

Detecção de Malwares Android: reprodução da seleção de características do SigPID

Joner Assolin, Vanderson Rocha, Guilherme Silveira, Gustavo Cardozo, Karina Casola, Eduardo Feitosa, Diego Kreutz

2021



UFAM



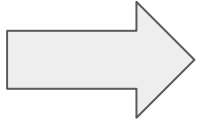
Detecção de malwares Android

Dataset



Detecção de malwares Android

Dataset



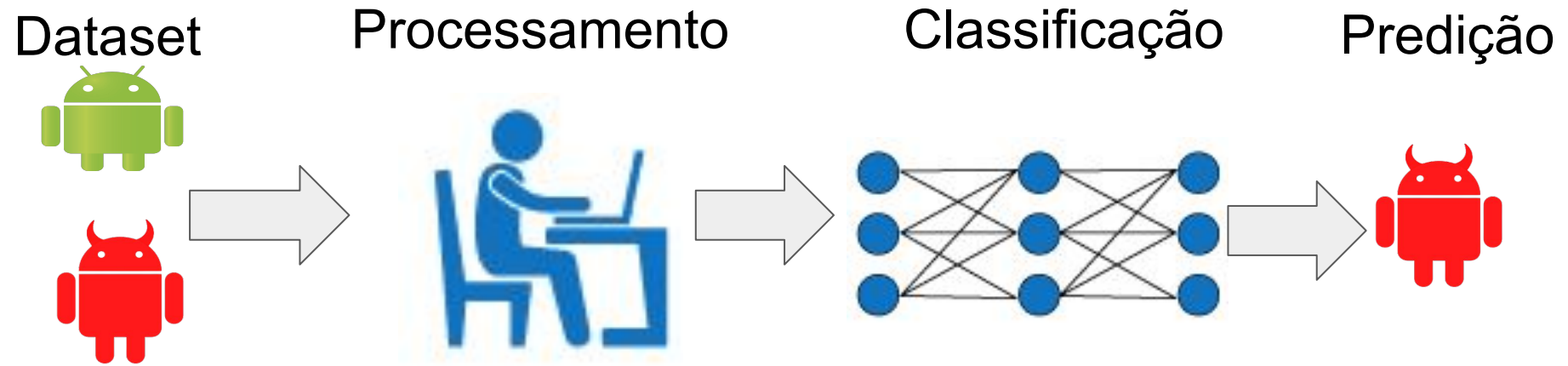
Processamento



Detecção de malwares Android

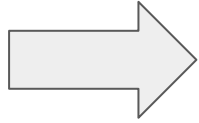
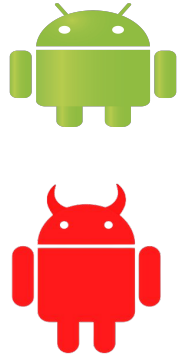


Detecção de malwares Android

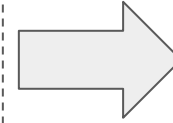


Detecção de malwares Android e o desafio da escalabilidade

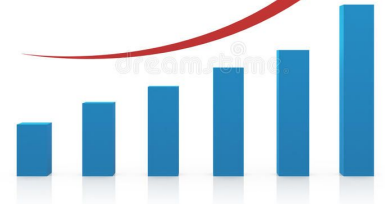
Dataset



Características



Tempo de Treino



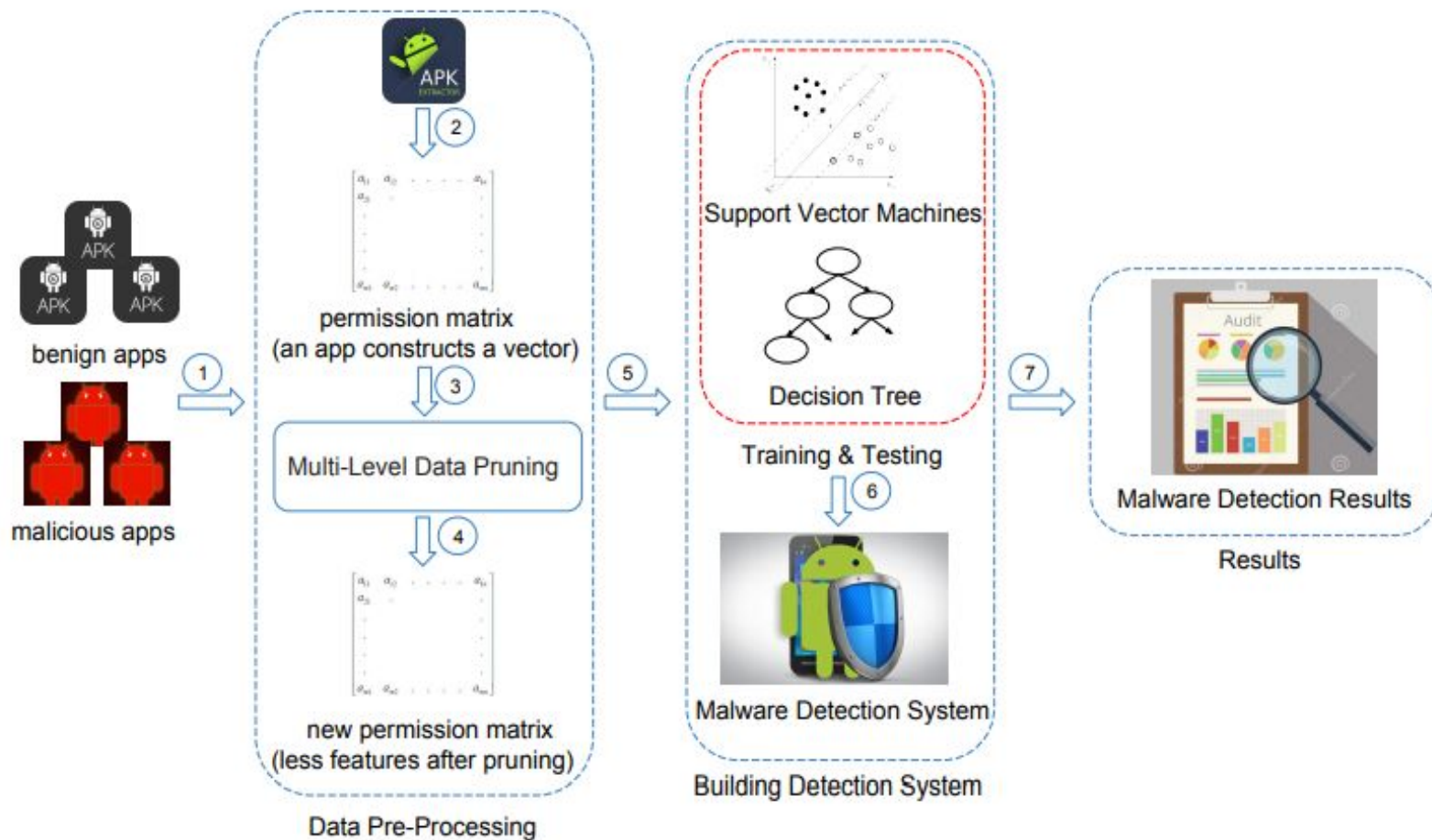


”

Será que existe trabalho na literatura focado na redução de características para aumentar o desempenho?



Seleção de características SigPID



Seleção de características SigPID

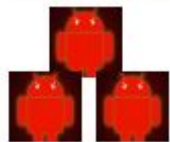


135

Permissões



benign apps



malicious apps



permission matrix
(an app constructs a vector)



Multi-Level Data Pruning

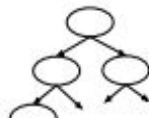


new permission matrix
(less features after pruning)

Data Pre-Processing



Support Vector Machines



Decision Tree

Training & Testing



Malware Detection System

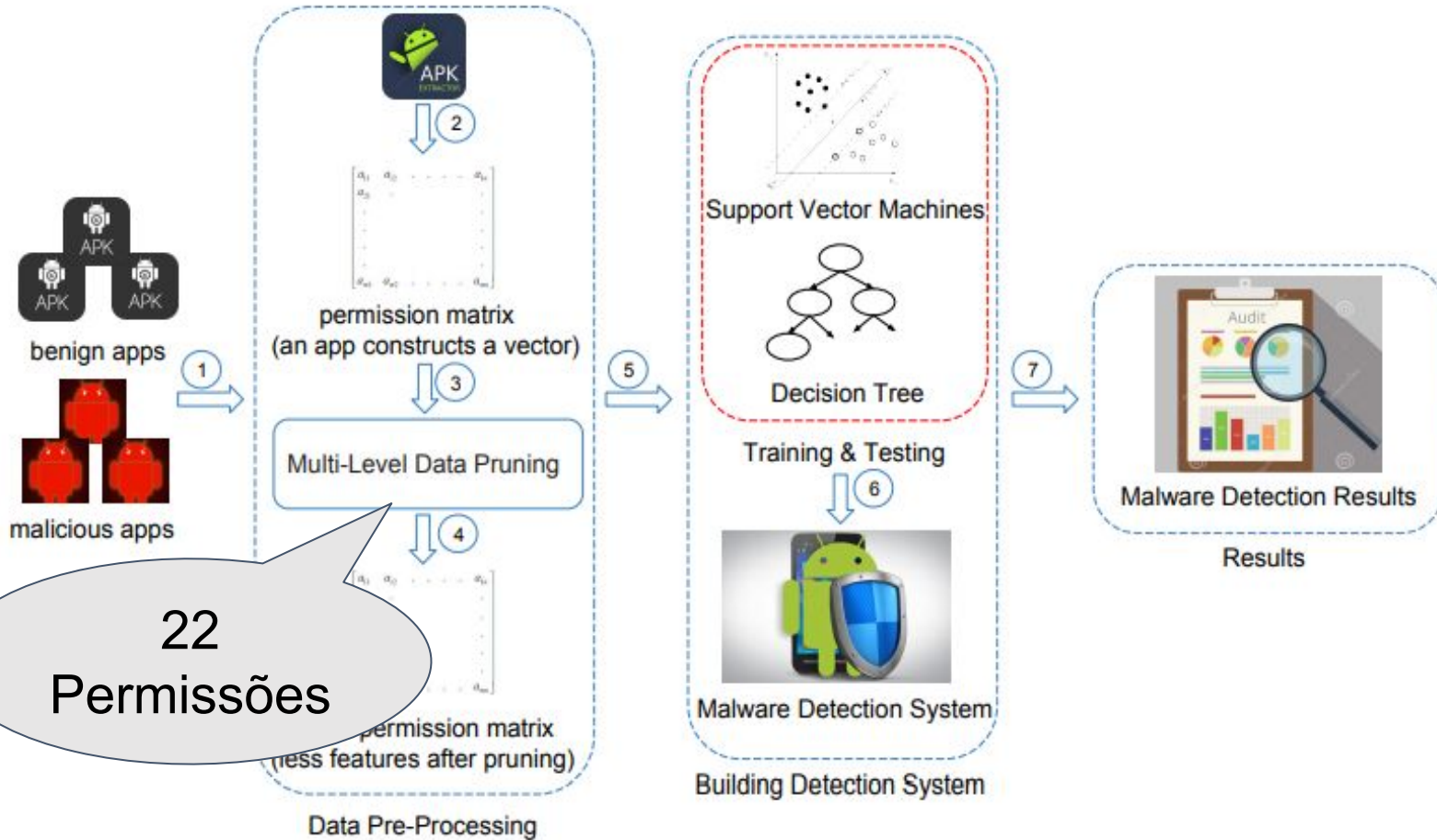
Building Detection System



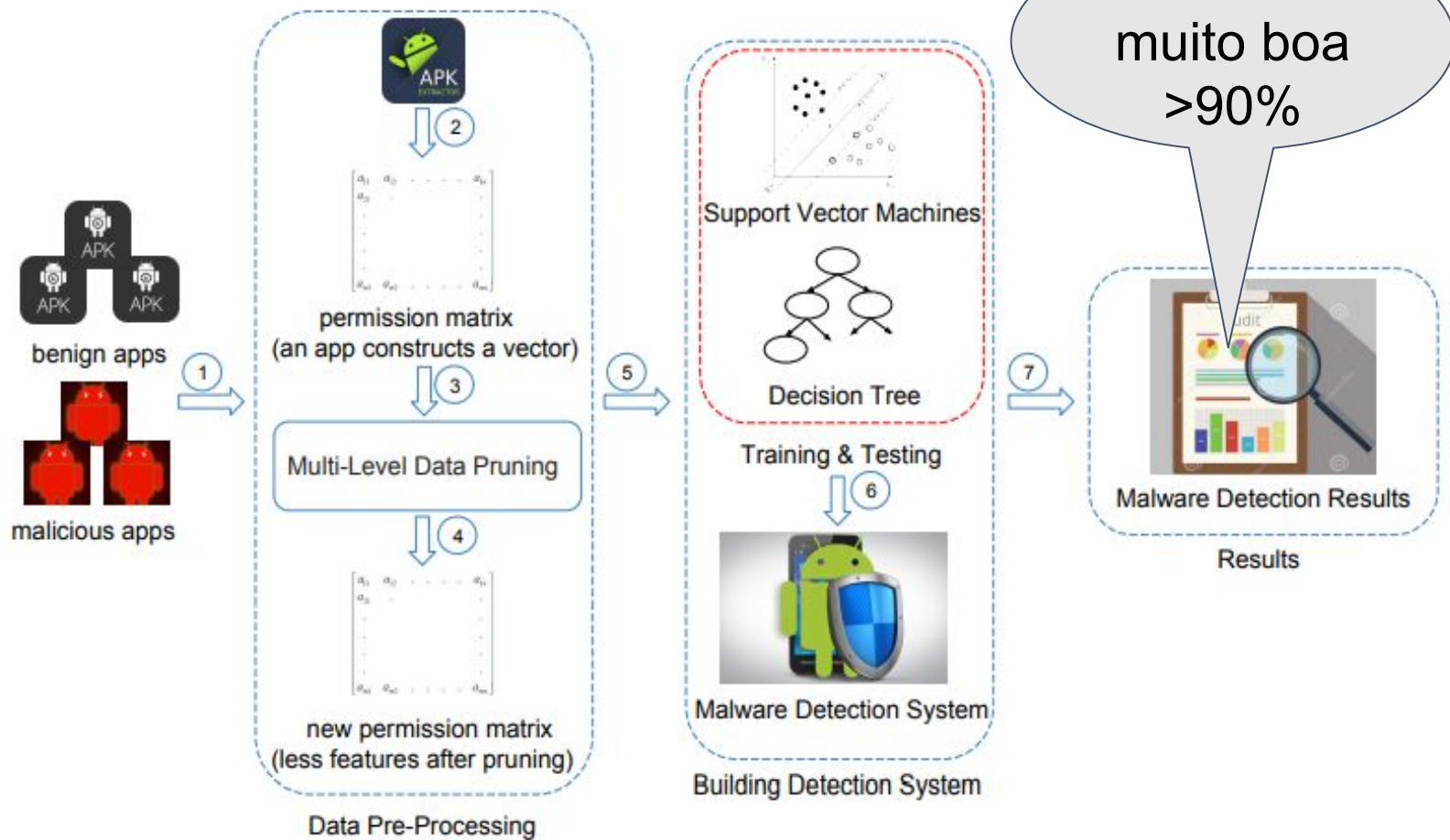
Malware Detection Results

Results

Seleção de características SigPID



Seleção de características SigPID



Objetivos

- **Replicar o SigPID com dataset conhecido**
- **Comparar com o trabalho original(SigPID)**
- **Comparar com outros conjuntos de permissões**



Etapas

- **Etapa 1 Análise de reprodutibilidade do SigPID**
- **Etapa 2 Reprodução do SigPID utilizando um dataset público**

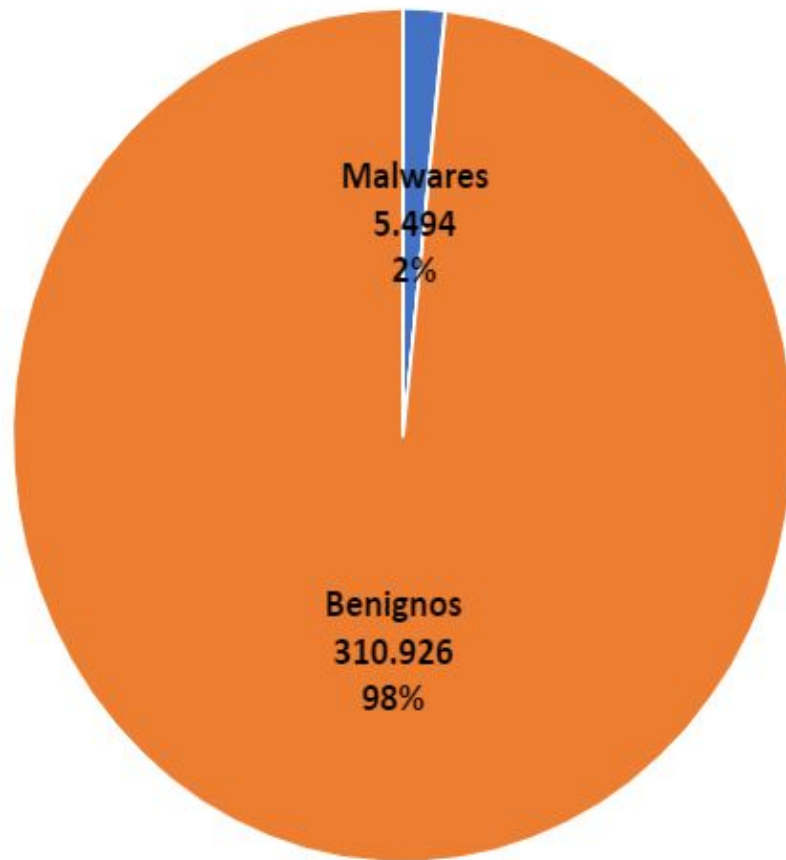




Etapa 1: Análise da reprodutibilidade

Dataset

- **SigPID**
- **Amostra**
 - Benignos
(310,926)
 - Malwares
(5.494)



Dataset

- **SigPID**
- **Disponibilidade**
 - Total 315,794 aplicativos
 - Google play(310,926) ✓*
 - Mal Zhou(1,260) ✗
 - Mal Com1(247) e Mal Com2(154) ✗
 - Mal VS(3,207) ✓*





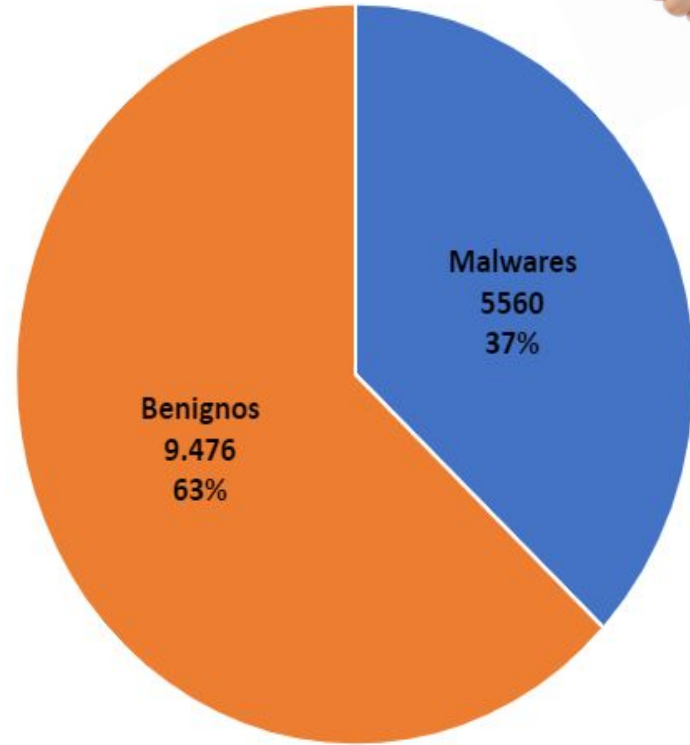
Etapa 2: Reprodução do SigPID

Dataset público escolhido (Drebin_215)



- **Amostra**

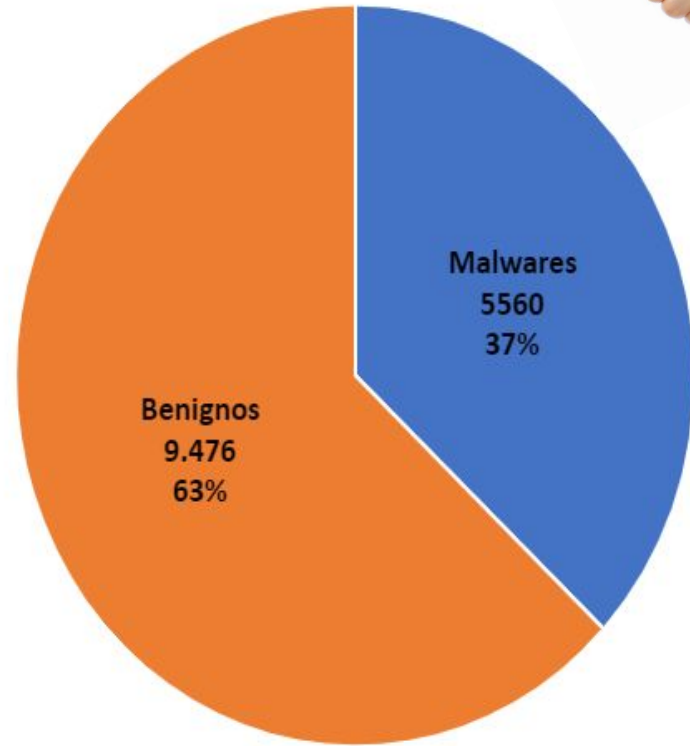
- 5.560 Malwares
- 9.476 Benignos



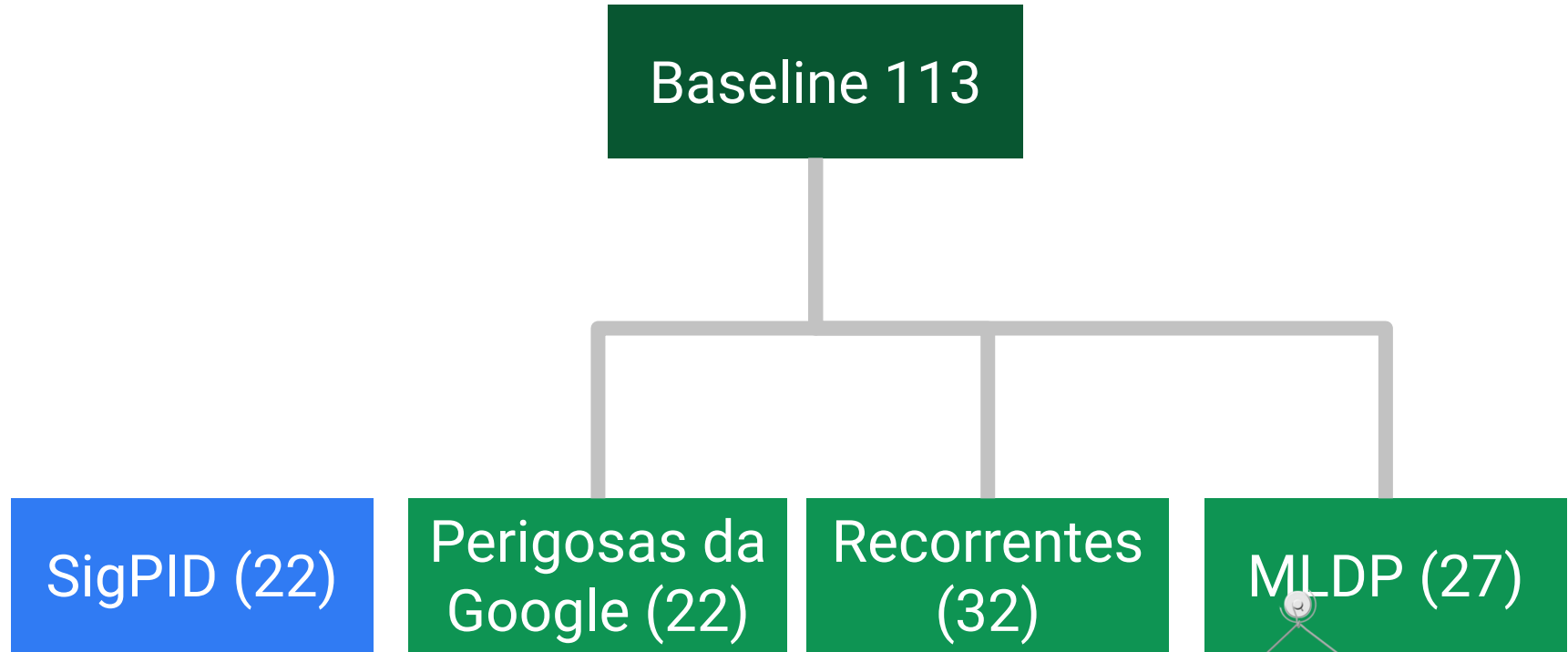
Dataset público escolhido (Drebin_215)



- **Amostra**
 - 5.560 Malwares
 - 9.476 Benignos
- **113 permissões**



Datasets





| Seleção de dados multinível (MLDP)

Os 3 níveis de seleção do MLDP

- PRNR
- SPR
- PMAR



PRNR

Classificação
de permissão
com taxa
negativa



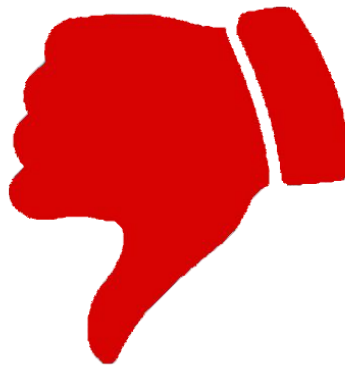
Classificação de permissão com taxa negativa (PRNR)

Ranking ≈ 1



Alto Risco

Ranking ≈ -1



Baixo risco

Ranking = 0



Irrelevante



Classificação de permissão com taxa negativa (PRNR)



Ordem
crescente

Permissões	R	Permissões	R
READ_CALENDAR	-1	WRITE_SMS	1
VIBRATE	-1	READ_SMS	1
RECORD_AUDIO	-0.3	CAMERA	0
CAMERA	0	INTERNET	0
INTERNET	0	RECORD_AUDIO	-0.3
READ_SMS	1	READ_CALENDAR	-1
WRITE_SMS	1	VIBRATE	-1

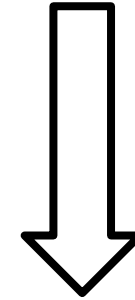


Ordem
decrescente



Sistema incremental de permissão(PIS)

Permissões	R	Permissões	R
READ_CALENDAR	-1	WRITE_SMS	1
VIBRATE	-1	READ_SMS	1
RECORD_AUDIO	-0.3	CAMERA	0
CAMERA	0	INTERNET	0
INTERNET	0	RECORD_AUDIO	-0.3
READ_SMS	1	READ_CALENDAR	-1
WRITE_SMS	1	VIBRATE	-1



SVM

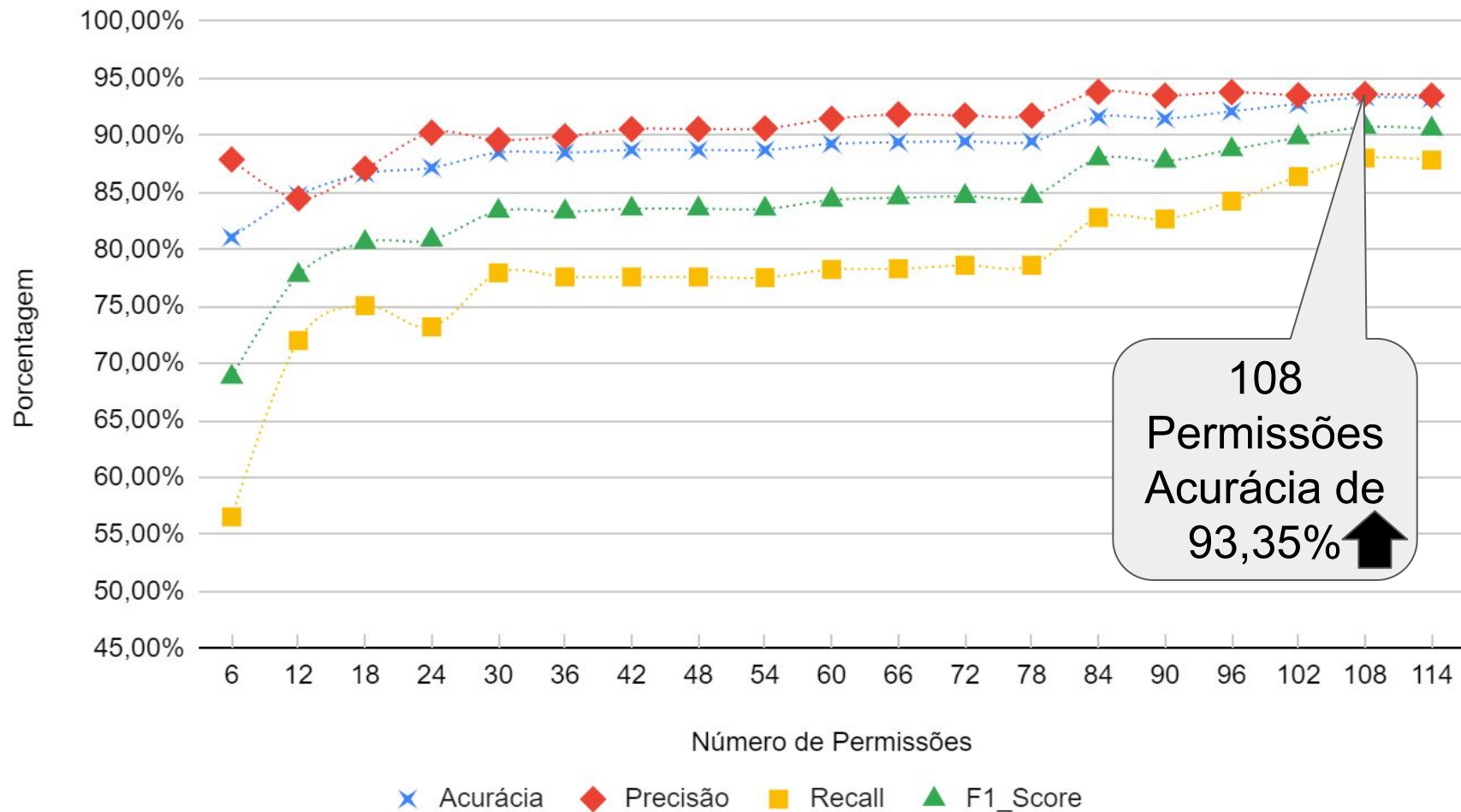
Acurácia

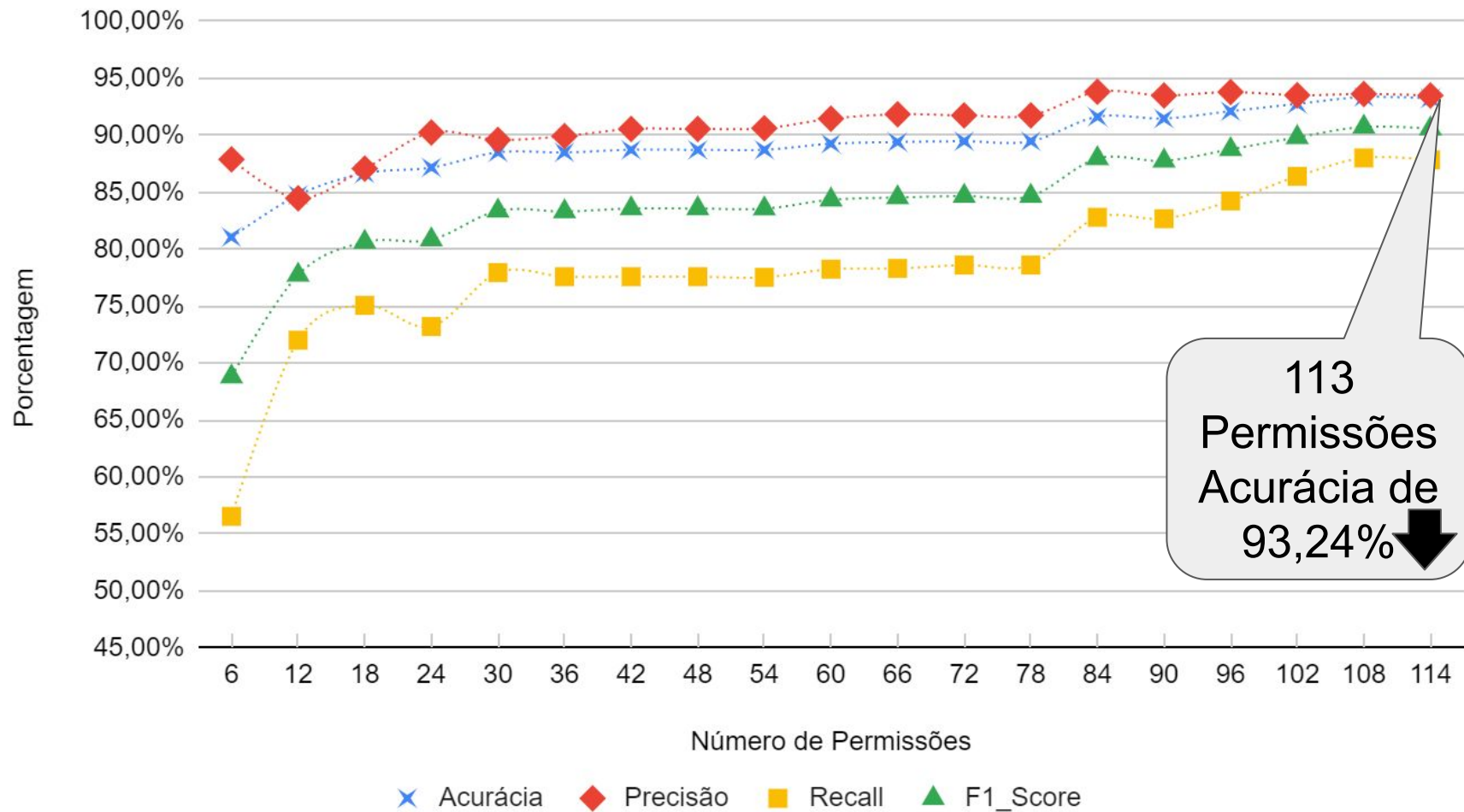
Precisão

Recall

F1_Score







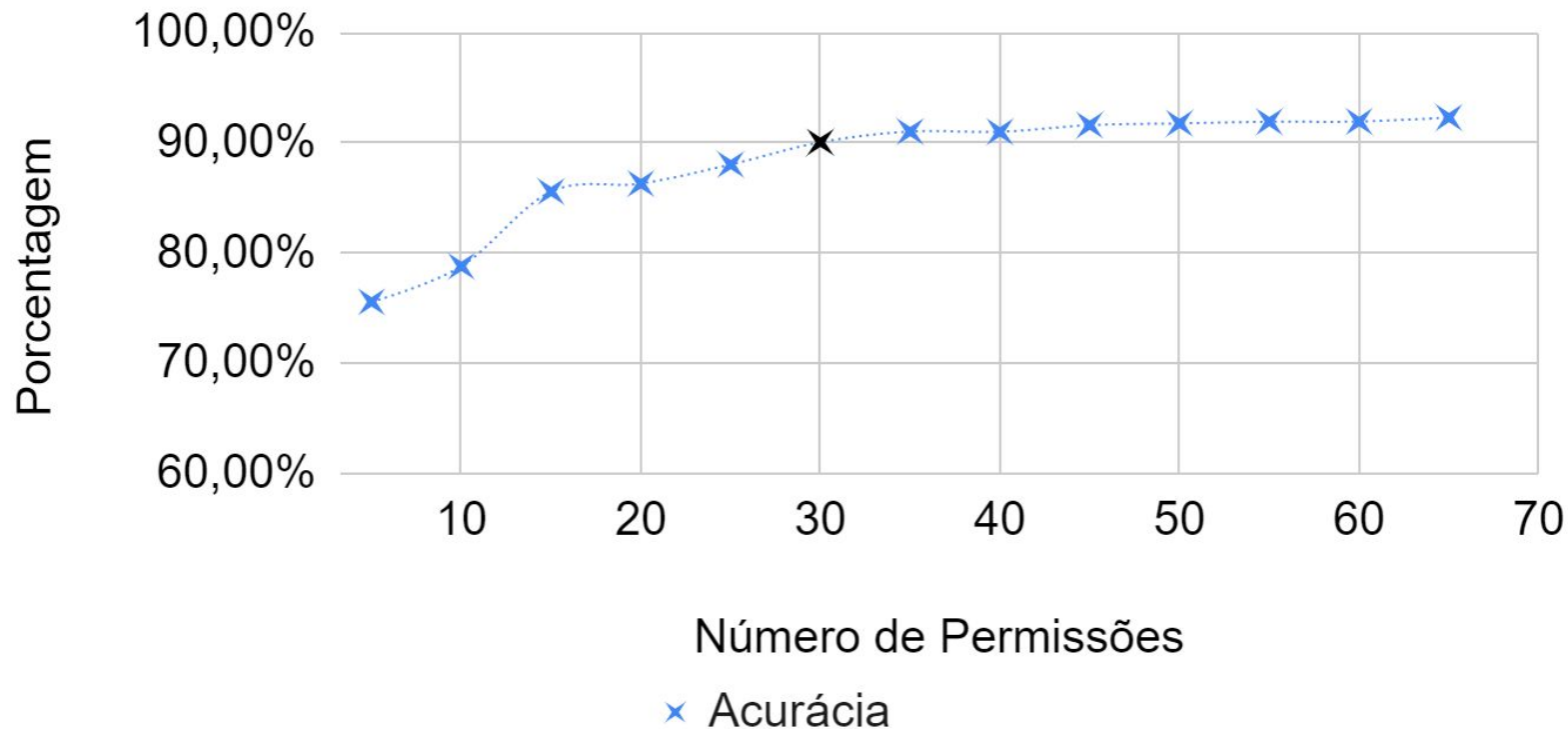
SPR

Classificação
de permissão
baseada em
suporte



PRNR 108

SPR 30



SVM

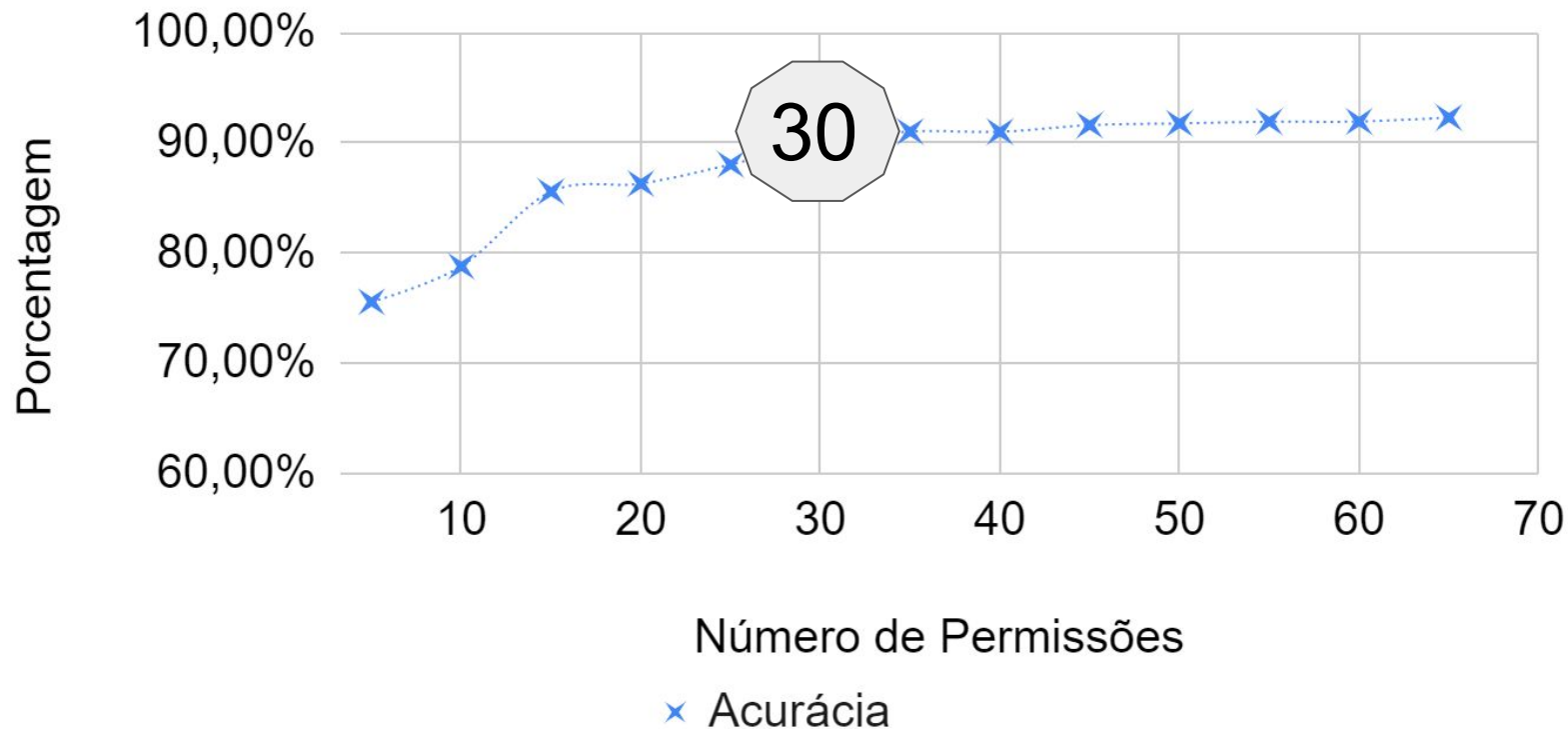
Acurácia

>90%



PRNR 108

SPR 30



SVM

Acurácia

>90%



PMAR

Mineração de
permissões com
regras de
associação



96,5% de confiança mínima e 10% de suporte mínimo

Antecedentes	Consequentes
CHANGE_WIFI_STATE	ACCESS_WIFI_STATE
MANAGE_ACCOUNTS	GET_ACCOUNTS
WRITE_SMS	READ_SMS





Treino e Teste



matplotlib

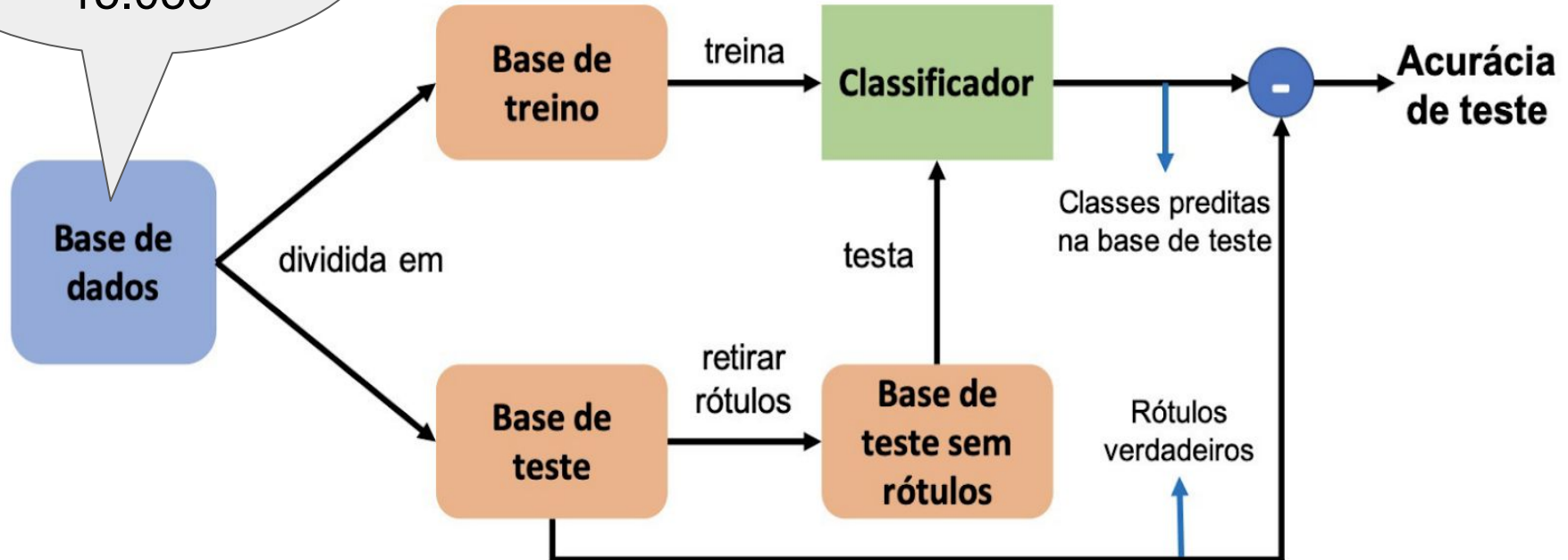


NumPy



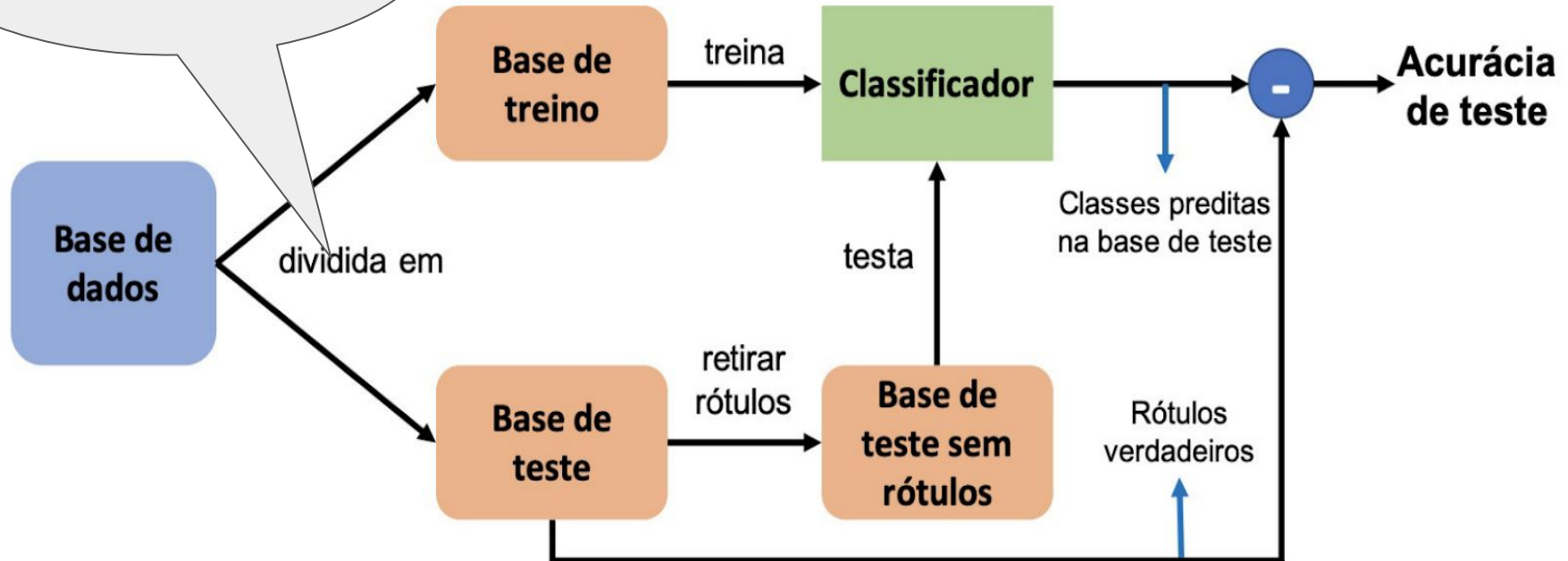
Treino e teste dos conjuntos de dados

Drebin_215
Tamanho
15.036



Treino e teste dos conjuntos de dados

70% - Treino
30% - Teste





Resultados

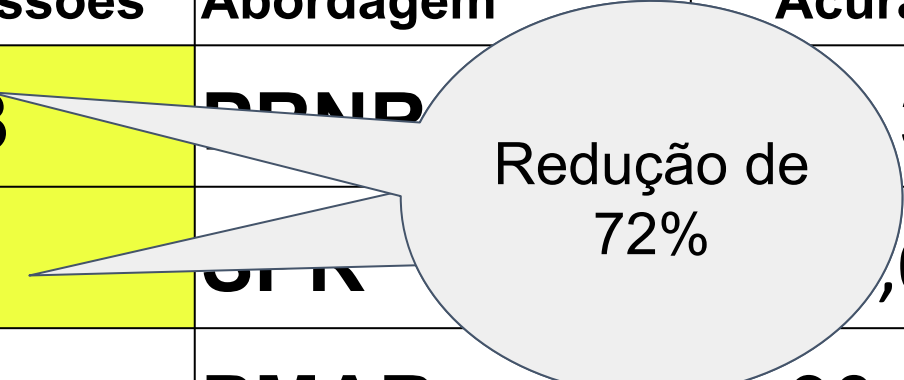
SVM

Nº de Permissões	Abordagem	Acurácia	Tempo Execução
108	PRNR	93,35	5,44
30	SPR	90,07	2,41
27	PMAR	90,05	2,26



SVM

Nº de Permissões	Abordagem	Acurácia	Tempo Execução
108	DDND	35	5,44
30	OTR	,07	2,41
27	PMAR	90,05	2,26



Redução de 72%

The diagram illustrates a significant reduction in the number of permissions. It features a light gray oval containing the text 'Redução de 72%'. Two white arrows point from this oval to the 'Nº de Permissões' column of the table. The first arrow points to the value '108' in the row for the 'DDND' approach. The second arrow points to the value '30' in the row for the 'OTR' approach, demonstrating a 72% decrease in permissions.



SVM

Nº de Permissões	Ordagem	Acurácia	Tempo Execução
108	PMAR	93,35	5,44s
30		90,07	2,41s
27		90,05	2,26s

Boa Acurácia



SVM

Nº de Permissões	Abordagem	Acurácia	Tempo Execução
108	PRNF	90,07	5,44s
30	SPR	90,07	2,41s
27	PMAR	90,05	2,26s

Boa
Redução no
tempo de
execução



SVM

Nº de Permissões	Abordagem	Acurácia	Tempo Execução
108	PRNR	35	5,44s
30	SPP	,07	2,41s
27	PMAR	90,05	2,26s

Redução de 75%



SVM

Nº de Permissões	Abordagem	Acurácia	Tempo Execução
108	PRNR	92,25	5,44s
30	SPR	92,07	2,41s
27	PMAR	90,05	2,26s

Diferença de 3.18s



SVM

Nº de Permissões	Abordagem	Acurácia	Tempo Execução
10	R	93,35	5,44s
3		90,07	2,41s
27	PMAR	90,05	2,26s

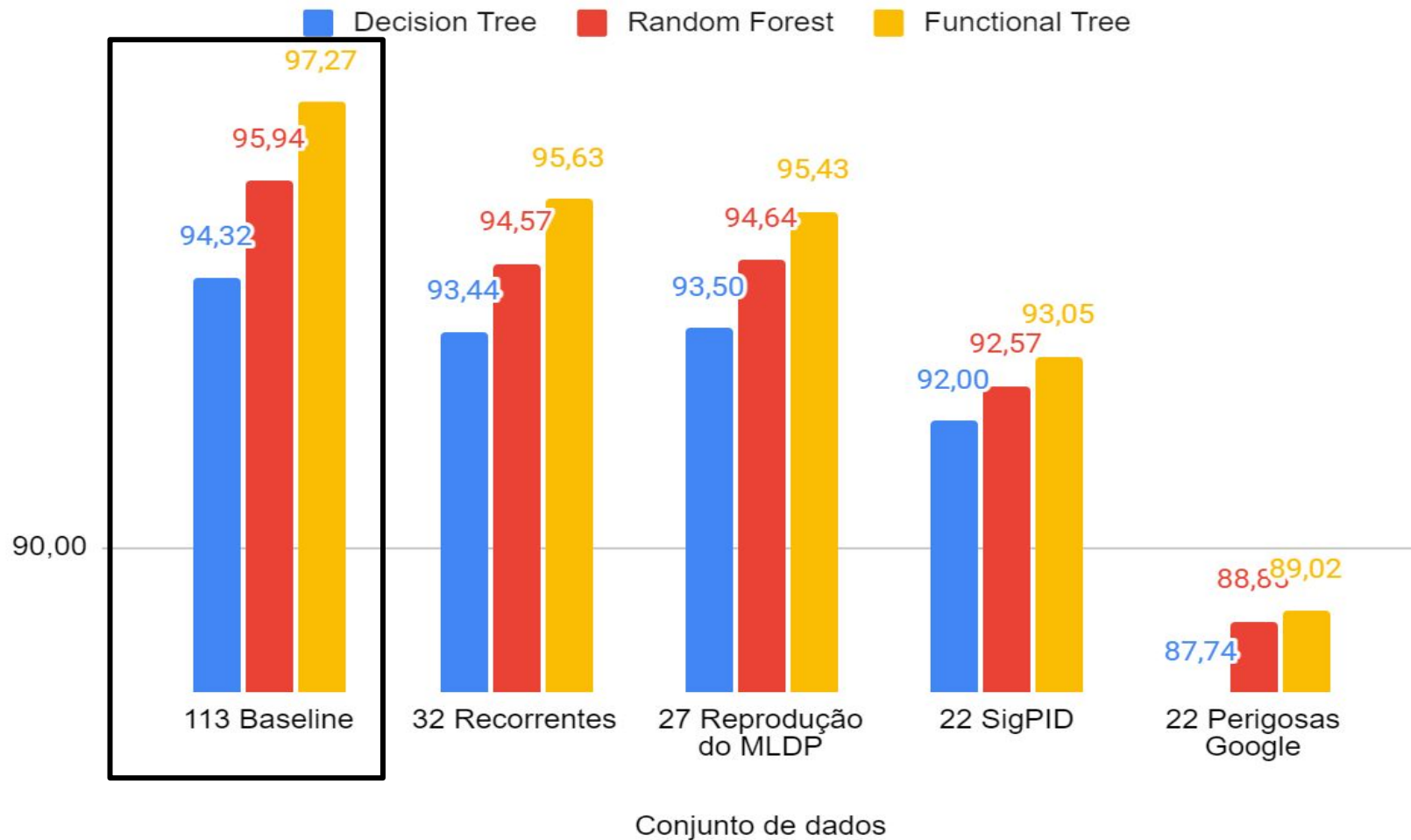
Boa
Acurácia



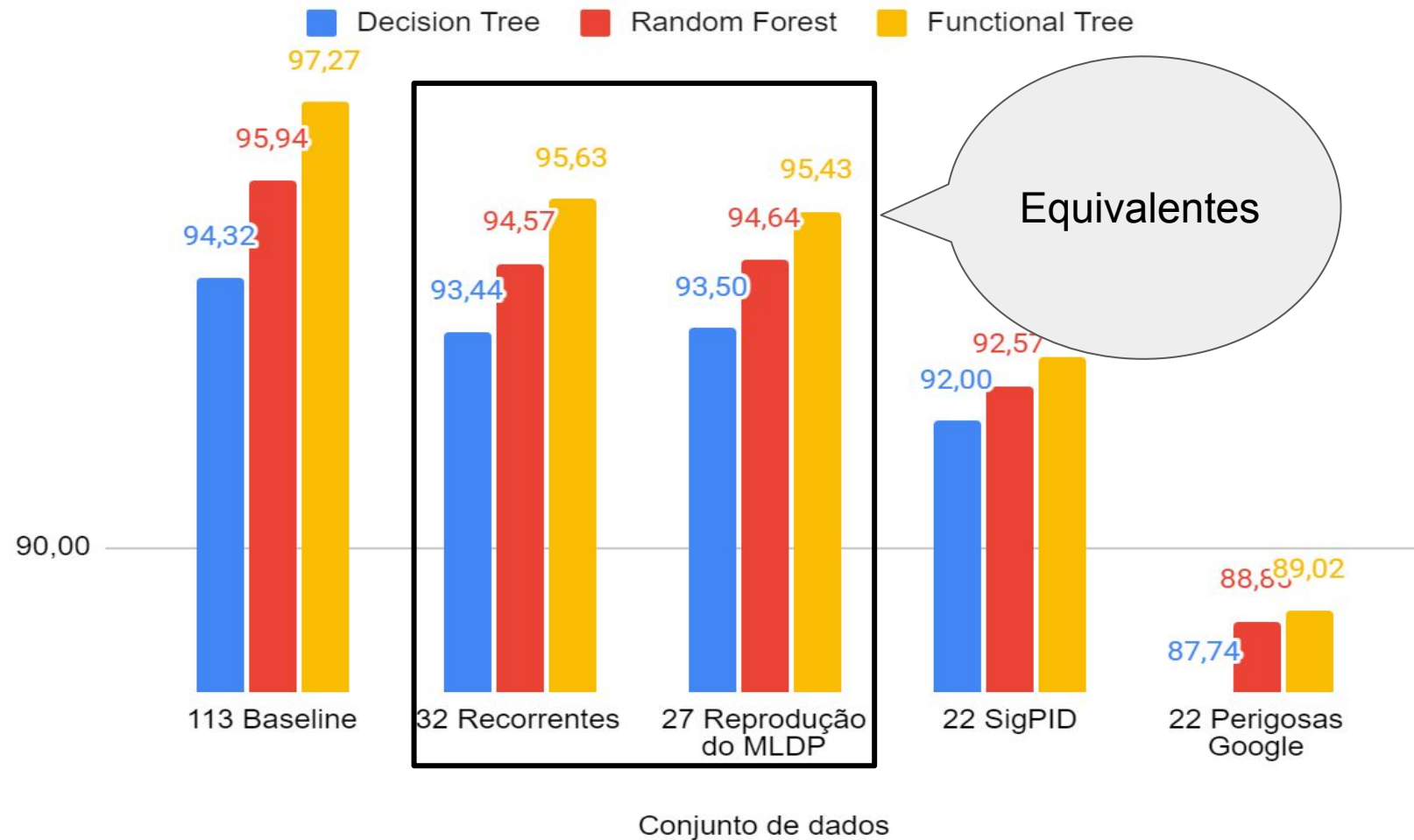
- **Decision Tree**
- **Random Forest**
- **Functional Trees**



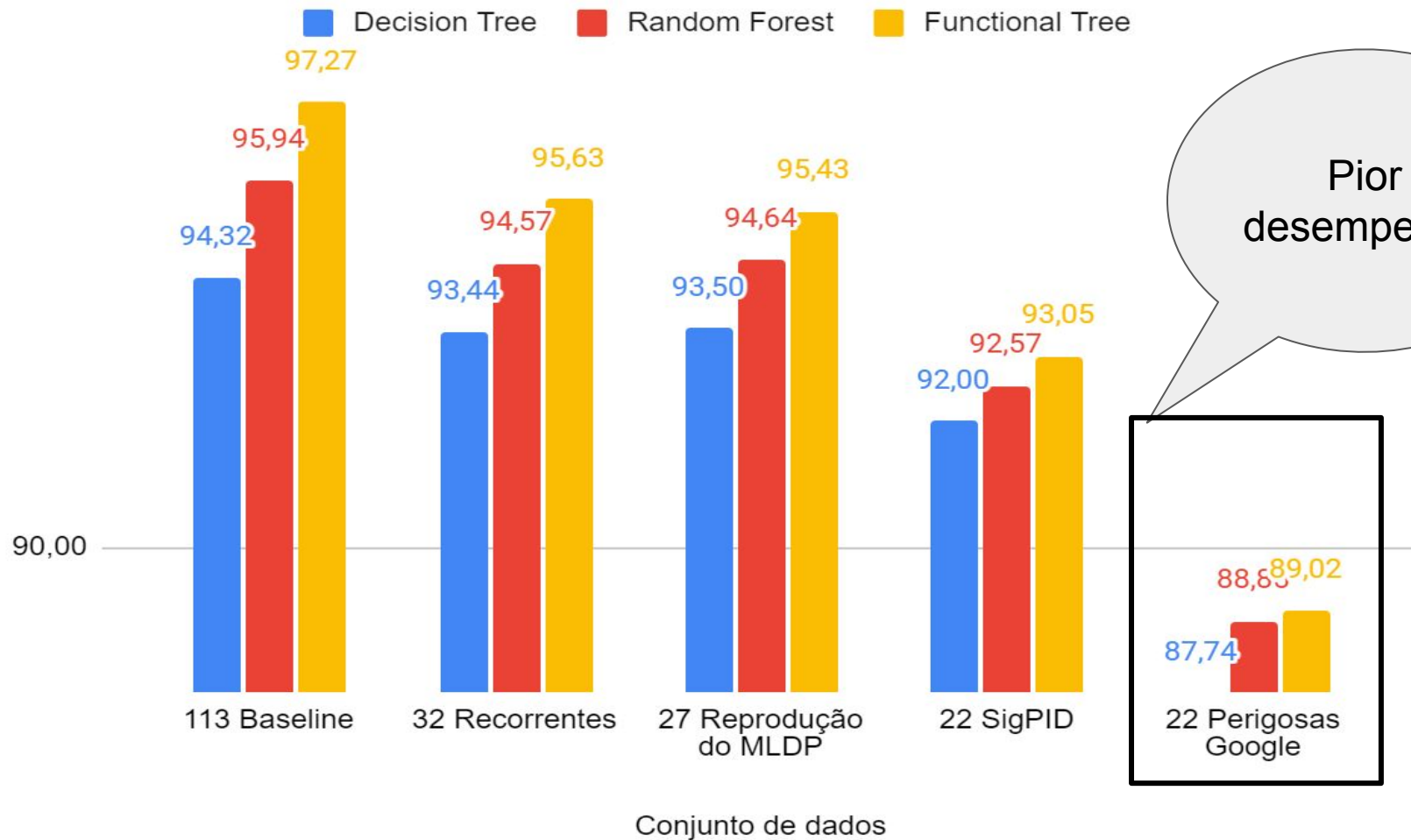
Acurácia



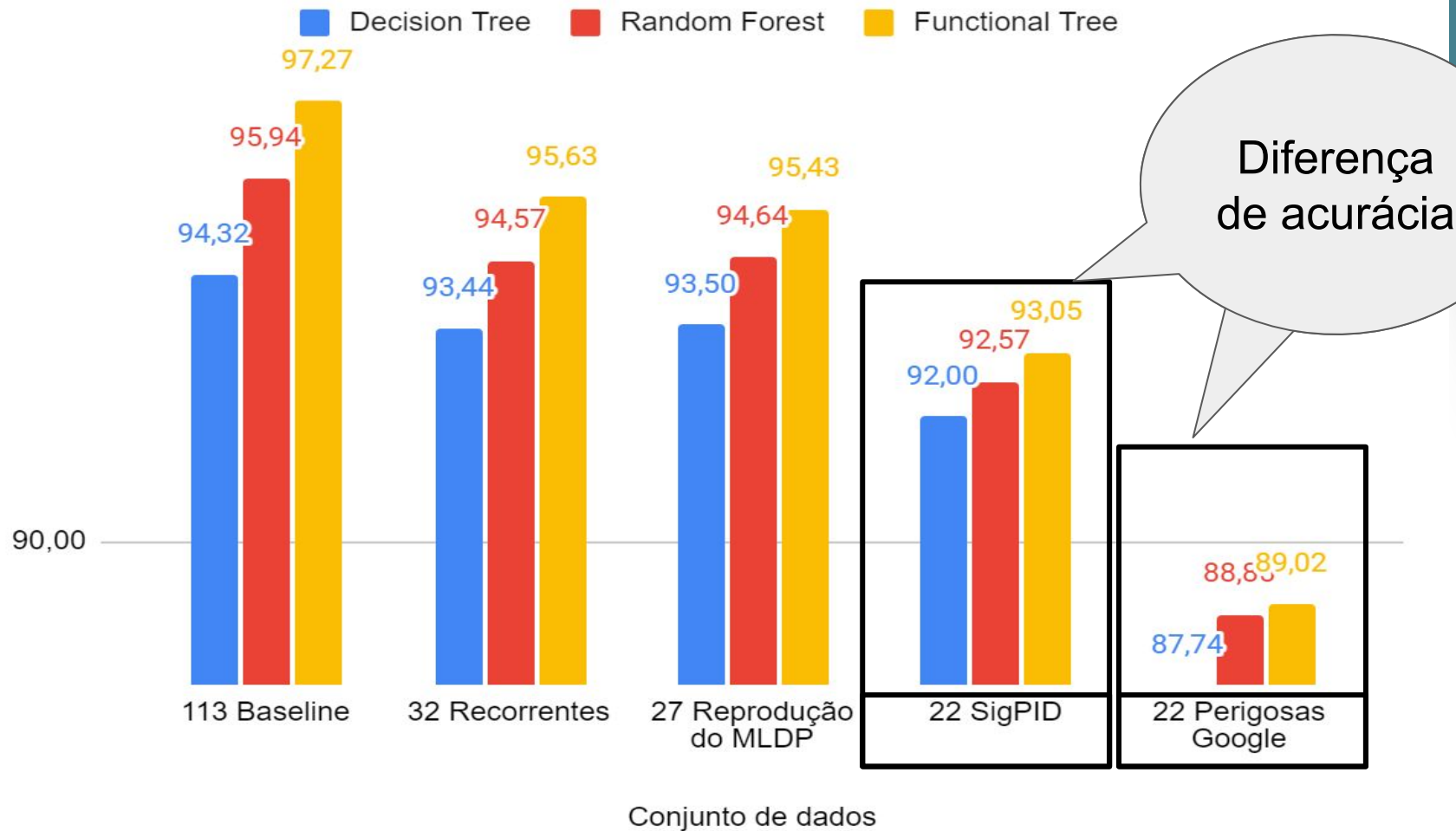
Acurácia



Acurácia



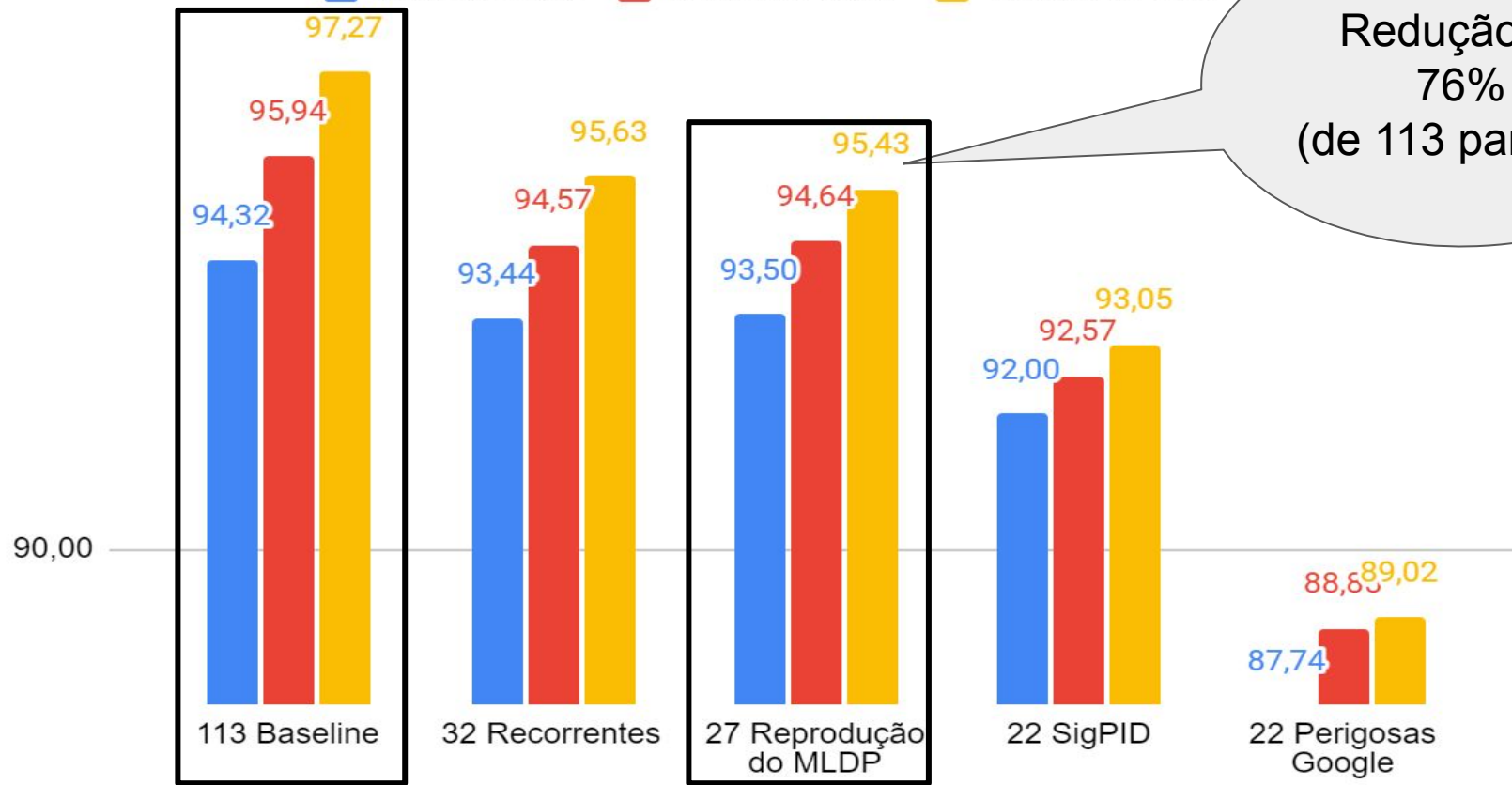
Acurácia



Acurácia

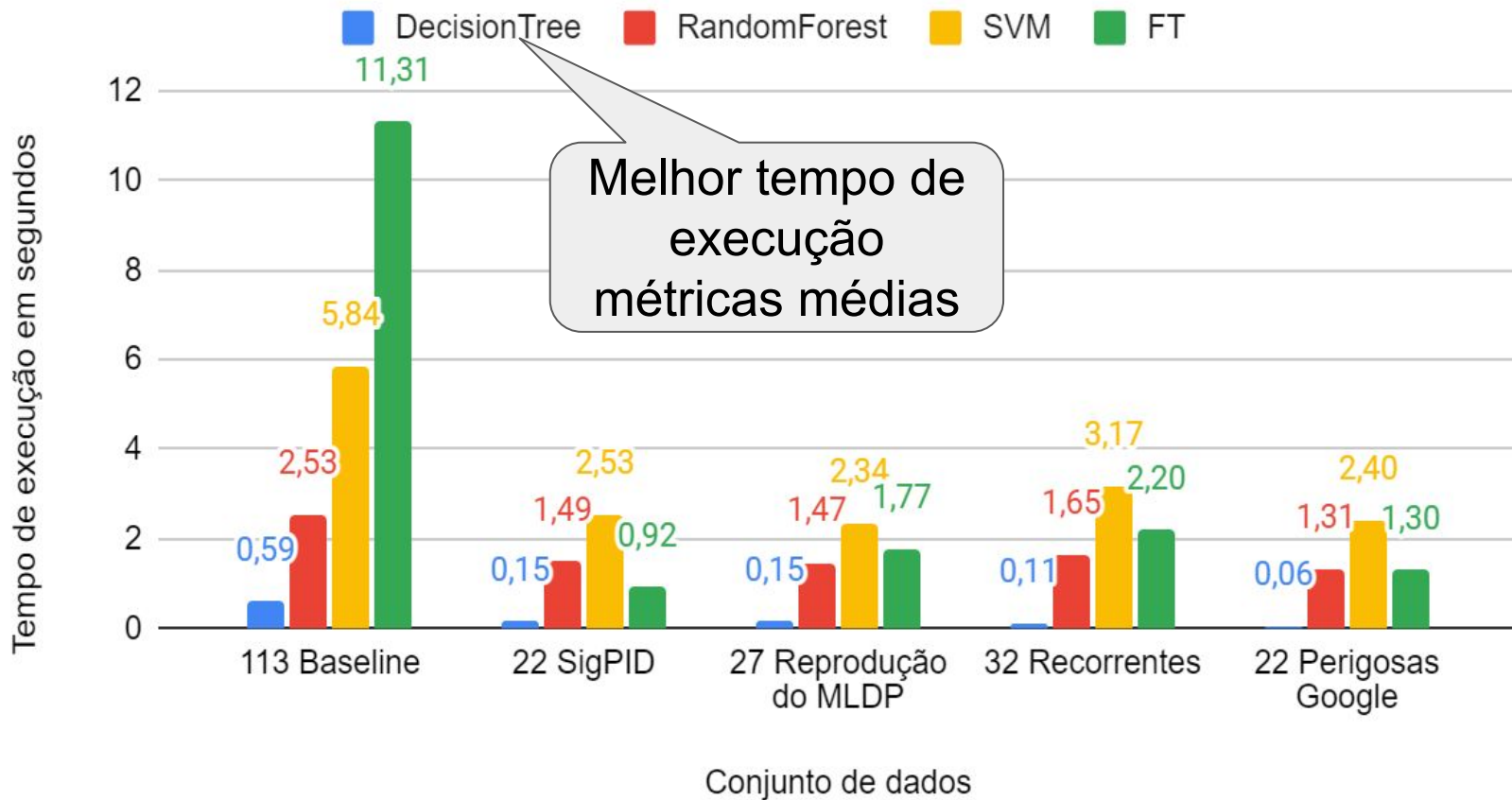


Decision Tree Random Forest Functional Tree

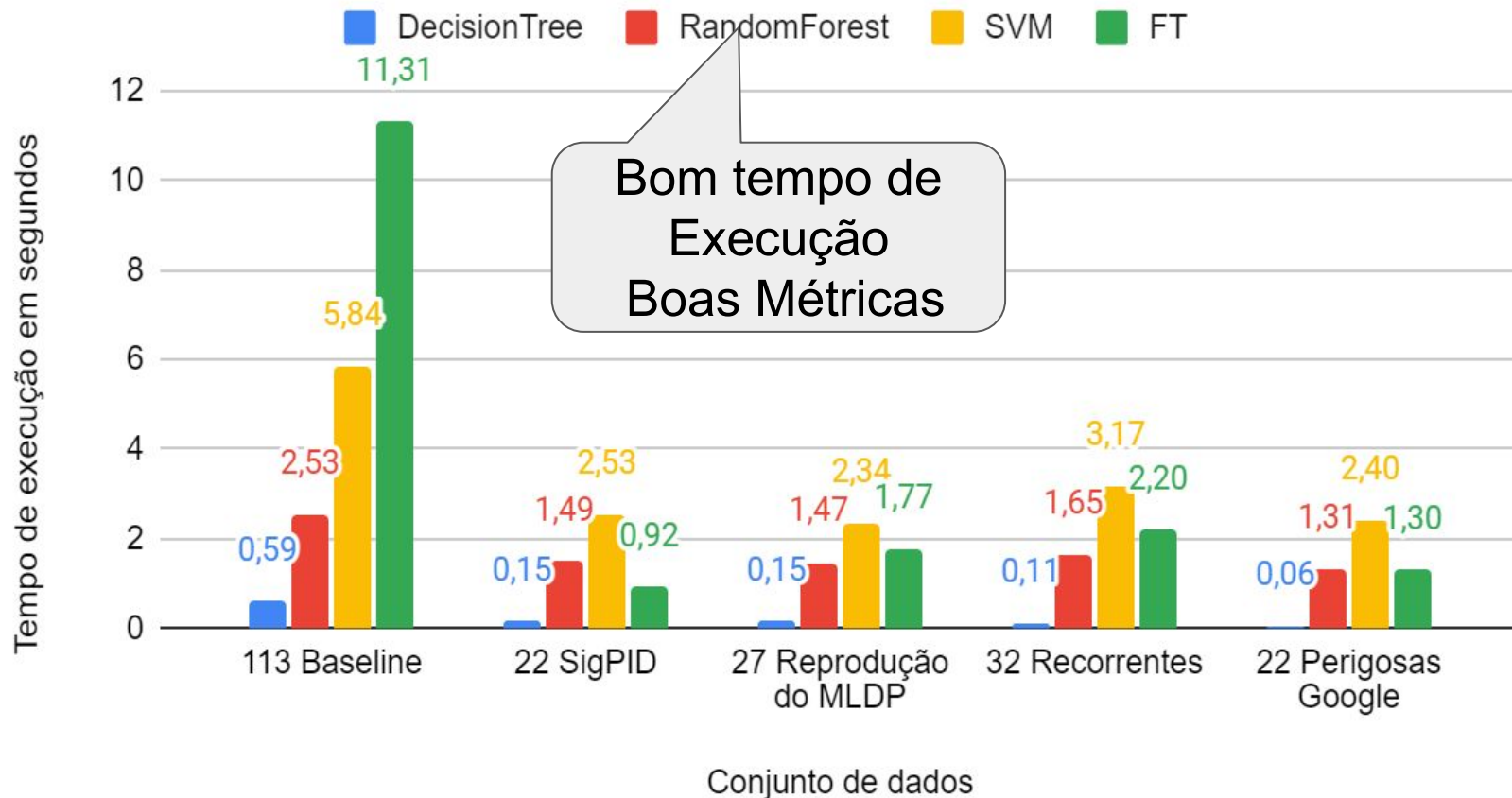


Conjunto de dados

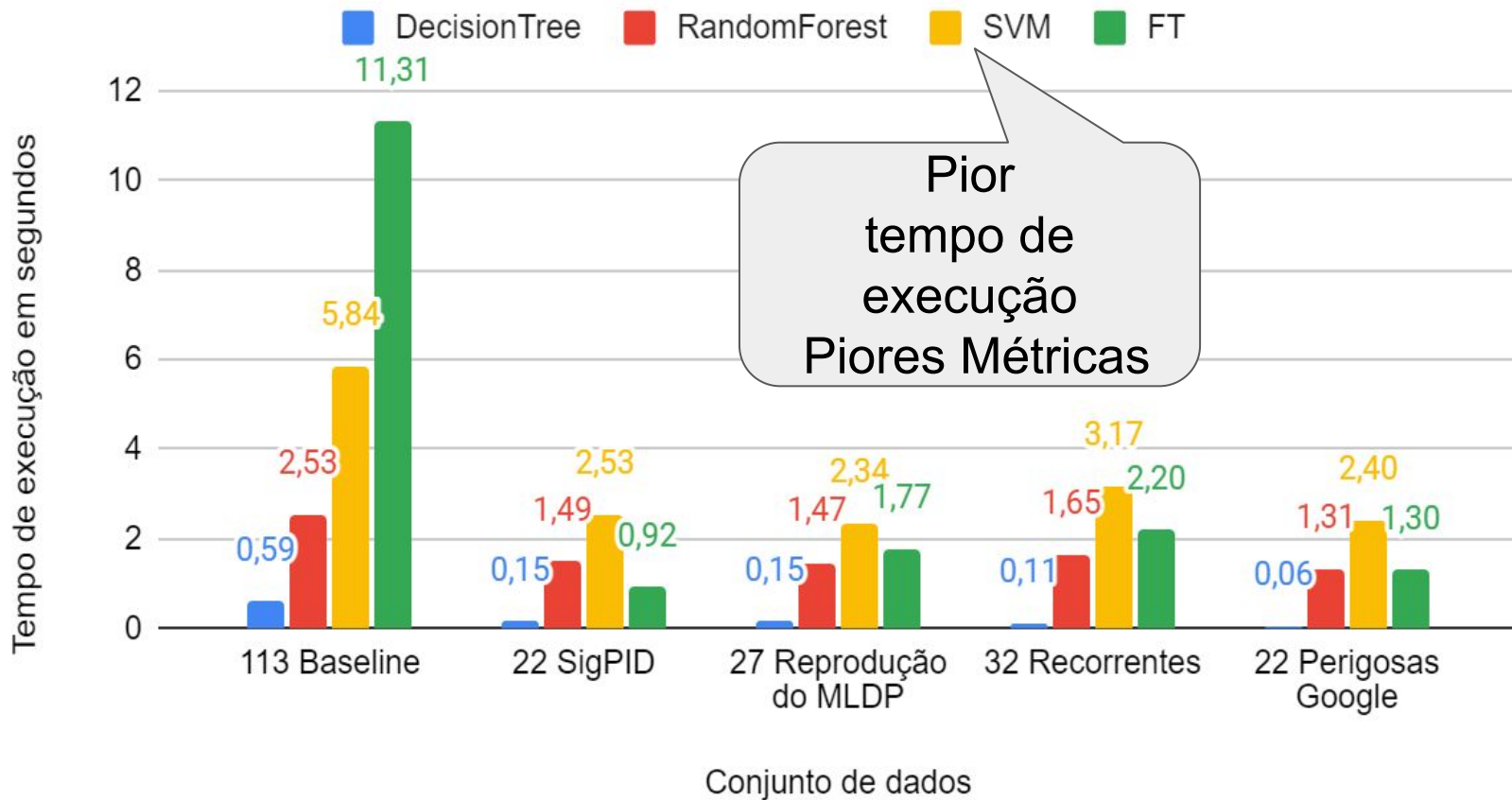
Tempo de execução



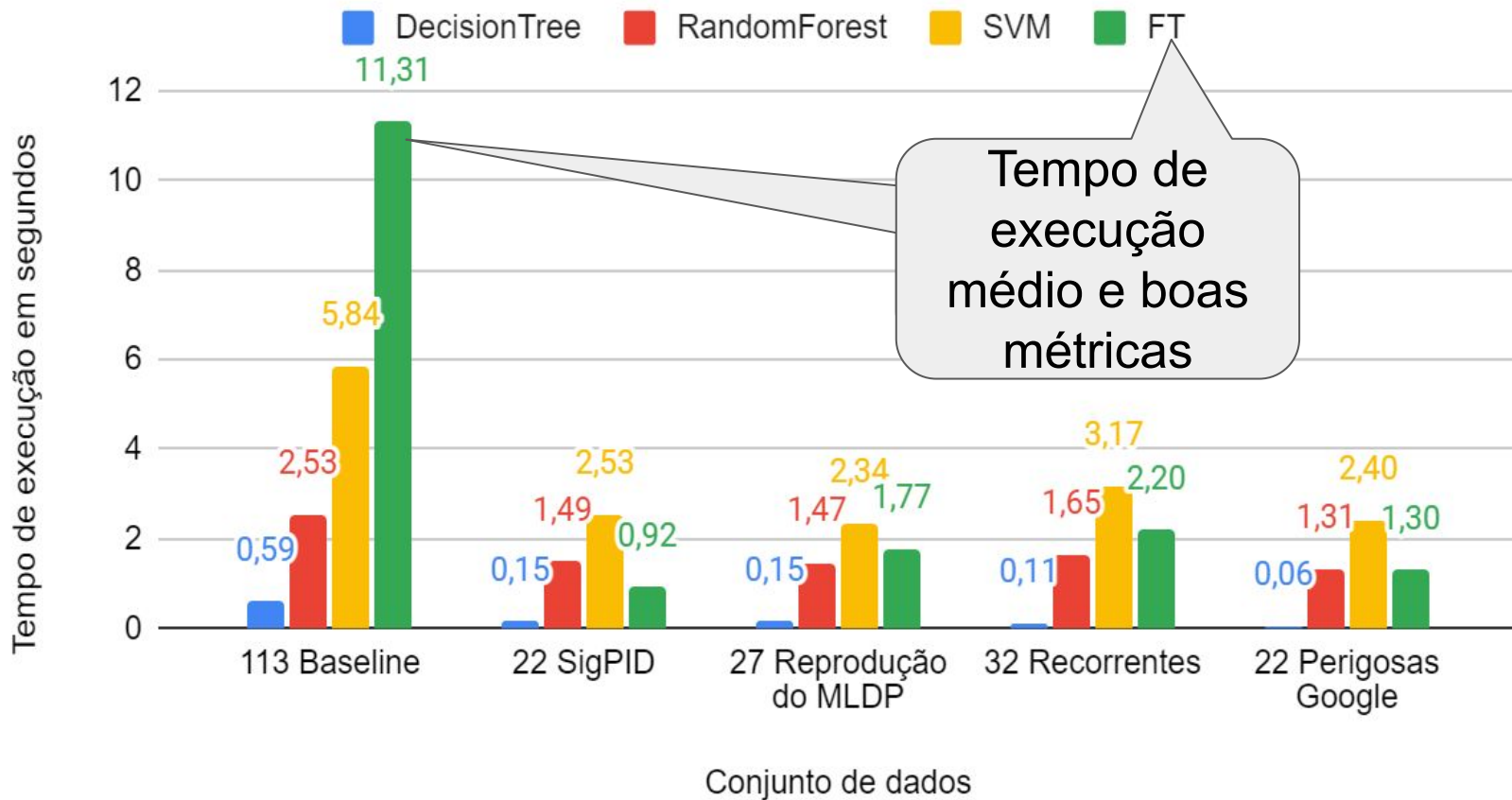
Tempo de execução



Tempo de execução



Tempo de execução



Conclusão



- Tempo de execução é mais relevante?
- Quantidade de dados impacta o tempo de execução
- Redução de características mantém boa acurácia
- Permissões perigosas por si só não são significativas

Trabalhos Futuros



- Testes com conjuntos de dados maiores
- Testes com conjuntos de dados atuais
- Avaliação dos níveis de seleção para outras características
- Otimização dos modelos
- Testar os modelos em smartphones modernos
- Criar um dataset com dados atualizados

Obrigado!



UFAM