Comprehensive Ransomware Detection: Optimization of Feature Selection through Machine Learning Algorithms and Explainable AI on Memory Analysis

Lucas Leonel, Diego Nunes Molinos,
Rodrigo Sanches Miani

NUSEC
NÚCLEO DE SEGURANÇA DA INFORMAÇÃO FACOM/UFU

UFU Universidade Federal de Uberlândia

Faculdade de Computação (FACOM)
Universidade Federal de Uberlândia (UFU)

Introdução

Ransomware is a type of malware that threatens to destroy or withhold a victim's critical data unless a ransom is paid to the attacker. Unfortunately, cyberattacks are on the rise as we see 71% year-over-year increase in cyberattacks.

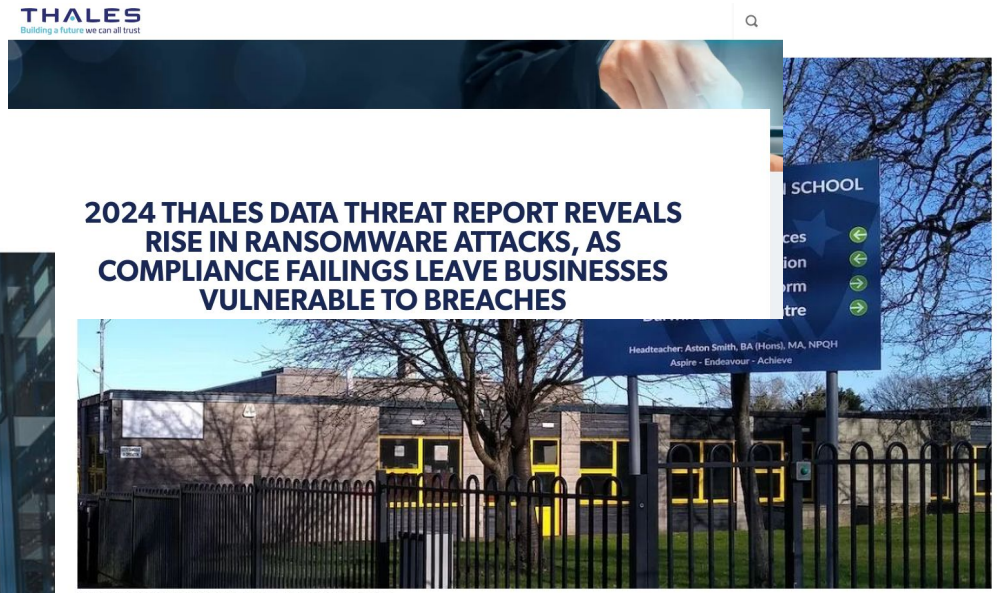


IMAGE: MX. GRANGER / WIKIMEDIA COMMONS / CC0 1.0

Alexander Martin
September 9th, 2024

Cybercrime News

Ransomware attack forces high school in London to close and send students home

Motivação

Muitos ransomwares utilizam técnicas de ofuscação e polimorfismo para evitar detecção baseada em assinaturas estáticas;

Através da análise de dados da memória é possível prever o comportamento do ransomware durante sua execução, e.g. DLLs, *Threads*, Arquivos, Mudanças no Registro etc.

Com o advento dos algoritmos de aprendizado de máquina (*Machine Learning*), a detecção de ransomwares tem evoluído consideravelmente.

Fato é que, uma compreensiva análise do conjunto de *features* juntamente com o apoio de ferramentas de IA explicável contribui significantemente para a eficiência dos modelos gerados.

Features no Contexto de Detecção de Ransomware com ML e XAI

Desempenho e Eficiência Computacional;

Risco de Overfitting;

Redução de Falsos Positivos;

Tempo de Treinamento e Atualização dos Modelos;

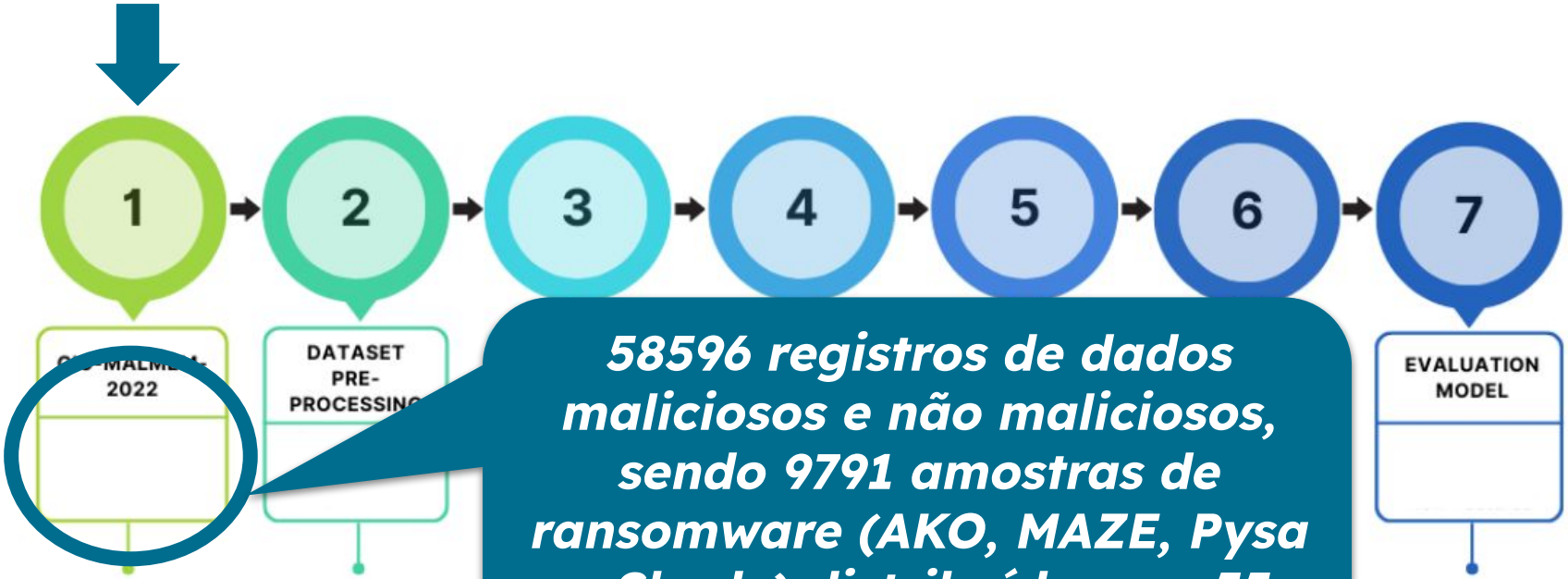
Desafio(s)

Destacar as características que melhor descrevem comportamentos maliciosos no contexto de Ransomware utilizando informações extraídas da memória.

Reduzir o número de características necessárias para detecção de ransomwares mantendo a confiabilidade do modelo;

Racionalizar o processo de decisão do modelo gerado (DT) para aumentar o nível de confiabilidade do processo de detecção;

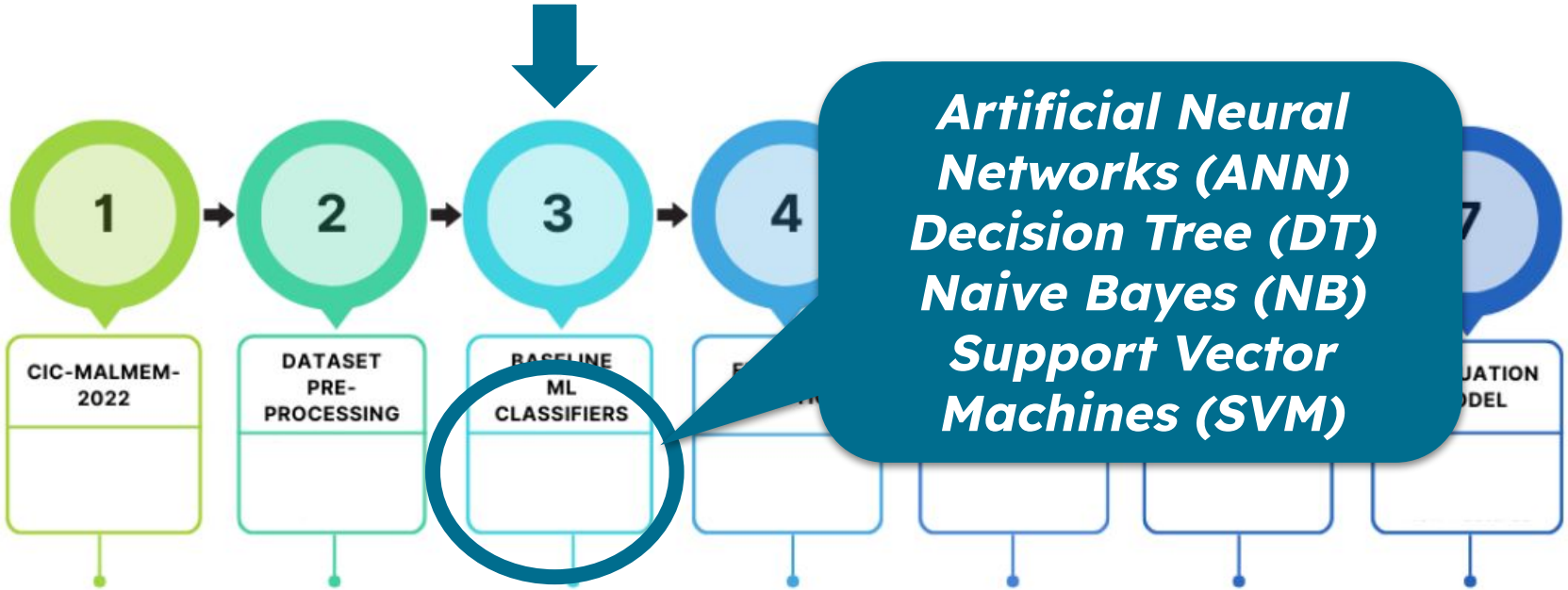
Método


1
2
3
4
5
6
7
2022
DATASET
PRE-
EVALUATION
MODEL
58596 registros de dados maliciosos e não maliciosos, sendo 9791 amostras de ransomware (AKO, MAZE, Pysa e Shade) distribuídas em 55 distintas features

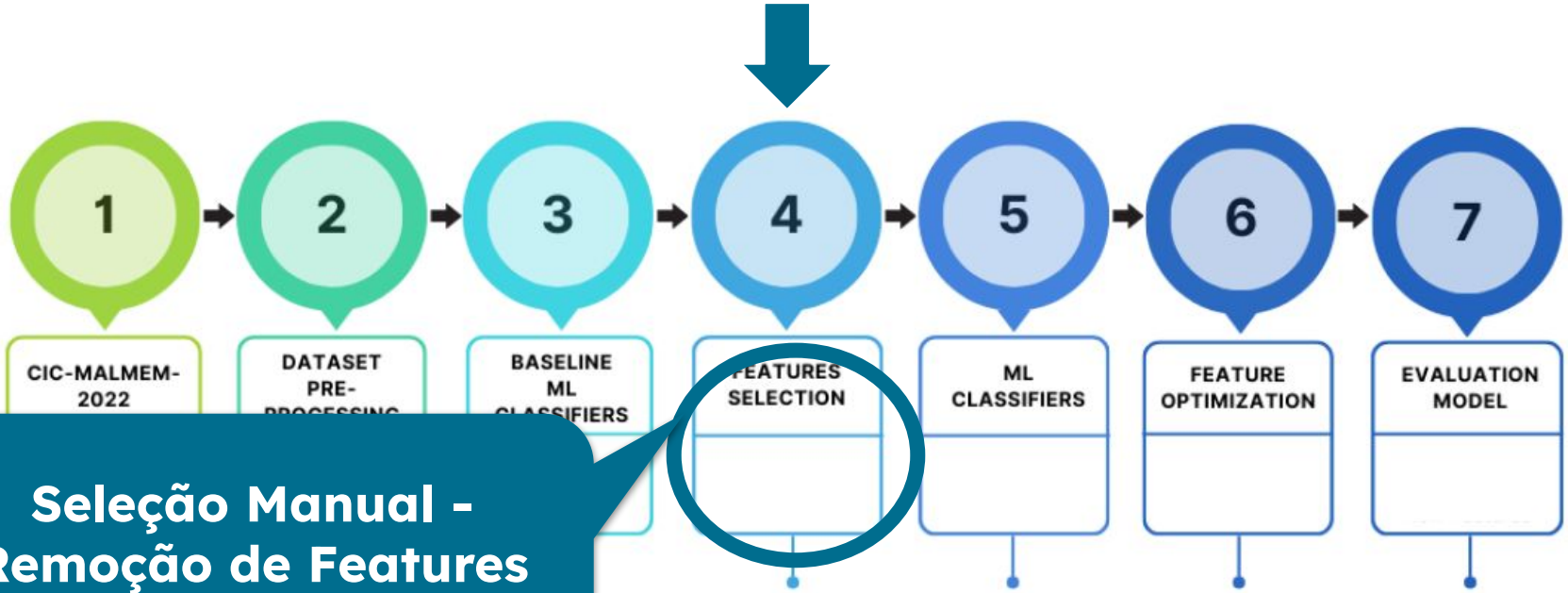
Método


1
2
3
7
CIC-MALMEM-2022
PRE-PROCESSING
EVALUATION MODEL
Estratificação de amostras maliciosas (Ransomware) e não maliciosas

Método

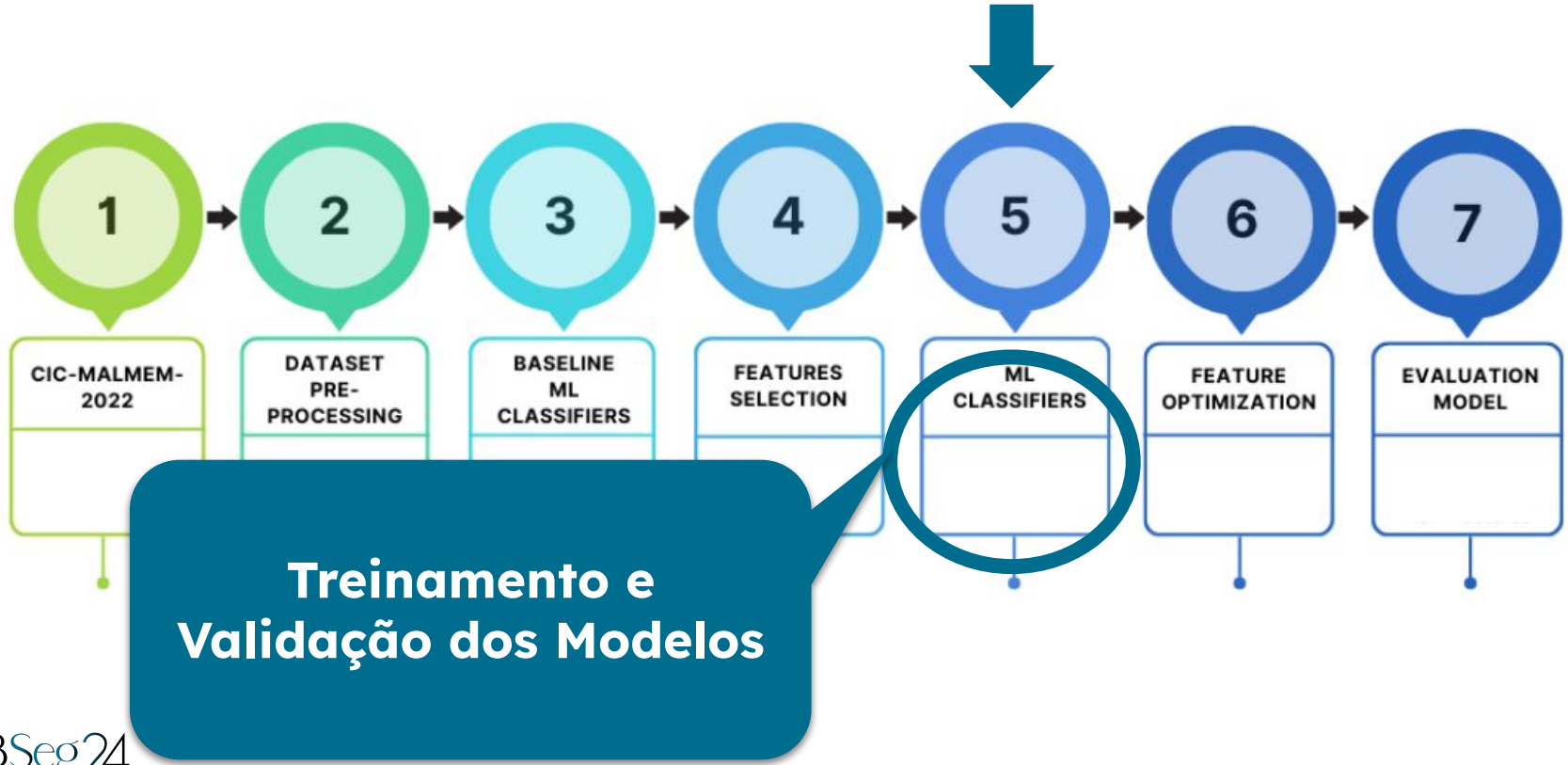

1
2
3
4
CIC-MALMEM-2022
DATASET PRE-PROCESSING
BASELINE ML CLASSIFIERS
Artificial Neural Networks (ANN)
Decision Tree (DT)
Naive Bayes (NB)
Support Vector Machines (SVM)

Método

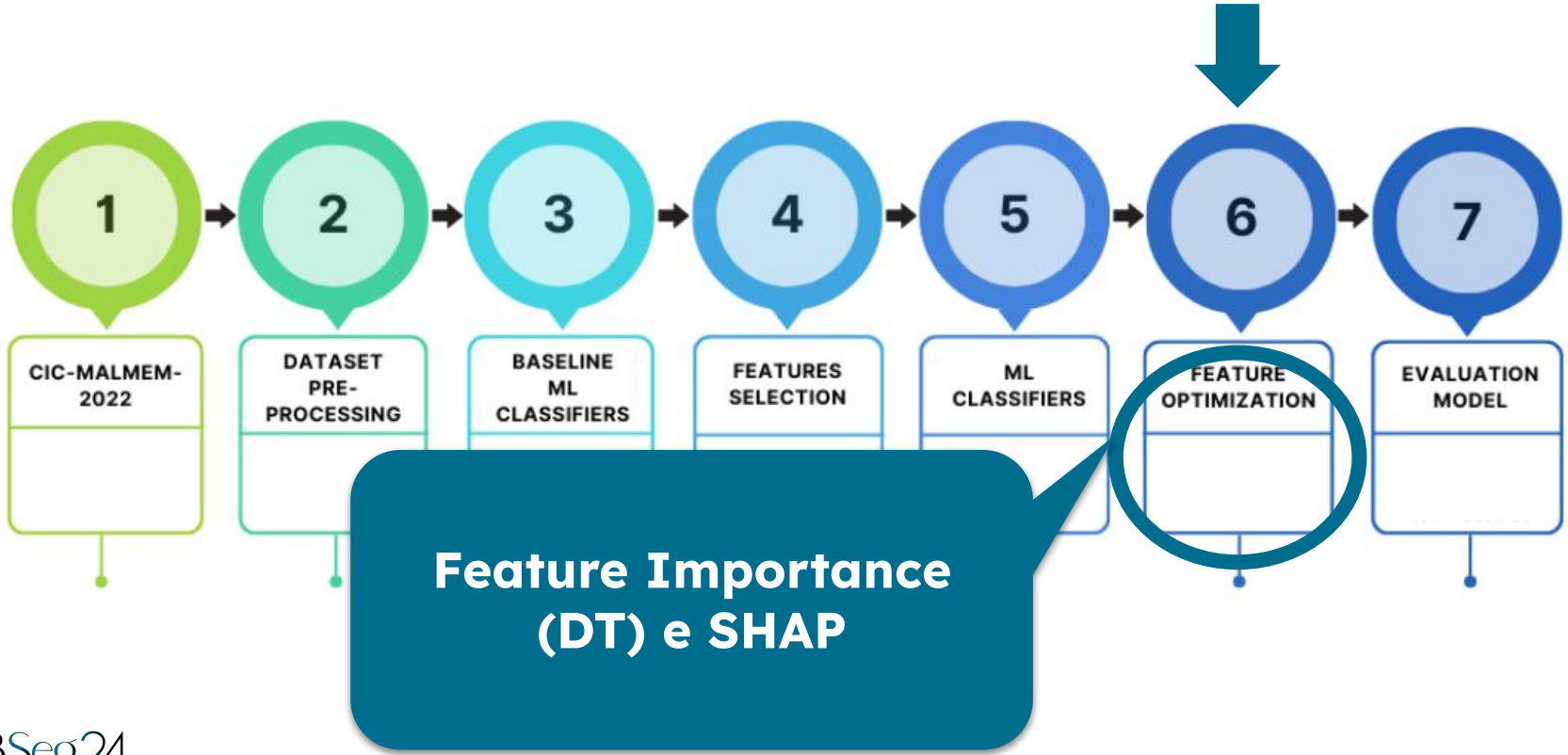


Seleção Manual - Remoção de Features menos significantes

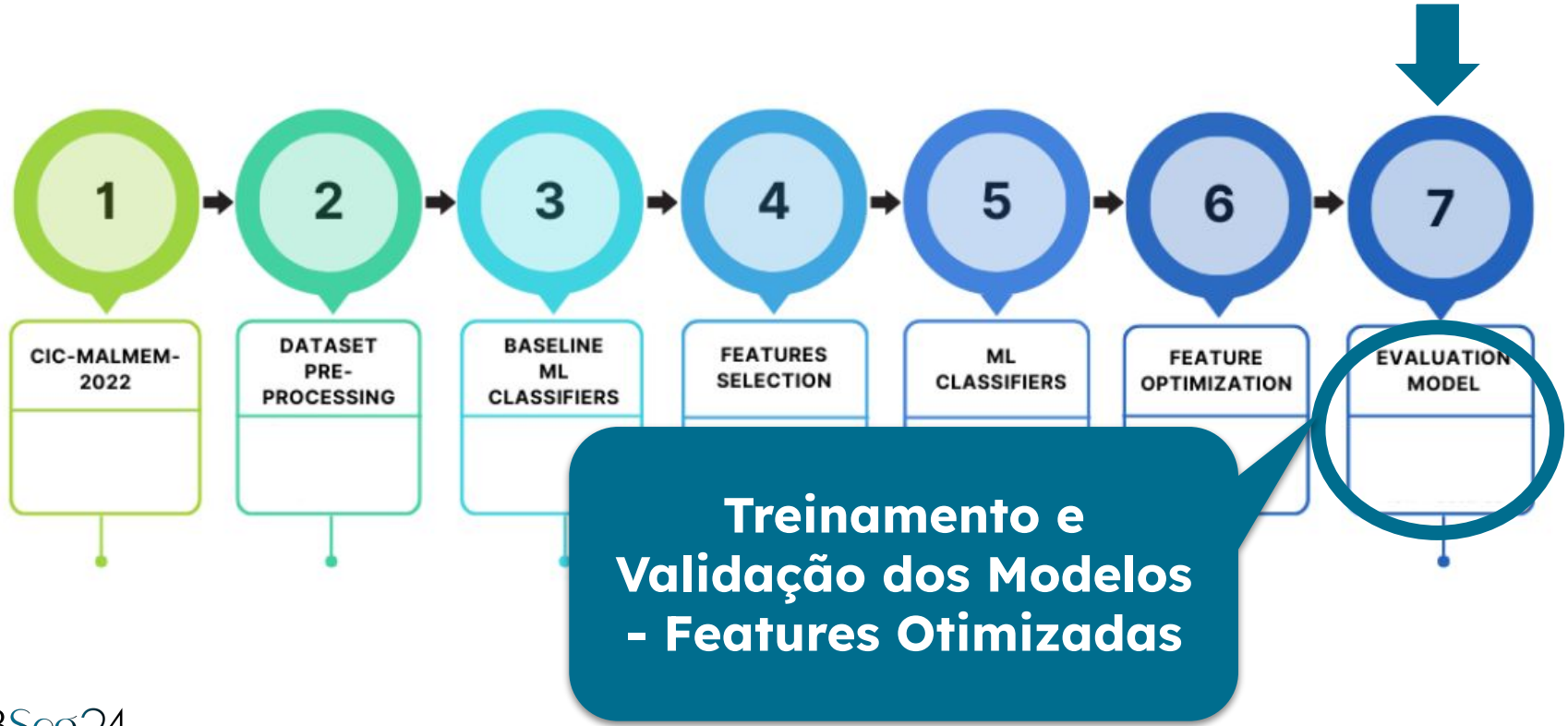
Método


1
2
3
4
5
6
7
CIC-MALMEM-2022
DATASET PRE-PROCESSING
BASELINE ML CLASSIFIERS
FEATURES SELECTION
ML CLASSIFIERS
FEATURE OPTIMIZATION
EVALUATION MODEL
Treinamento e Validação dos Modelos

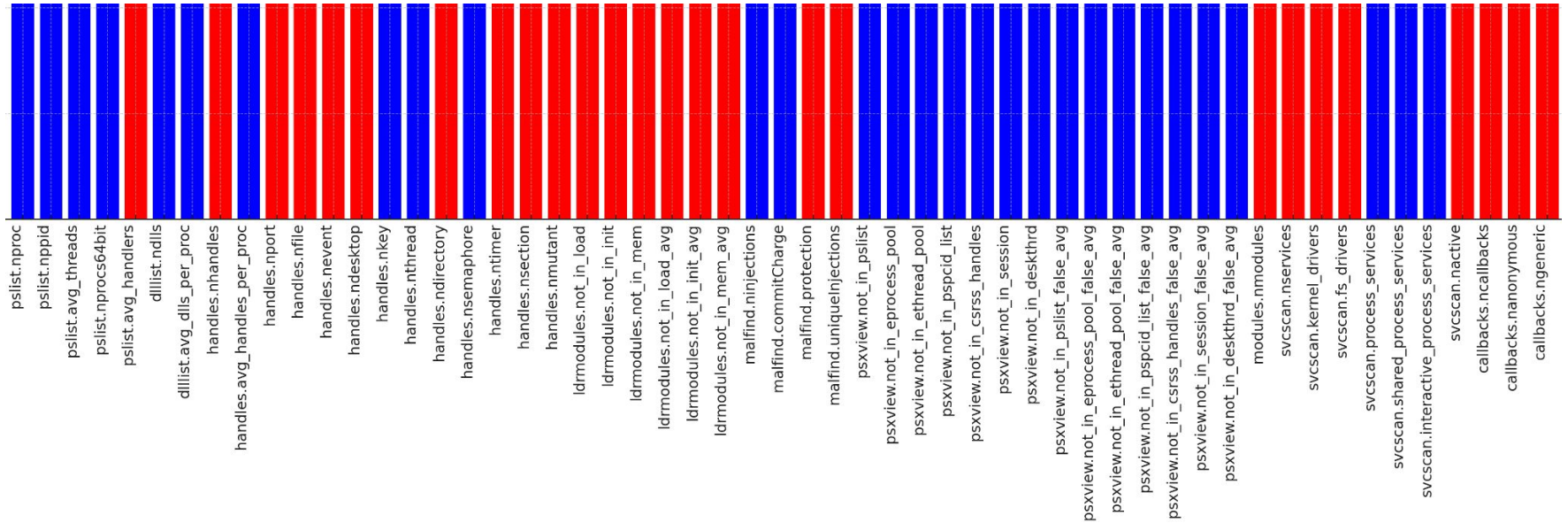
Método


1
2
3
4
5
6
7
CIC-MALMEM-2022
DATASET PRE-PROCESSING
BASELINE ML CLASSIFIERS
FEATURES SELECTION
ML CLASSIFIERS
FEATURE OPTIMIZATION
EVALUATION MODEL
Feature Importance (DT) e SHAP

Método

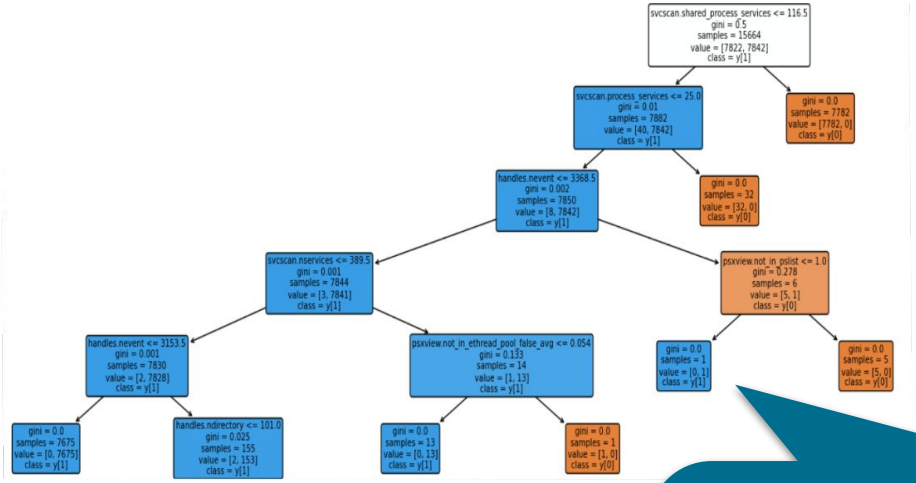

1
CIC-MALMEM-2022
2
DATASET PRE-PROCESSING
3
BASELINE ML CLASSIFIERS
4
FEATURES SELECTION
5
ML CLASSIFIERS
6
FEATURE OPTIMIZATION
7
EVALUATION MODEL
Treinamento e Validação dos Modelos - Features Otimizadas

Avaliação

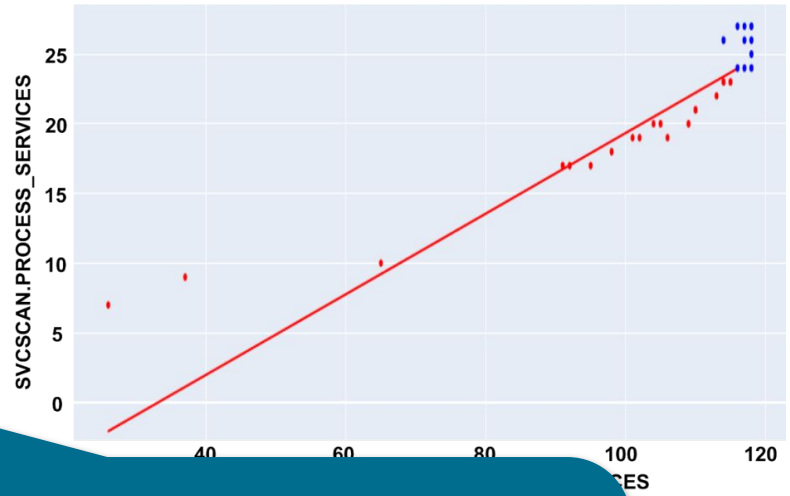


a) Terminated Processes, b) DLL's Records, c) Registry Modifications, d) Active Network Connections e, e) Running Services

Avaliação



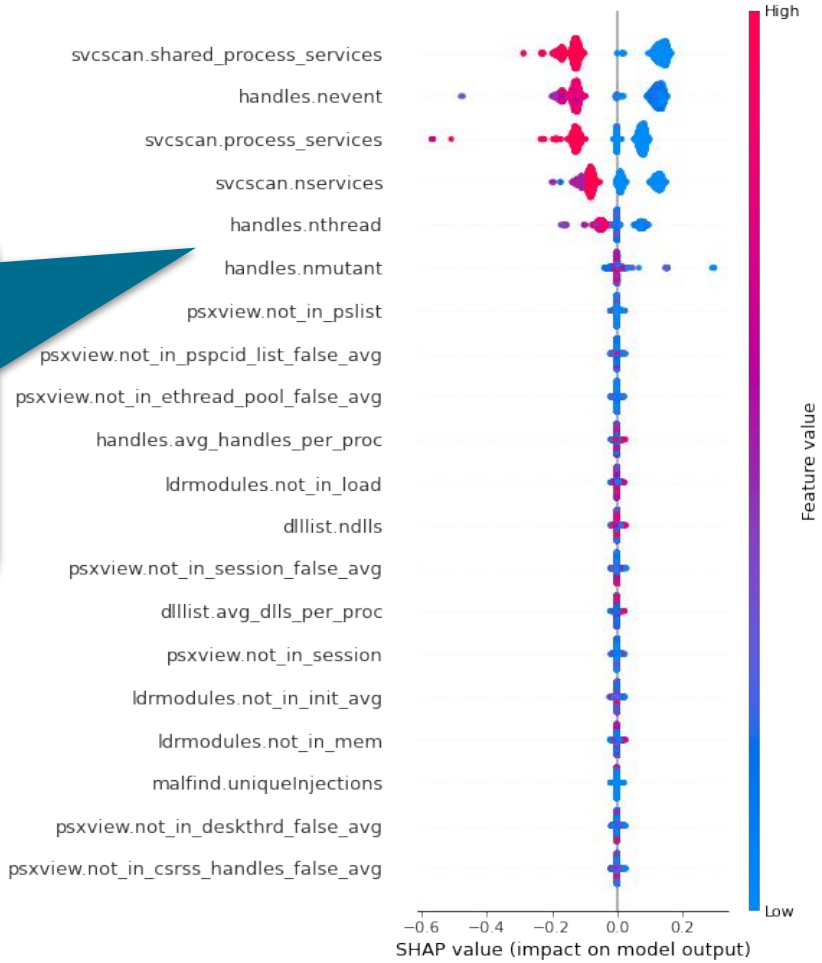
(a)



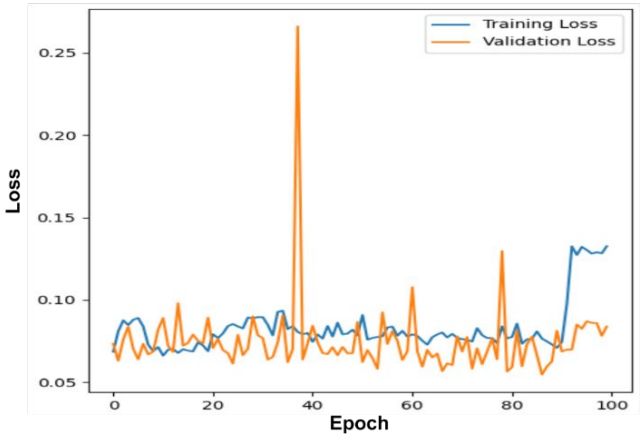
(a) svscan˙shared process services,
(b) svscan.process services,
(c) handles.nevent,
(d) svscan.nservices.

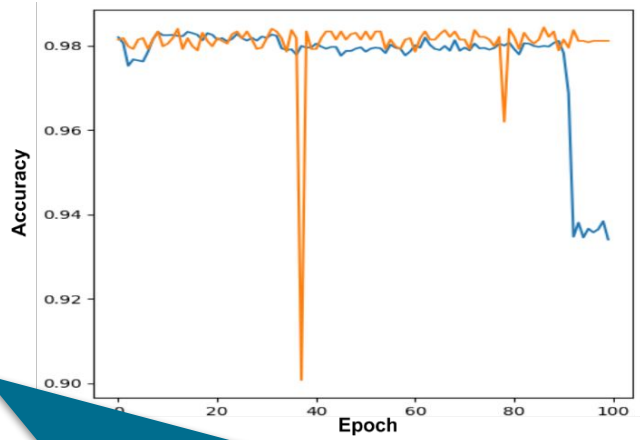
Avaliação

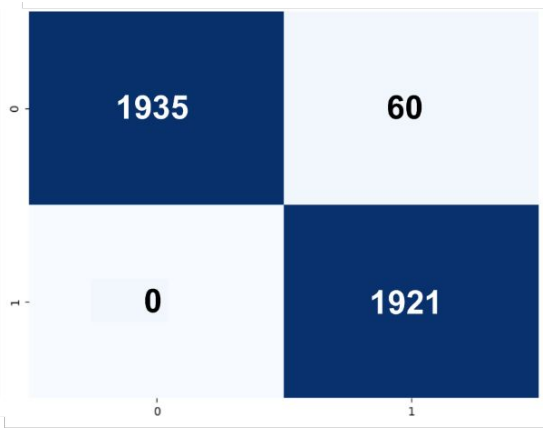
(a) svs-can.shared_process_services,
(b) handles.nevent,
(c) svscan.process_services,
(d) svscan.nservices,
(e) handles.nthread



Resultados

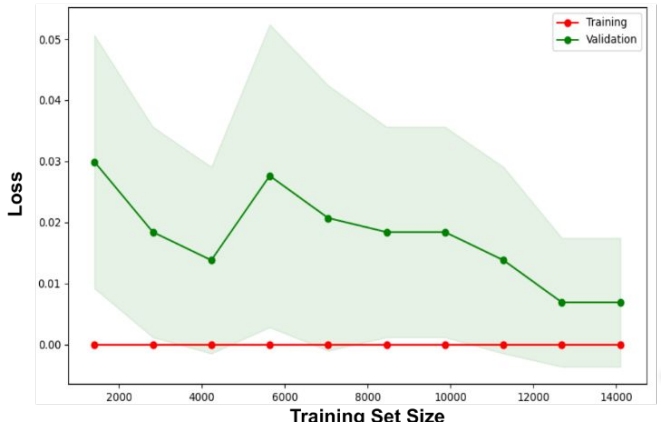


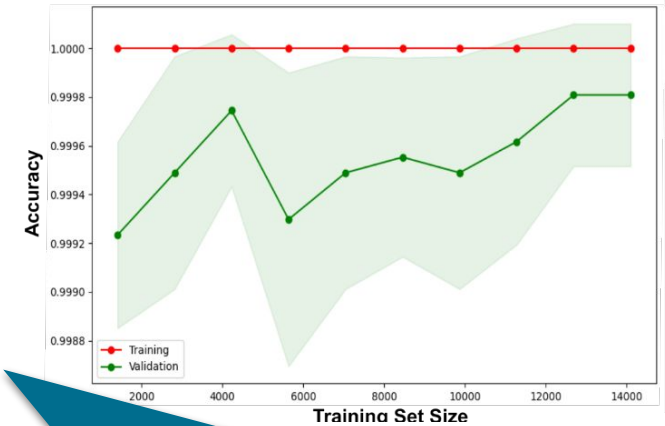


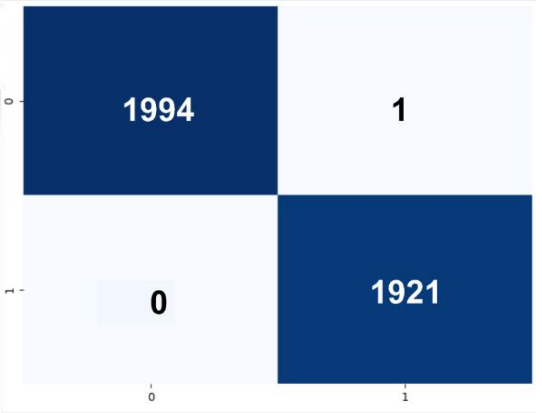


ANN exibe picos de perda ocasionais durante a fase de validação. No entanto, o modelo demonstra aprendizagem e convergência à medida que o número de *Epochs* aumenta.

Resultados

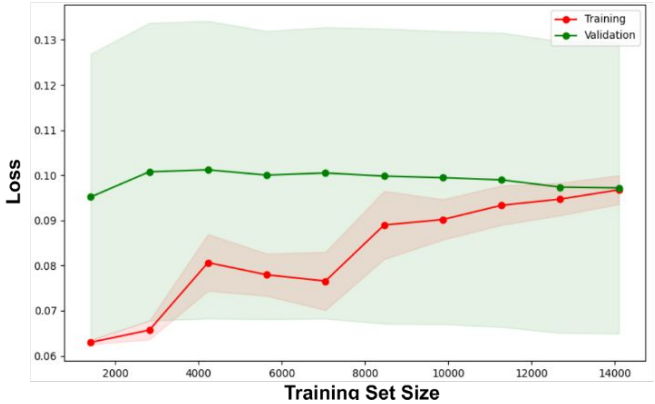


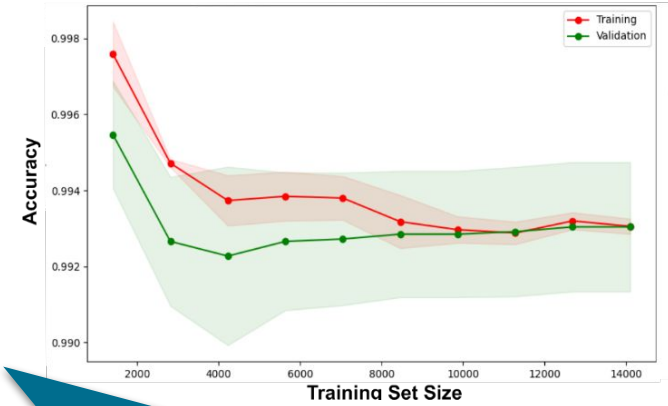


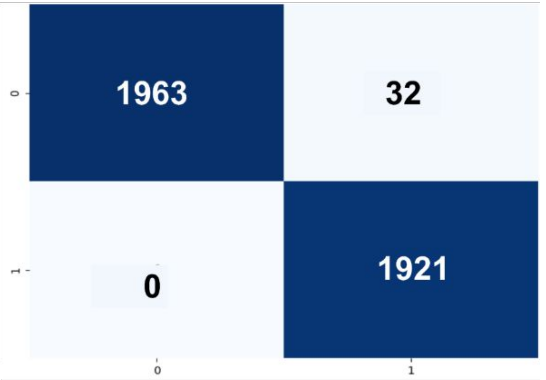


DT apresenta estabilidade tanto no treinamento quanto na validação, mesmo com o aumento de amostras, indicando desempenho confiável do modelo.

Resultados

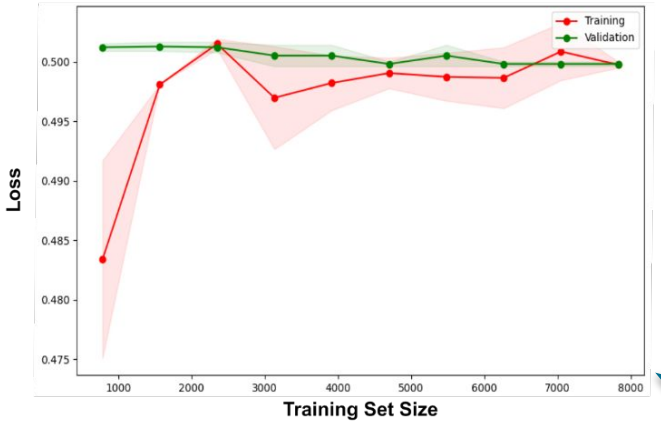


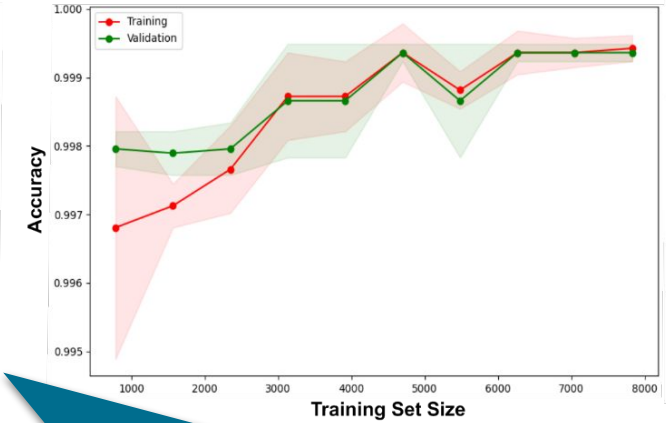


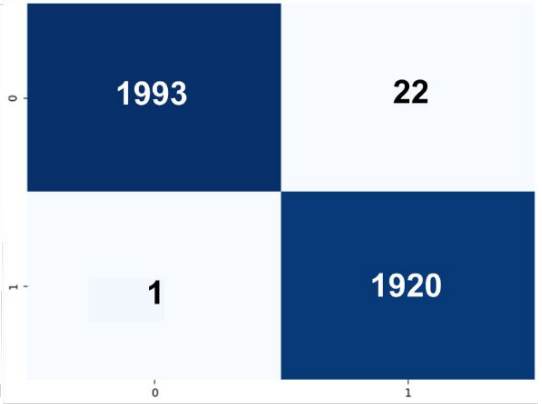


Assim como o DT, o NB apresenta estabilidade mesmo com o aumento nas amostras.

Resultados







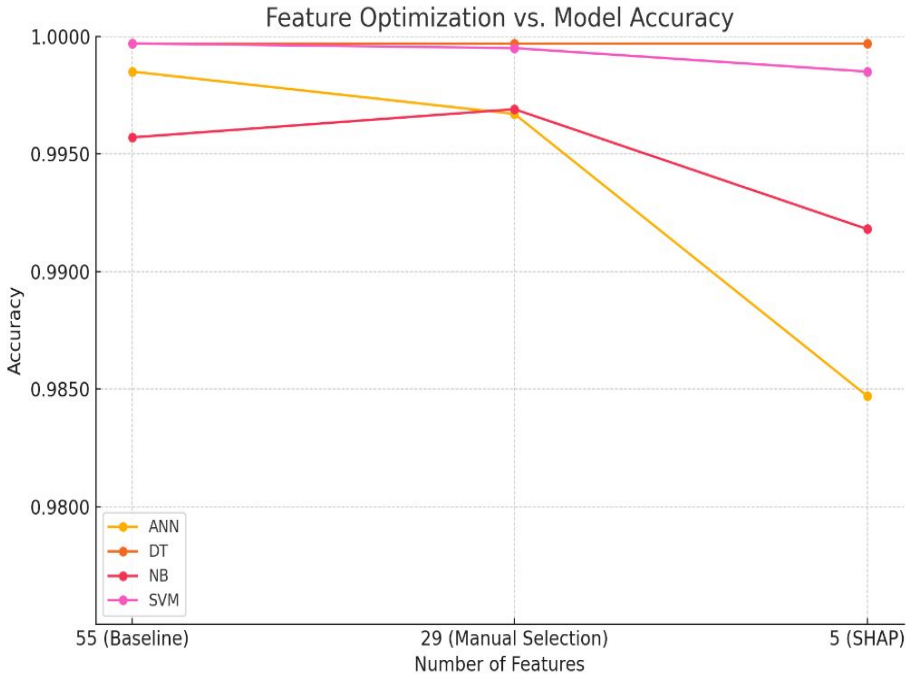
SVM inicialmente aparenta instabilidade, contudo, eventualmente alcança uma convergência à medida que o volume de amostras aumenta.

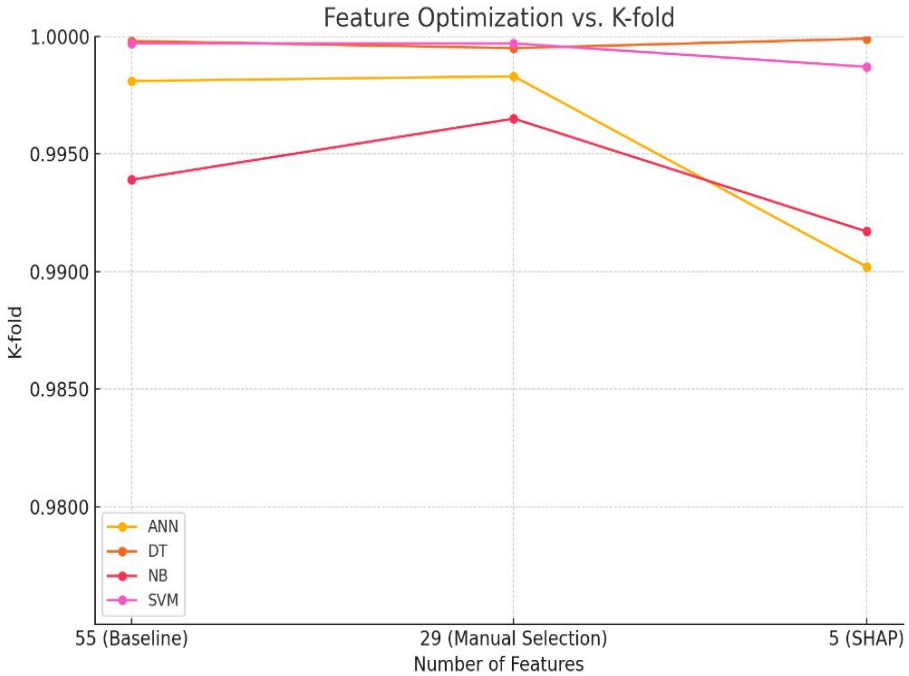
Resultados

Em todos os casos, os modelos parecem ter uma alta taxa de acerto com acurácia próxima de 1.0, tanto no conjunto de treinamento quanto de validação.

A matriz de confusão indica uma excelente separação entre as classes, com erros mínimos, o que é promissor para uma eventual aplicação prática do modelo.

Resultados


Feature Optimization vs. Model Accuracy
Accuracy
1.0000
0.9950
0.9900
0.9850
0.9800
ANN
DT
NB
SVM
55 (Baseline)
29 (Manual Selection)
5 (SHAP)
Number of Features


Feature Optimization vs. K-fold
K-fold
1.0000
0.9950
0.9900
0.9850
0.9800
ANN
DT
NB
SVM
55 (Baseline)
29 (Manual Selection)
5 (SHAP)
Number of Features

Considerações finais

- Integração de ML e XAI melhorou a detecção de ransomware com maior transparência e otimização, redução de 55 para 5 features.
- A integração de técnicas de XAI foi um ponto fundamental, permitindo a transparência no processo de decisão do modelo gerado pelo DT.

Trabalhos futuros

- Ampliar o conjunto de features, e.g. quantidade de *swap* utilizada e chamadas de alocação dinâmica de memória;
- Aplicar o método utilizado para outros tipos de malwares;
- Adicionar novos classificadores, e.g. algoritmos de boosting, tais como: *AdaBoost* e *Gradient Boosting Machines*.

Obrigado!

Lucas Leonel
lucas.leonelcosta@gmail.com

Diego Nunes Molinos
diego.molinos@ufu.br

Rodrigo Sanches Miani
miani@ufu.br

Patrocinadores do SBSeg 2024!

nic.br

cgi.br

Google

Tempest

CAPES

SiDi

FAPESP

zscaler™

BugHunt

CNPq

C.E.S.A.R

FACULDADE IBPTECH