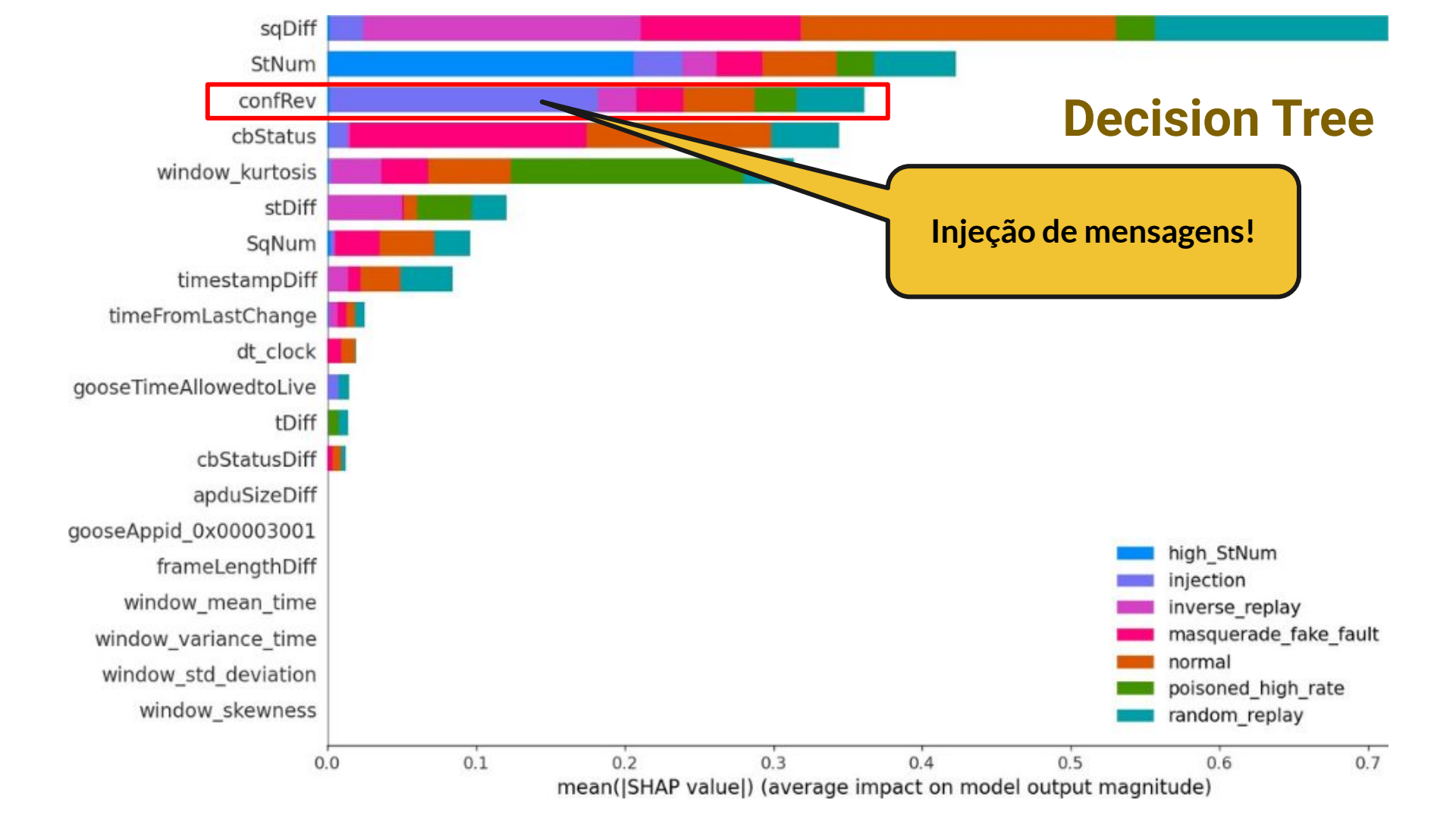

Reset SqNum & Increase StNum!
SqNum = 6
SqNum = 5
SqNum = 4
SqNum = 3
SqNum = 2
SqNum = 1
StNum = 2
StNum = 2
StNum = 2
StNum = 2
StNum = 2
StNum = 2

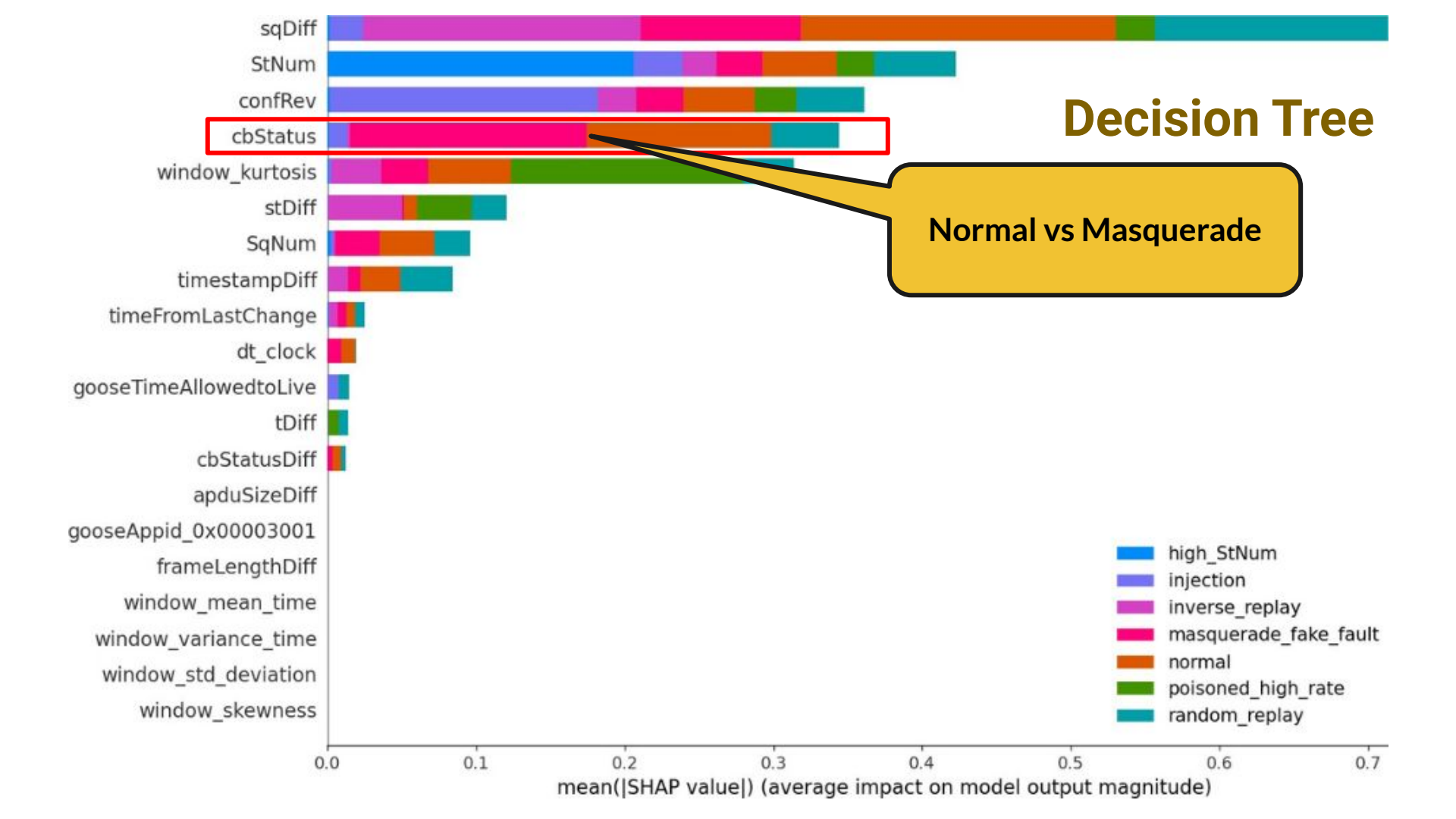
Resultados

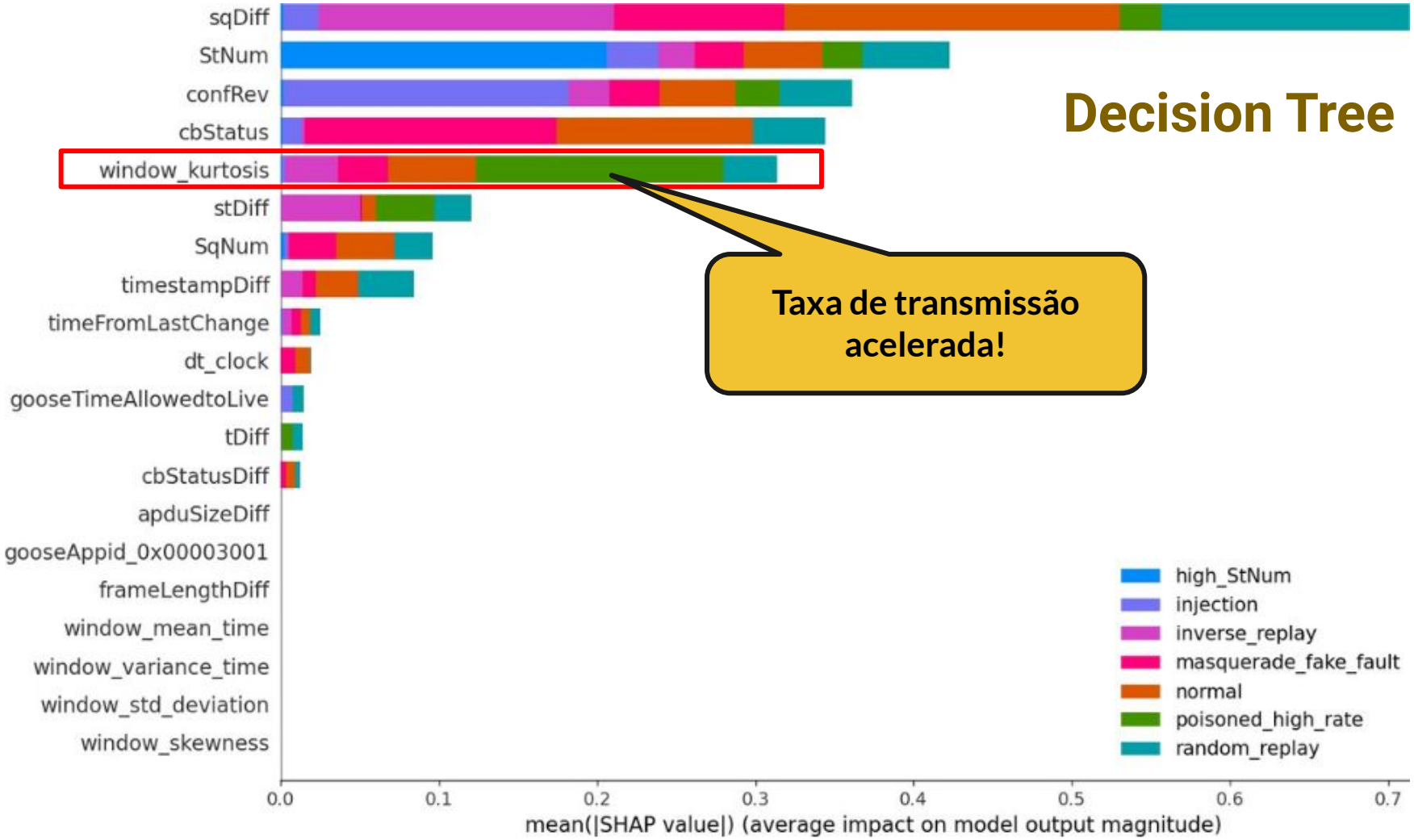

Preamble
Start of Frame
MAC Header
Priority Tagged
Type
Eth-type PDU
APDU
APPID
Length
Reserved 1
Reserved 2
ASDU N
GOOSE PDU
FCS
gocbRef
TTL
datSet
goID
timestamp
stNum
sqNum
test
confRev
ndsCom
numDatSet
allData
data
...
data

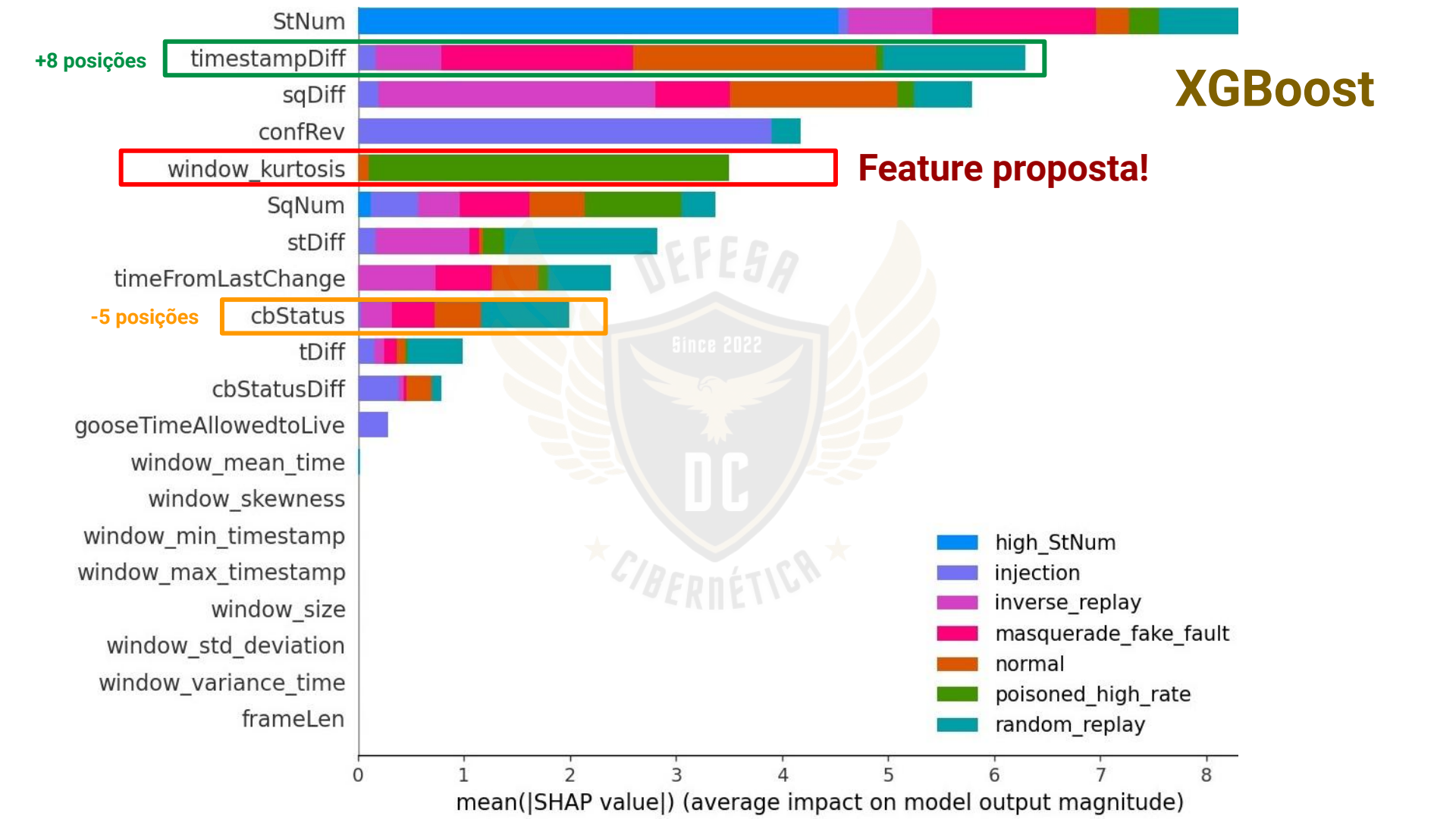

SqNum = 3
StNum = 1
Max Time
SqNum = 2
StNum = 1
Max Time
SqNum = 1
StNum = 1

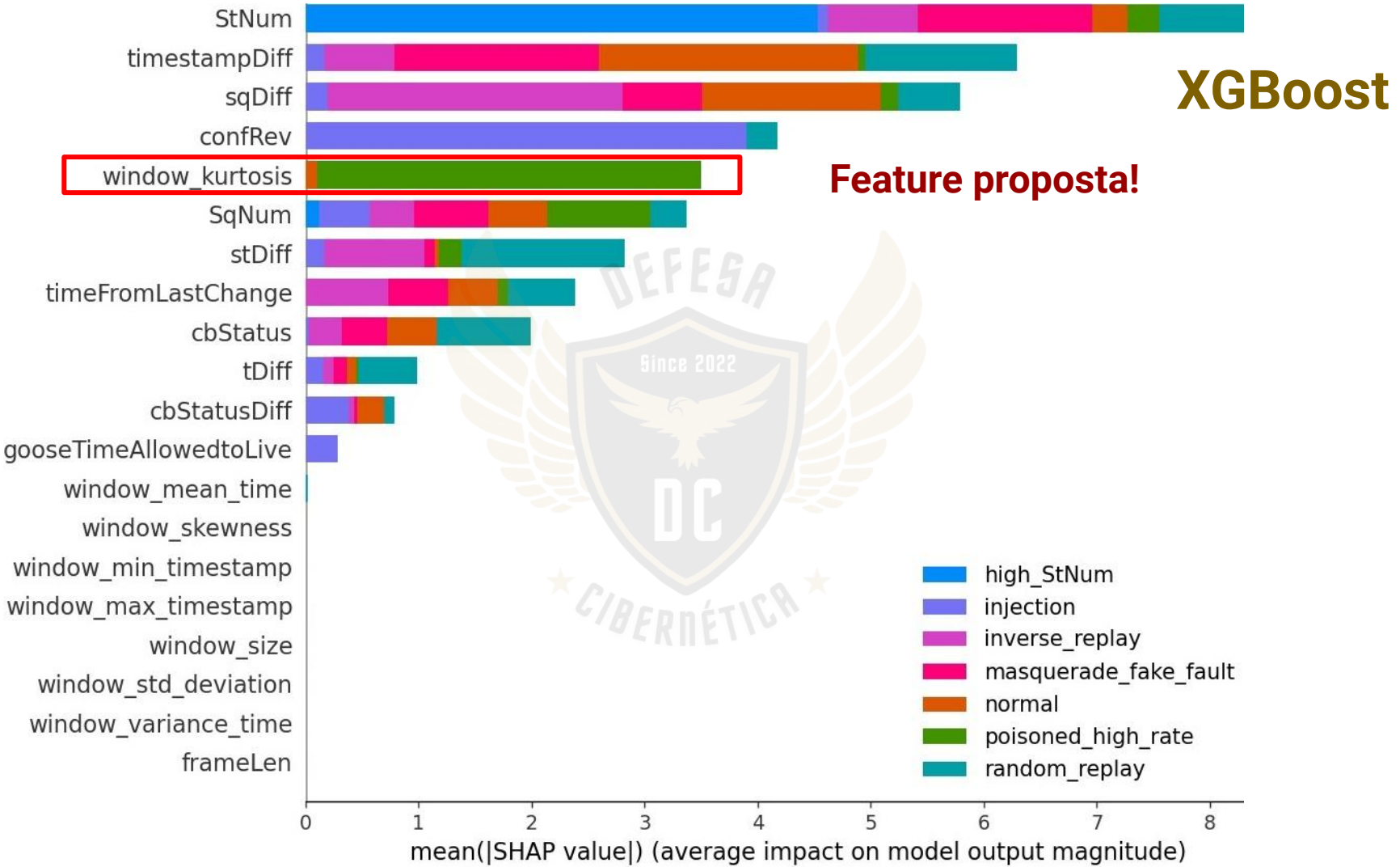

sqDiff
StNum
confRev
cbStatus
window_kurtosis
stDiff
SqNum
timestampDiff
timeFromLastChange
dt_clock
gooseTimeAllowedtoLive
tDiff
cbStatusDiff
apduSizeDiff
gooseAppid_0x00003001
frameLengthDiff
window_mean_time
window_variance_time
window_std_deviation
window_skewness
Decision Tree
Manipulação do número de status (StNum)
high_StNum
injection
inverse_replay
masquerade_fake_fault
normal
poisoned_high_rate
random_replay
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
mean(|SHAP value|) (average impact on model output magnitude)


sqDiff
StNum
confRev
cbStatus
window_kurtosis
stDiff
SqNum
timestampDiff
timeFromLastChange
dt_clock
gooseTimeAllowedtoLive
tDiff
cbStatusDiff
apduSizeDiff
gooseAppid_0x00003001
frameLengthDiff
window_mean_time
window_variance_time
window_std_deviation
window_skewness
Decision Tree
Injeção de mensagens!
high_StNum
injection
inverse_replay
masquerade_fake_fault
normal
poisoned_high_rate
random_replay
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
mean(|SHAP value|) (average impact on model output magnitude)


sqDiff
StNum
confRev
cbStatus
window_kurtosis
stDiff
SqNum
timestampDiff
timeFromLastChange
dt_clock
gooseTimeAllowedtoLive
tDiff
cbStatusDiff
apduSizeDiff
gooseAppid_0x00003001
frameLengthDiff
window_mean_time
window_variance_time
window_std_deviation
window_skewness
Decision Tree
Normal vs Masquerade
high_StNum
injection
inverse_replay
masquerade_fake_fault
normal
poisoned_high_rate
random_replay
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
mean(|SHAP value|) (average impact on model output magnitude)

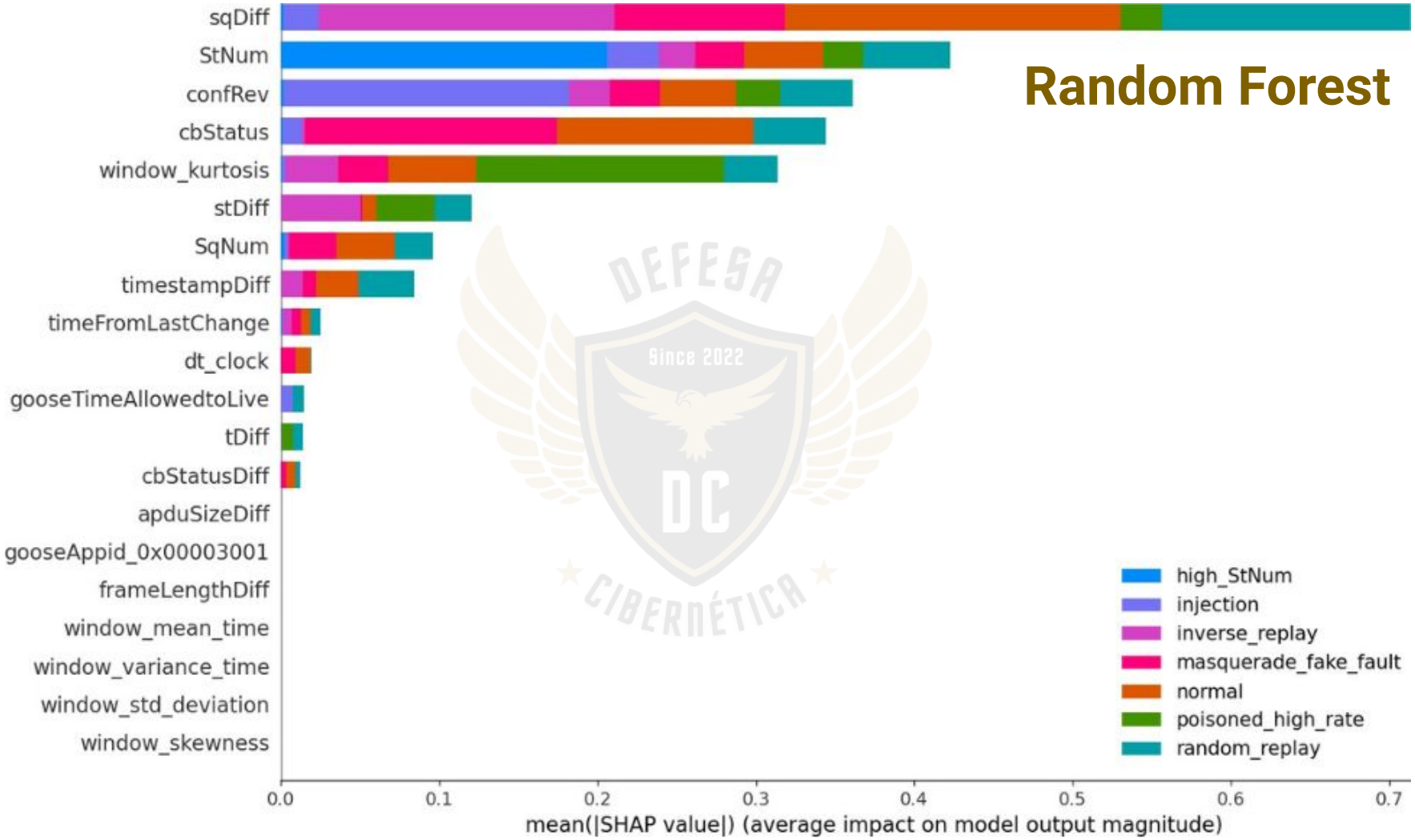

sqDiff
StNum
confRev
cbStatus
window_kurtosis
stDiff
SqNum
timestampDiff
timeFromLastChange
dt_clock
gooseTimeAllowedtoLive
tDiff
cbStatusDiff
apduSizeDiff
gooseAppid_0x00003001
frameLengthDiff
window_mean_time
window_variance_time
window_std_deviation
window_skewness
Decision Tree
Taxa de transmissão acelerada!
high_StNum
injection
inverse_replay
masquerade_fake_fault
normal
poisoned_high_rate
random_replay
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
mean(|SHAP value|) (average impact on model output magnitude)

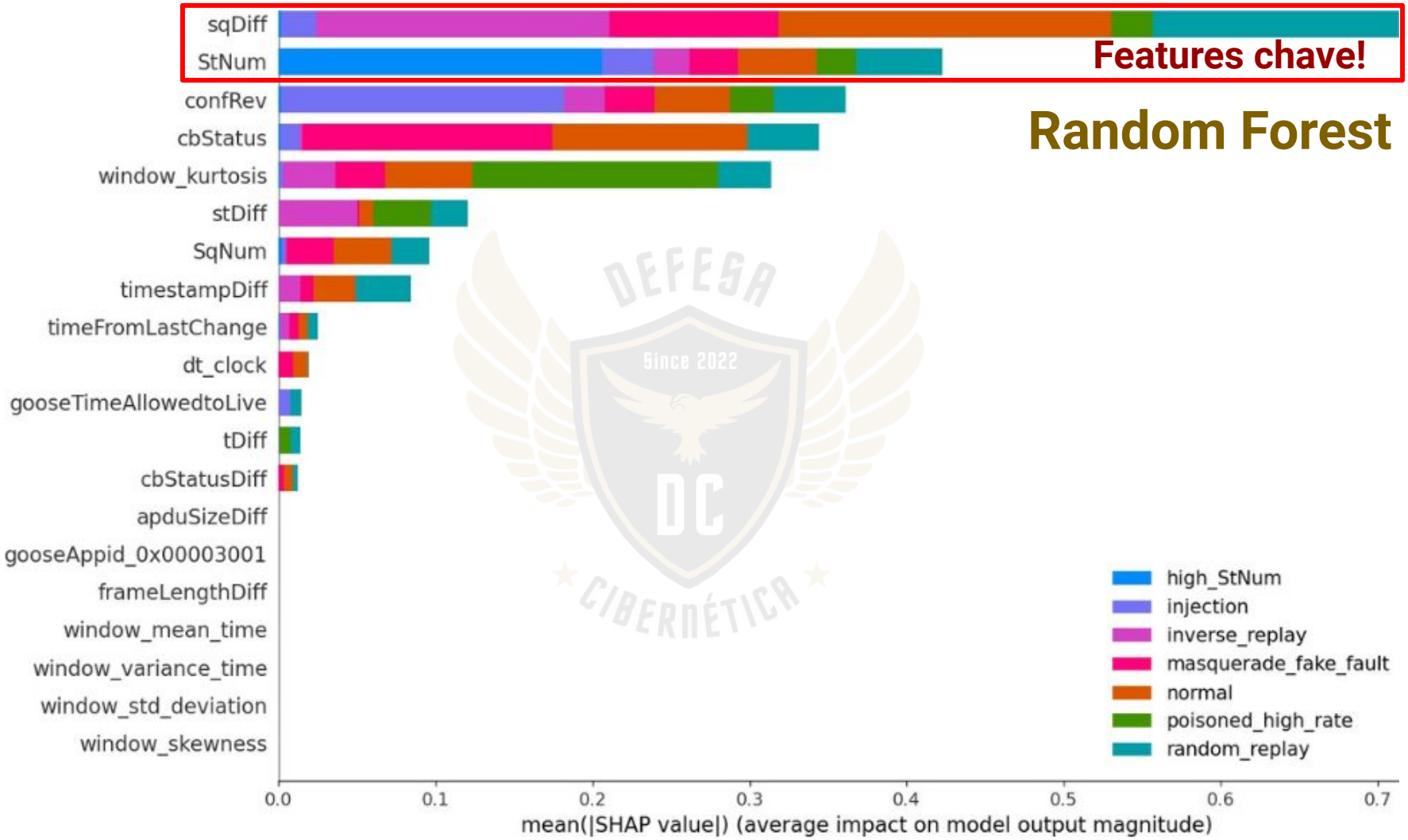

StNum
+8 posições
timestampDiff
sqDiff
confRev
window_kurtosis
SqNum
stDiff
timeFromLastChange
-5 posições
cbStatus
tDiff
cbStatusDiff
gooseTimeAllowedtoLive
window_mean_time
window_skewness
window_min_timestamp
window_max_timestamp
window_size
window_std_deviation
window_variance_time
frameLen
XGBoost
Feature proposta!
high_StNum
injection
inverse_replay
masquerade_fake_fault
normal
poisoned_high_rate
random_replay
0
1
2
3
4
5
6
7
8
mean(|SHAP value|) (average impact on model output magnitude)

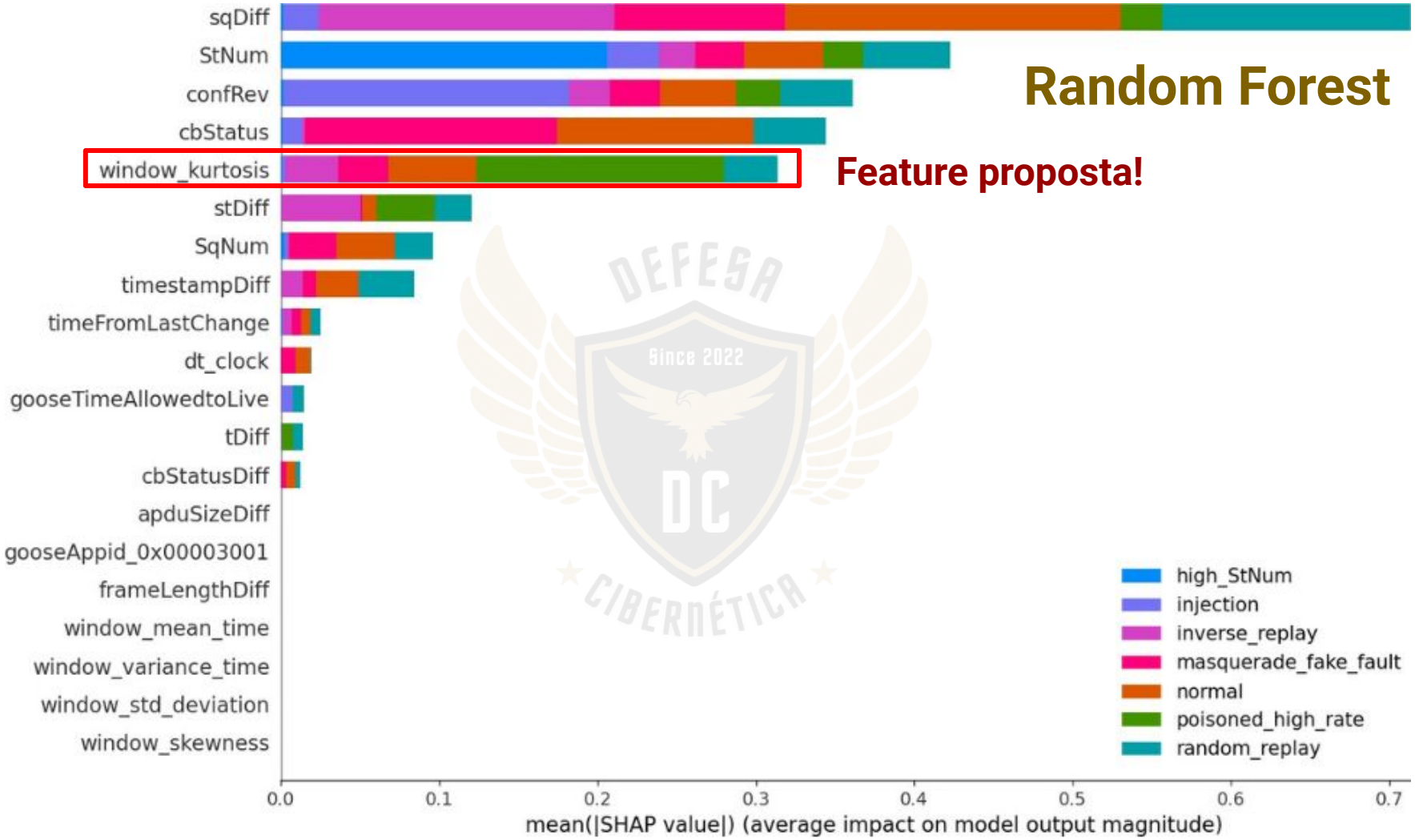


XGBoost

Feature proposta!

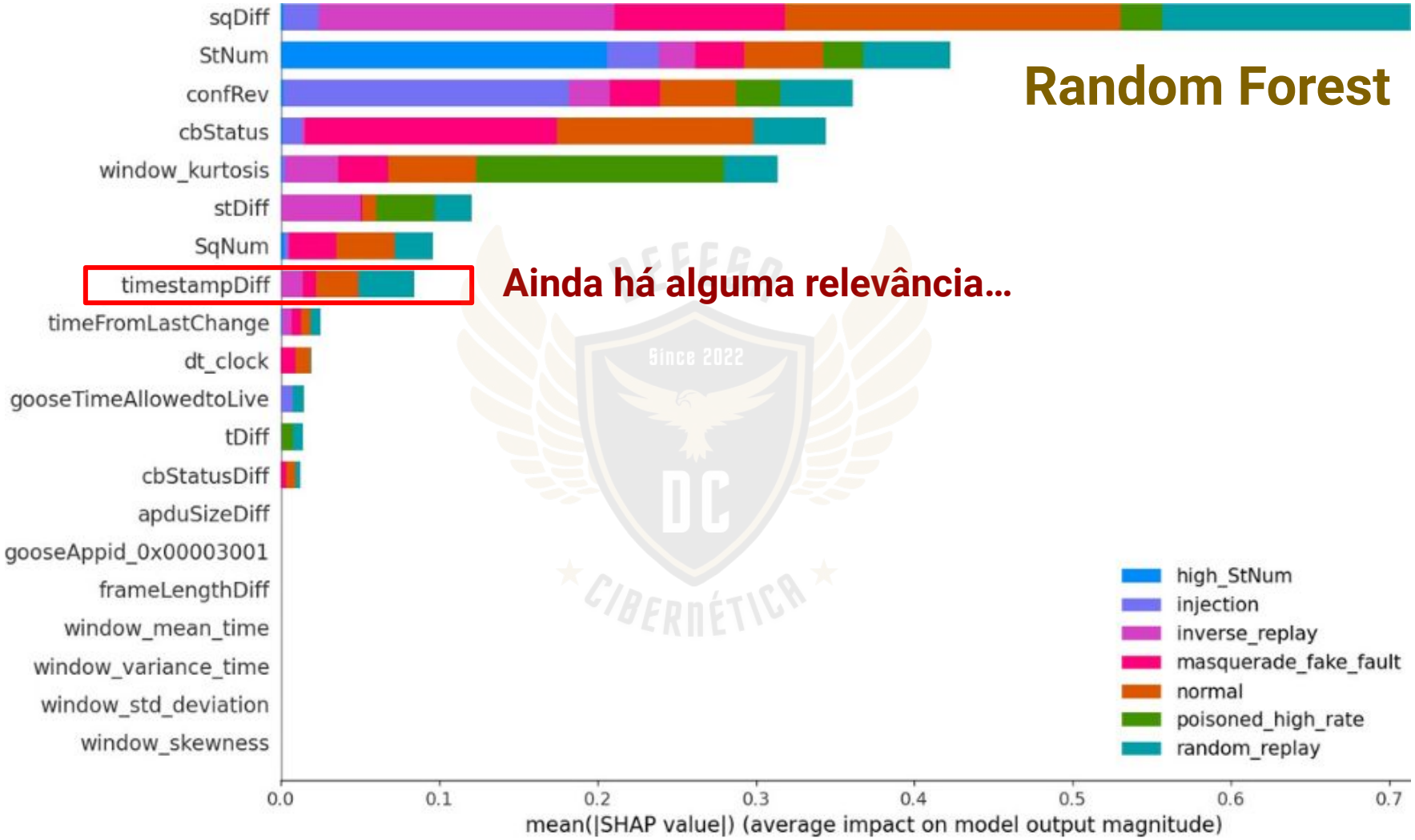

Random Forest
sqDiff
StNum
confRev
cbStatus
window_kurtosis
stDiff
SqNum
timestampDiff
timeFromLastChange
dt_clock
gooseTimeAllowedtoLive
tDiff
cbStatusDiff
apduSizeDiff
gooseAppid_0x00003001
frameLengthDiff
window_mean_time
window_variance_time
window_std_deviation
window_skewness
high_StNum
injection
inverse_replay
masquerade_fake_fault
normal
poisoned_high_rate
random_replay
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
mean(|SHAP value|) (average impact on model output magnitude)


sqDiff
StNum
confRev
cbStatus
window_kurtosis
stDiff
SqNum
timestampDiff
timeFromLastChange
dt_clock
gooseTimeAllowedtoLive
tDiff
cbStatusDiff
apduSizeDiff
gooseAppid_0x00003001
frameLengthDiff
window_mean_time
window_variance_time
window_std_deviation
window_skewness
Features chave!
Random Forest
high_StNum
injection
inverse_replay
masquerade_fake_fault
normal
poisoned_high_rate
random_replay
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
mean(|SHAP value|) (average impact on model output magnitude)



Random Forest

Feature proposta!



Random Forest

Ainda há alguma relevância...

Considerações Finais

- **Explicabilidade**
  - O valor SHAP está diretamente associado ao modo de operação!
- **Nuances**
  - Algoritmos baseados em árvores (**Random Forest** e **Decision Tree**) priorizaram as features fundamentais (sqDiff e StNum)
  - **XGBoost** “valorizou mais” o TimestampDiff!
- **Novas features**
  - Uma delas teve um impacto significativo (window_kurtosis)

Código-Fonte no GitHub

https://qr-codes.io/JFbZNe

SBSeg24 SÃO JOSÉ DOS CAMPOS | unipampa | Universidade Federal Fluminense | UFU

Obrigado!

Perguntas?

silvioquincozes@unipampa.edu.br

@defesacibernetica.dc

Referências

- https://tiinside.com.br/18/07/2024/quaduplicam-custos-para-empresas-de-infraestrutura-se-recuperarem-de-ciberataques/
- https://oglobo.globo.com/economia/companhia-eletrica-em-porto-rico-sofre-ataque-cibernetico-incendio-milhares-ficam-sem-luz-25056417

Obrigado!

silvioquincozes@unipampa.edu.br