Detecção de Malwares Android: datasets e reprodutibilidade

6º Workshop Regional de Segurança da Informação e de Sistemas Computacionais

Tainá Soares, Guilherme Siqueira, Lucas Barcellos,
Renato Sayyed, Luciano Vargas,
Gustavo Rodrigues, Joner Assolin, Jonas Pontes,
Diego Kreutz e Eduardo Feitosa

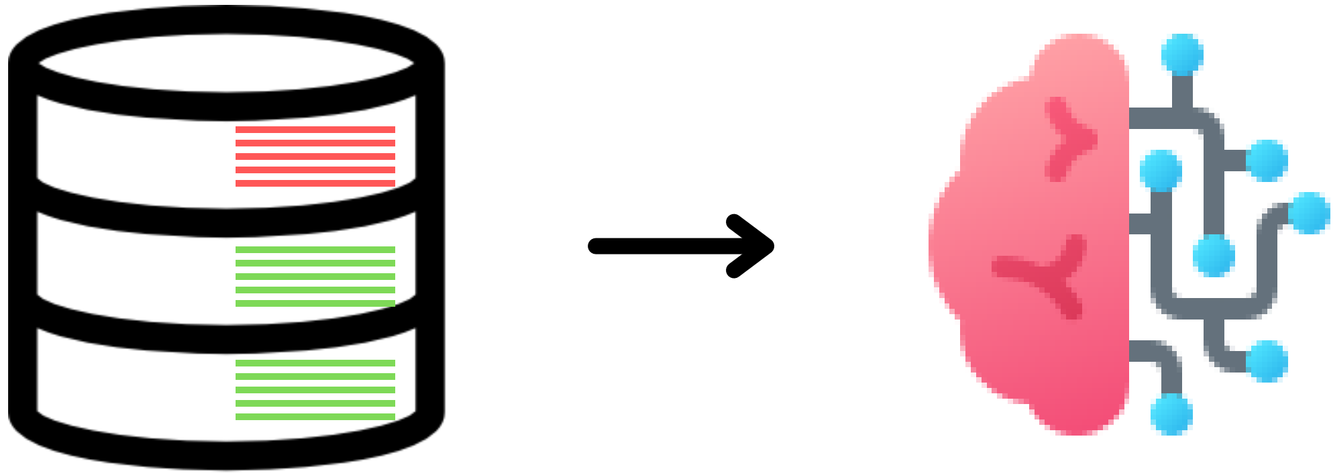
UFAM

A importância do dataset

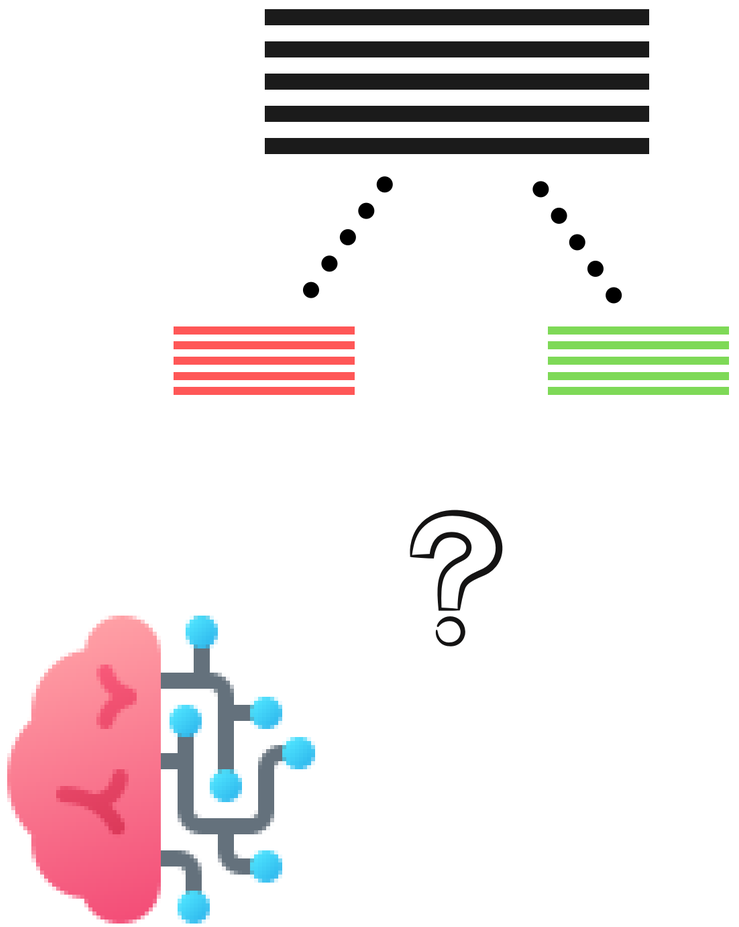
← aplicações maliciosas

← aplicações benignas

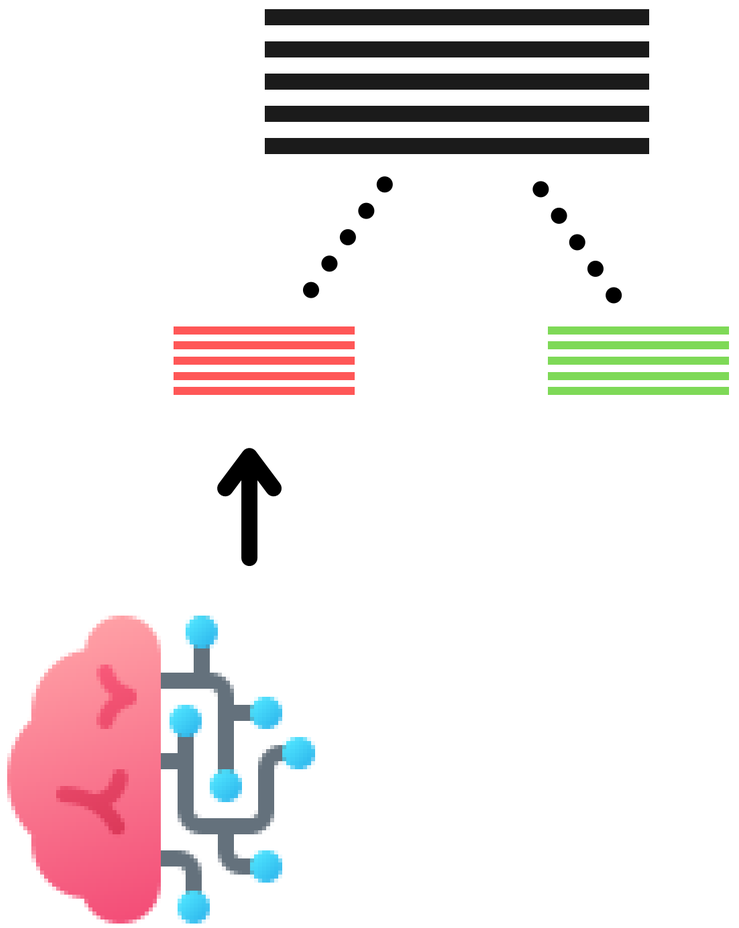
A importância do dataset



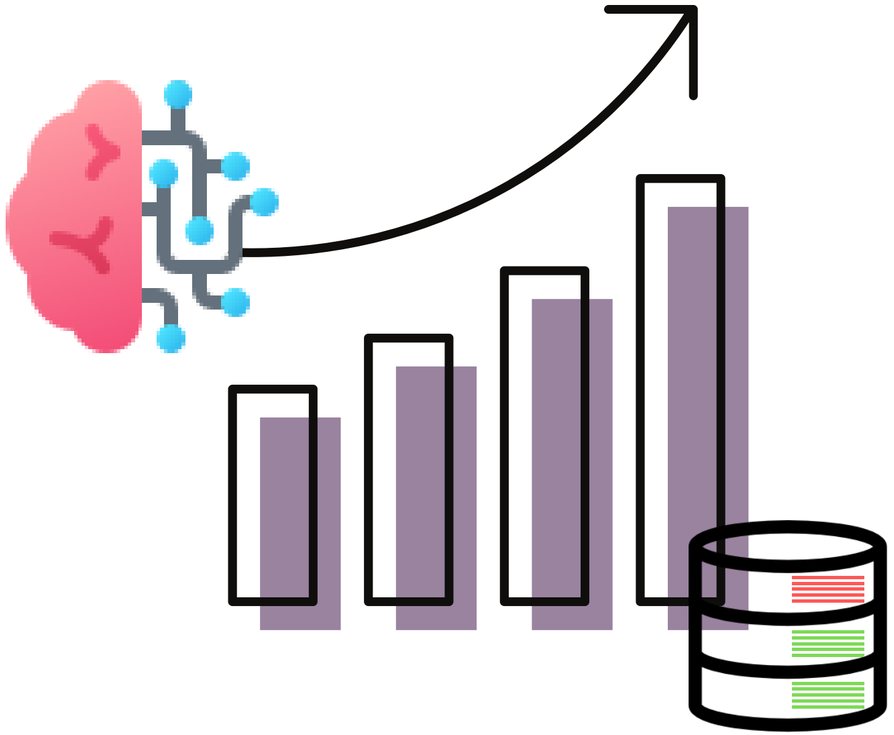
A importância do dataset



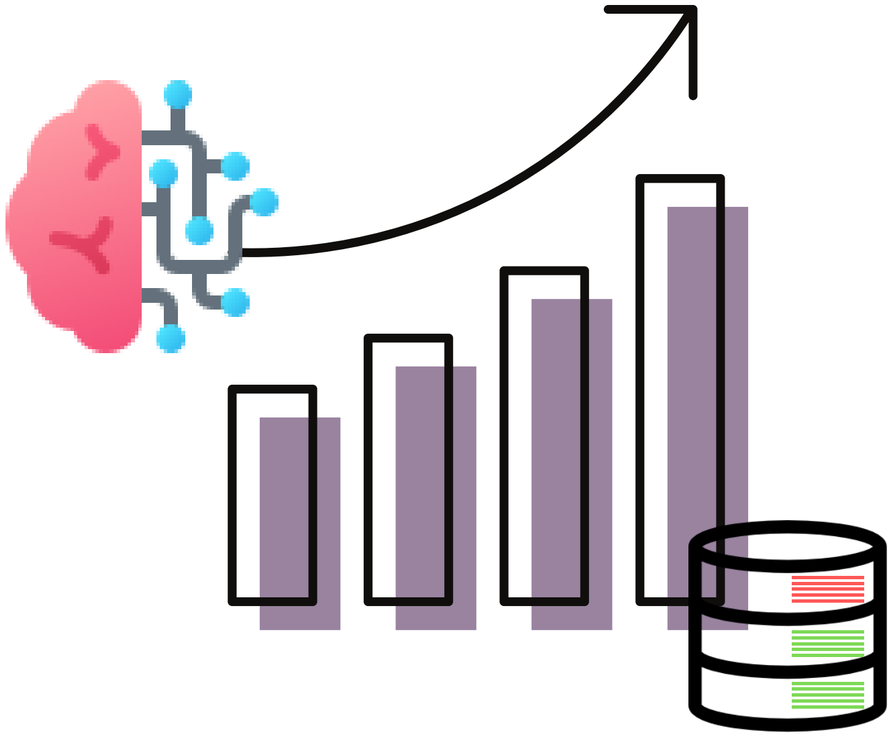
A importância do dataset



A importância do dataset



Reprodutibilidade do dataset



- validação
- comparação

Roteiro

- Objetivo

Roteiro

- Objetivo
- Metodologia

Roteiro

- Objetivo
- Metodologia
- Resultados

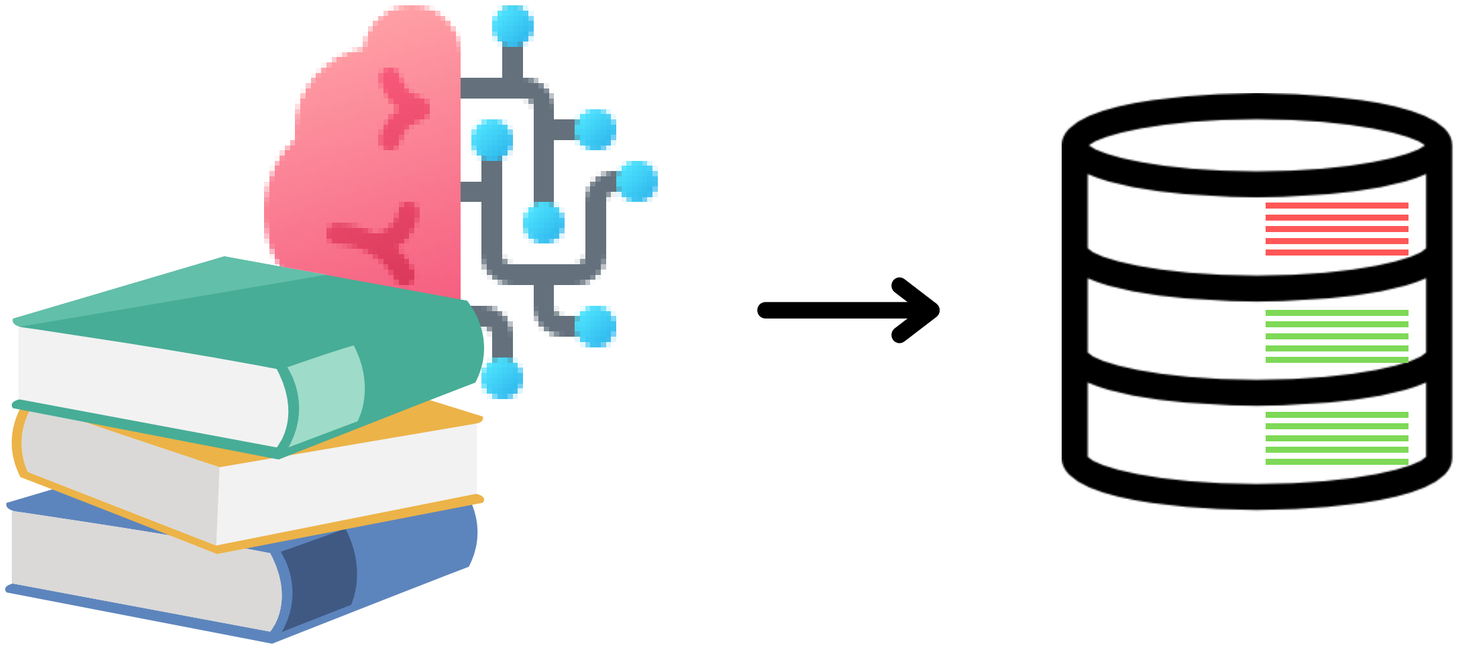
Roteiro

- Objetivo
- Metodologia
- Resultados
- Conclusão

Roteiro

- Objetivo
- Metodologia
- Resultados
- Conclusão
- Trabalhos Futuros

Objetivo



Metodologia

- Seleção dos trabalhos
- Análise dos trabalhos

Metodologia – seleção dos trabalhos

G1 -> survey ou revisão sistemática de literatura

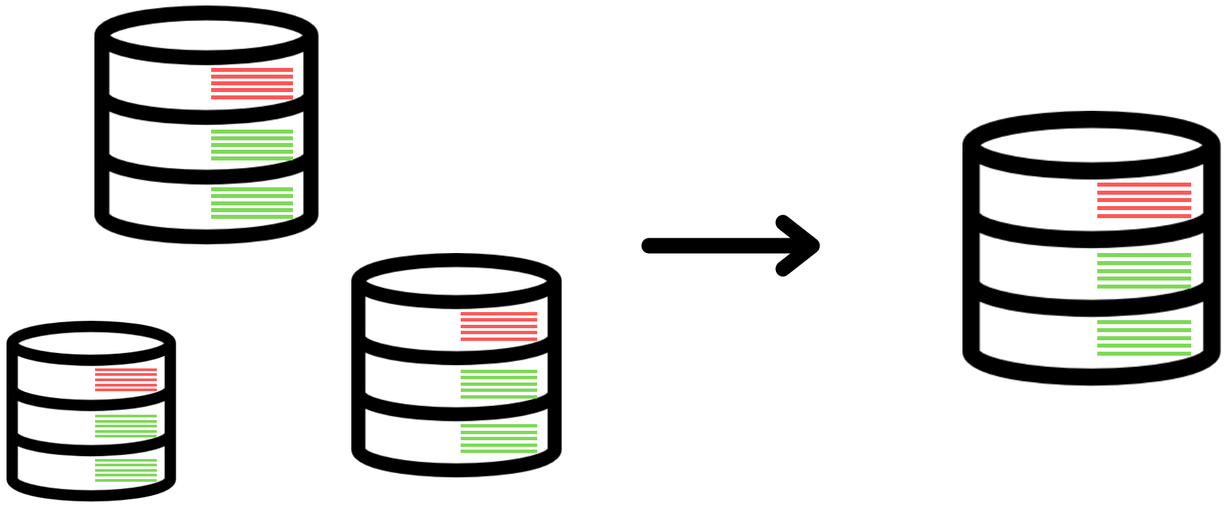
G2 -> 40 citações - Google Scholar

G3 -> conferências de segurança

G4 -> conferências de inteligência artificial

38

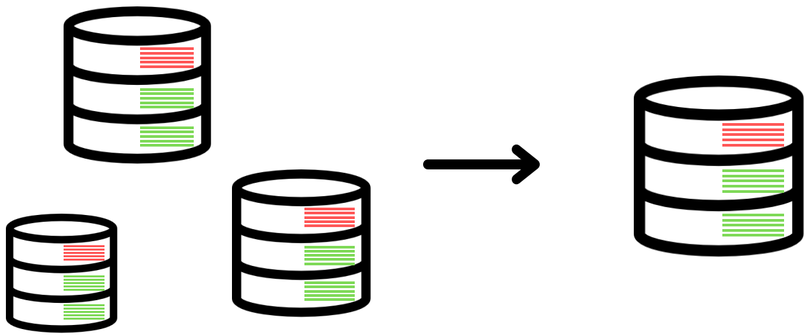
Metodologia - análise dos trabalhos



Metodologia – análise dos trabalhos

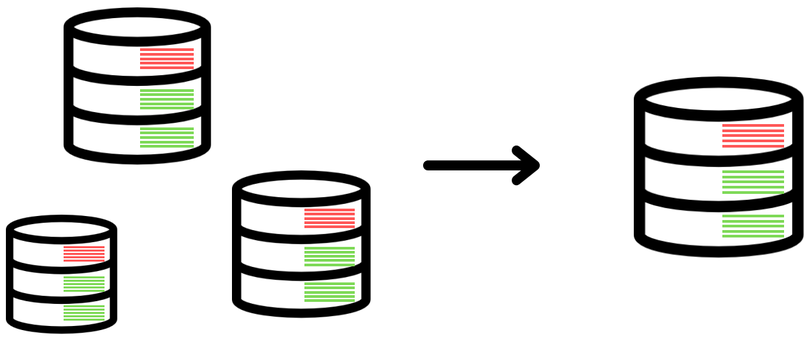
- *"Detection and Mitigation of Android Malware through Hybrid Approach"*

- DroidKin e ContagioDump



Metodologia – análise dos trabalhos

- *"Dynamic Permissions based Android Malware Detection using Machine Learning Techniques"*



- Genome, AndMalShare, DroidKin, Android Botnet e várias lojas

Metodologia – análise dos trabalhos

- Fontes

Metodologia – análise dos trabalhos

- Fontes
- Acessível

Metodologia – análise dos trabalhos

- Fontes
- Acessível
- Detalhamento

Metodologia – análise dos trabalhos

- 2 ou 3 co-autores

Metodologia – análise dos trabalhos

- 2 ou 3 co-autores
- Divergência: 1, 2 ou 3 revisores diferentes

Resultados

- Referência à origem
- Quantidades

80%

Resultados

- Referência à origem
- Quantidades

80%

1. *"A machine learning approach to Android malware detection"*

- 91 malwares - ?
- 2.081 apps benignos - ?

Resultados

- Referência à origem
- Quantidades

80%

2. *"Permission-based Android malware detection"*

- 700 (total) - ?

Resultados

- Referência à origem
- Quantidades
- Disponibilidade

90%

Resultados

- Referência à origem
- Quantidades
- Disponibilidade

90%

3. *"Machine learning in wavelet domain for electromagnetic emission based malware analysis"*

Resultados

- Referência à origem
- Quantidades
- Disponibilidade
- Detalhamento

100%

Resultados

- Referência à origem
- Quantidades
- Disponibilidade
- Detalhamento

100%

4. *"Andromaly: a behavioral malware detection framework for Android devices"*

Conclusão

- Todos os datasets são irreprodutíveis

Conclusão

- Todos os datasets são irreprodutíveis
  - Nenhum trabalho descreve suficientemente

Conclusão

- Todos os datasets são irreprodutíveis
  - Nenhum trabalho descreve suficientemente
  - Nenhum trabalho disponibiliza dataset

Conclusão

- Recomendações
  - Fontes públicas
  - Detalhamento
  - Disponibilização

Trabalho Futuro

- Reprodutibilidade dos modelos de aprendizado de máquina

Obrigada pela atenção!