

# O impacto da seleção de permissões na detecção de malwares Android

Curso de Graduação em Engenharia de Software

**Discente: Joner Mello<sup>1</sup>**

**Orientador: Prof. Diego Kreutz<sup>1</sup>**

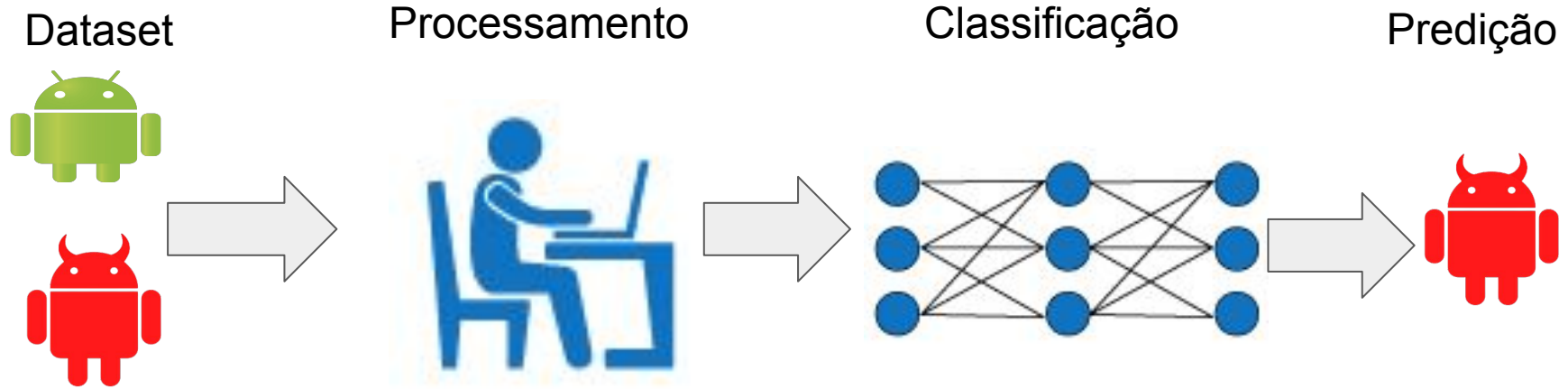
**Coorientador: Gustavo Cardozo<sup>1</sup>**

<sup>1</sup>Universidade Federal do Pampa  
Campus Alegrete

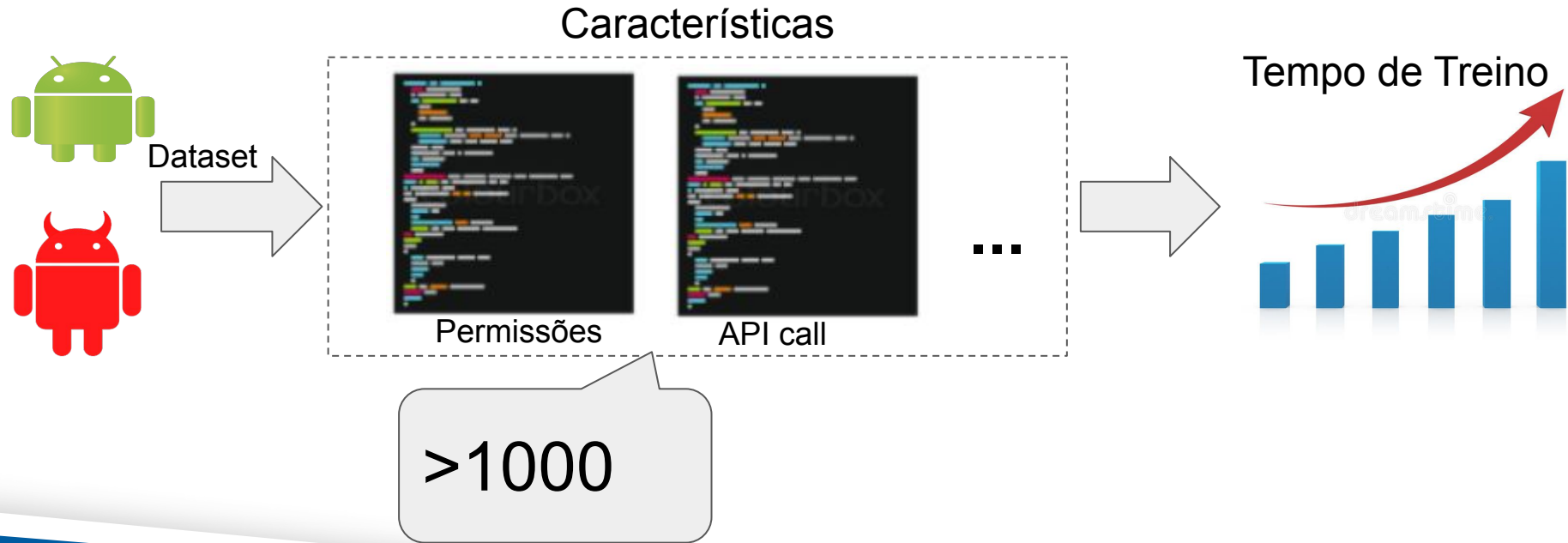
2021



# Detecção de malwares Android



# Detecção de malwares Android e o desafio da escalabilidade



Datasets com muitas características impactam o tempo de treinamento dos modelos



**DefenseDroid 1.490 features levou 11,3 segundos**  
**Drebin 215 features levou 3,2 segundos**



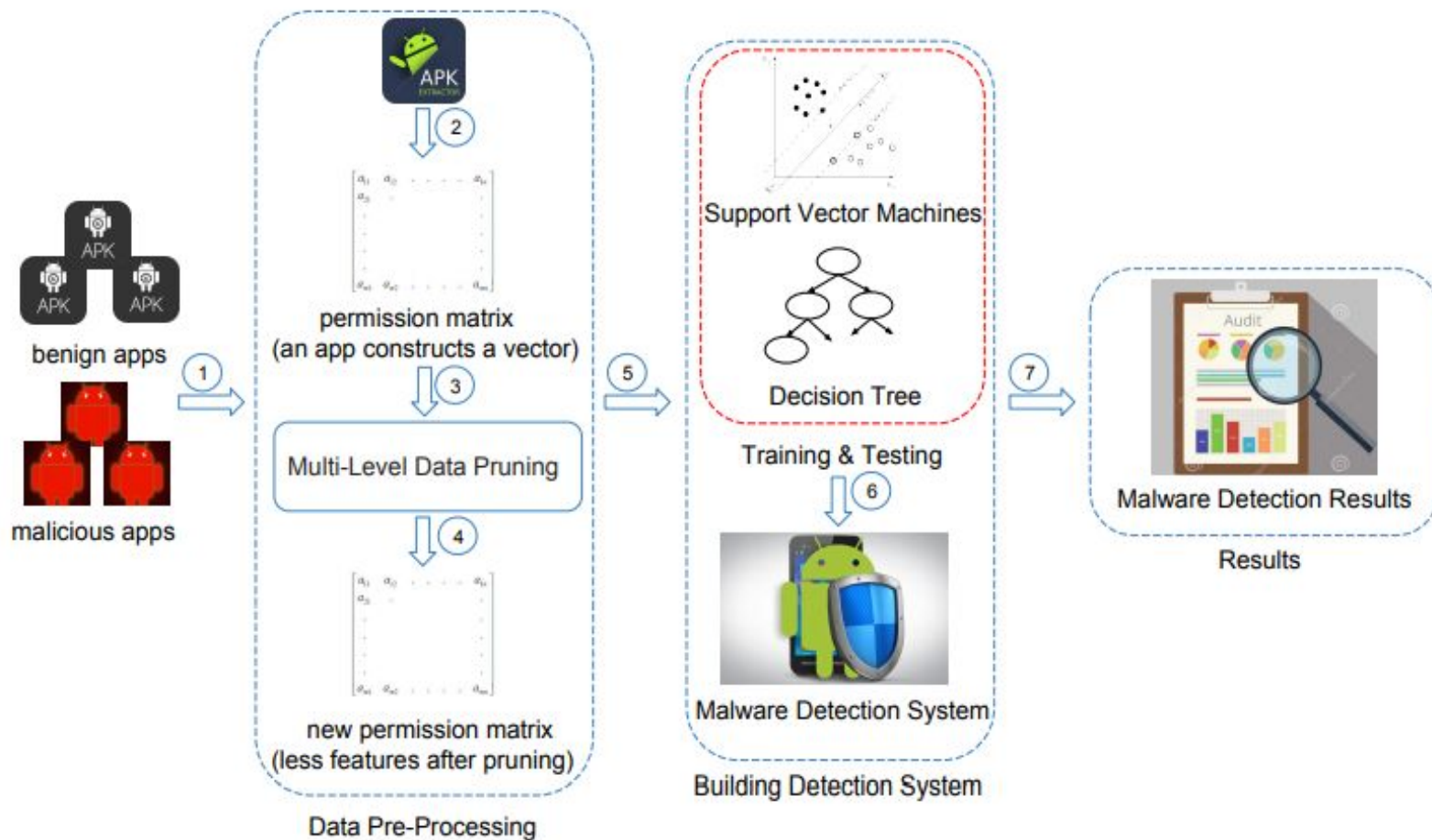


”

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



# Seleção de características SigPID



# Seleção de características SigPID

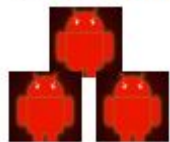


135

Permissões



benign apps



malicious apps



permission matrix  
(an app constructs a vector)



Multi-Level Data Pruning

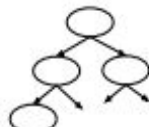


new permission matrix  
(less features after pruning)

Data Pre-Processing



Support Vector Machines



Decision Tree

Training & Testing



Malware Detection System

Building Detection System

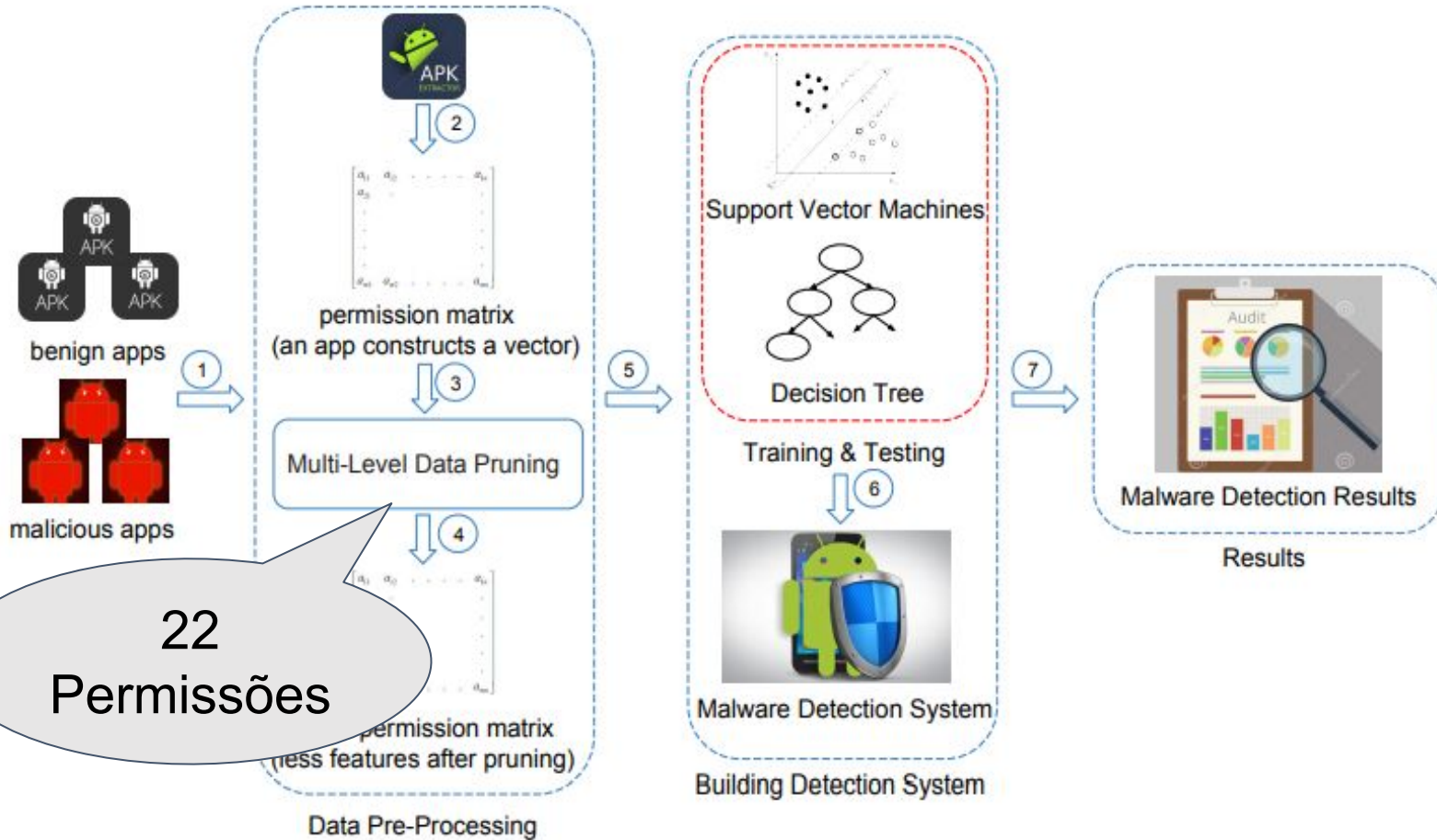


Malware Detection Results

Results

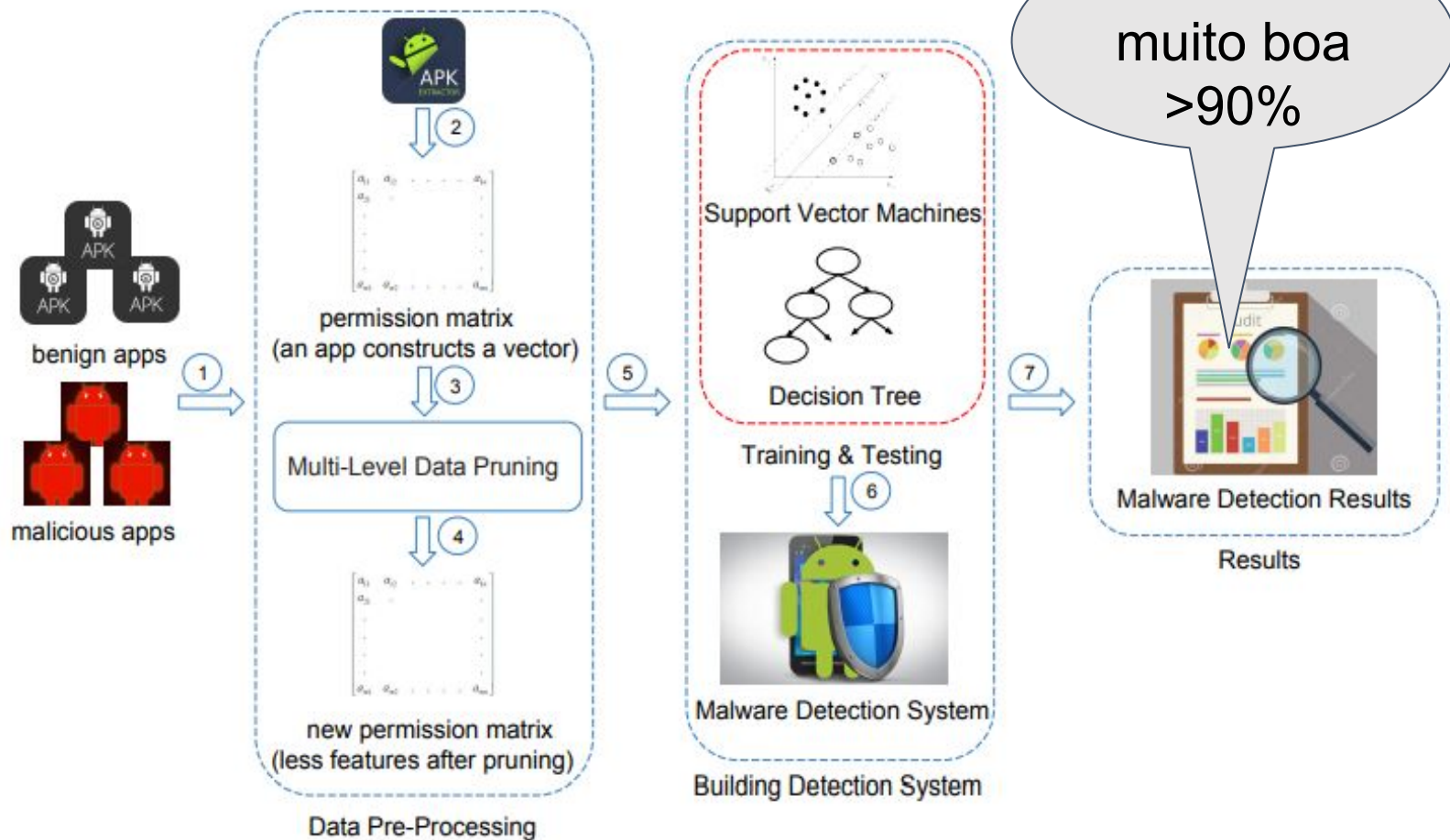


# Seleção de características SigPID





# Seleção de características SigPID



# Objetivos

- **Replicar o SigPID com dataset conhecido**
- **Comparar com o trabalho original(SigPID)**
- **Comparar com outros conjuntos de permissões**
  - **Permissões recorrentes em malwares**
- **Criar uma API Web**



# Etapas

- **Etapa 1 levantamento das permissões recorrentes**
- **Etapa 2 Análise de reprodutibilidade do SigPID**
- **Etapa 3 Reprodução do SigPID utilizando um dataset público**
- **Etapa 4 Criação da API Web**



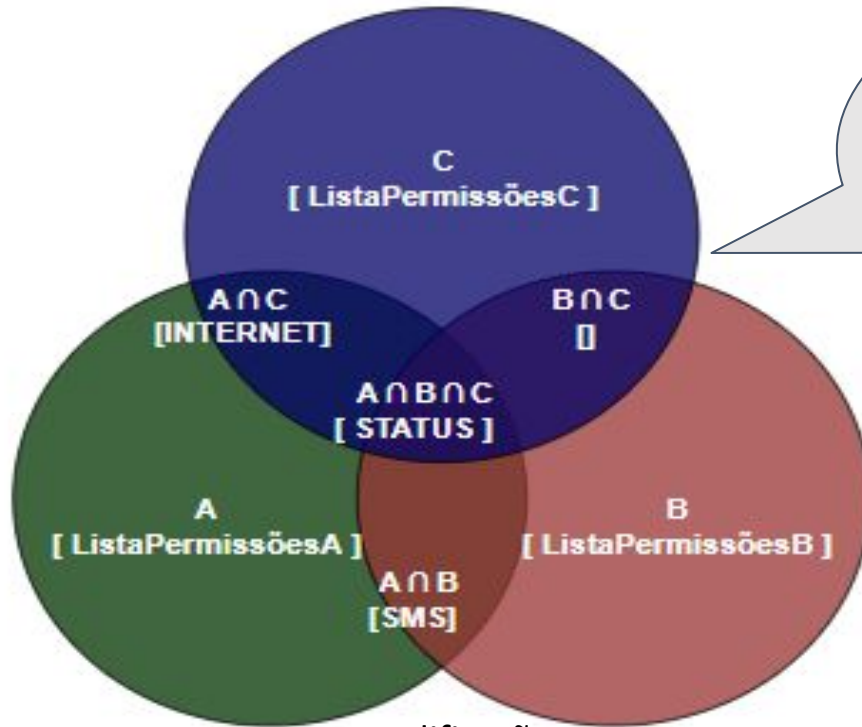


# Etapa 1

**Será que a abordagem de seleção de características do SigPID é a melhor alternativa?**



# Seleção de permissões recorrentes



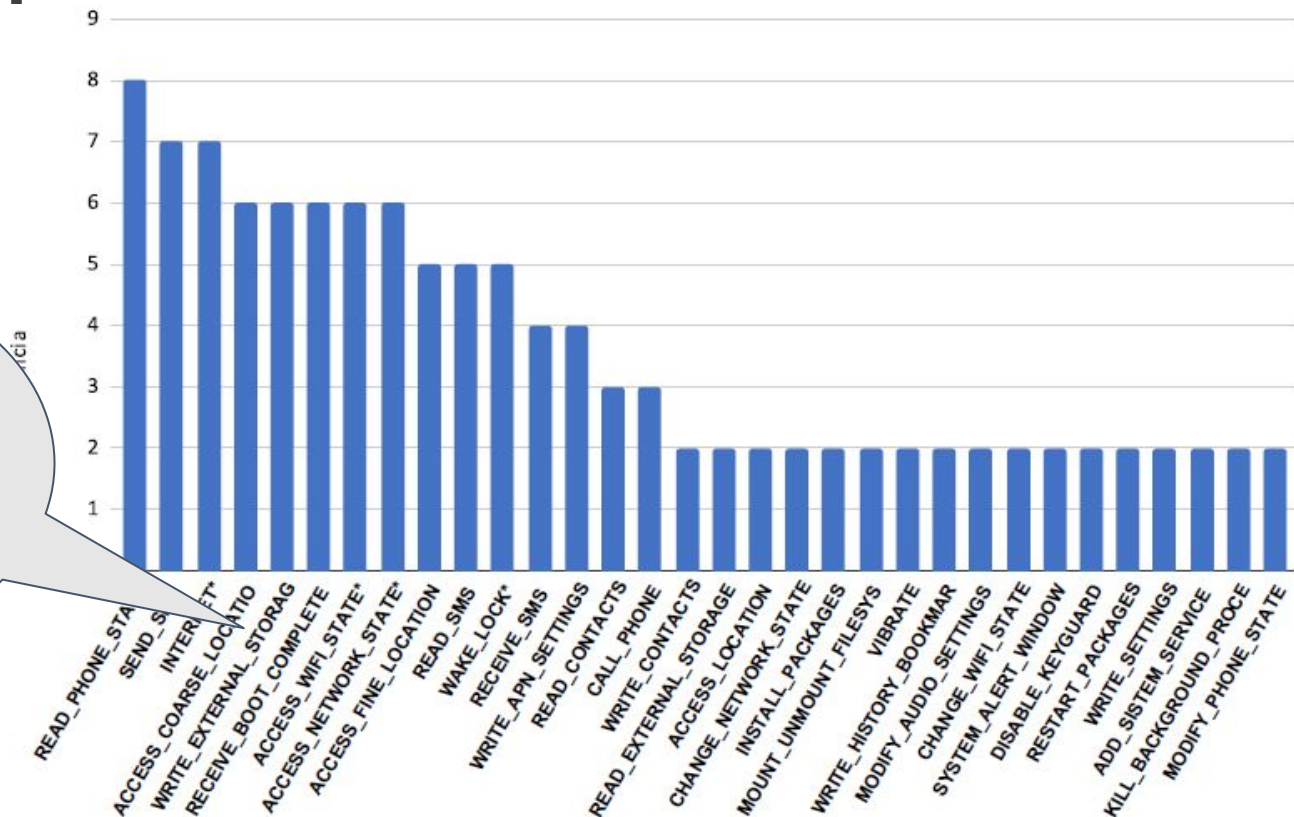
Intersecção de  
permissões

exemplificação



# Seleção de permissões recorrentes

32  
Permissões  
recorrentes





# Seleção de permissões recorrentes

90,6% (29 de 32)

API-1

**READ\_PHONE\_STATE**

Ler as informações atuais da rede celular.

API-1

**SEND\_SMS**

Permite que um aplicativo envie mensagens SMS.

API-1

**INTERNET**

Permite que os aplicativos abram sockets de rede.

API-1

**ACCESS\_COARSE\_LOCATION**

Permite que um aplicativo acesse a localização aproximada.



# Seleção de permissões recorrentes

API-1

**READ\_PHONE\_STATE**

Ler as informações atuais da rede celular.

API-1

**SEND\_SMS**

Permite que um aplicativo envie mensagens SMS.

API-1

**INTERNET**

Permite que os aplicativos estabeleçam uma conexão sockets

API-1

**ACCESS\_COARSE\_LOCATION**

Permite que um aplicativo acesse a localização aproximada.

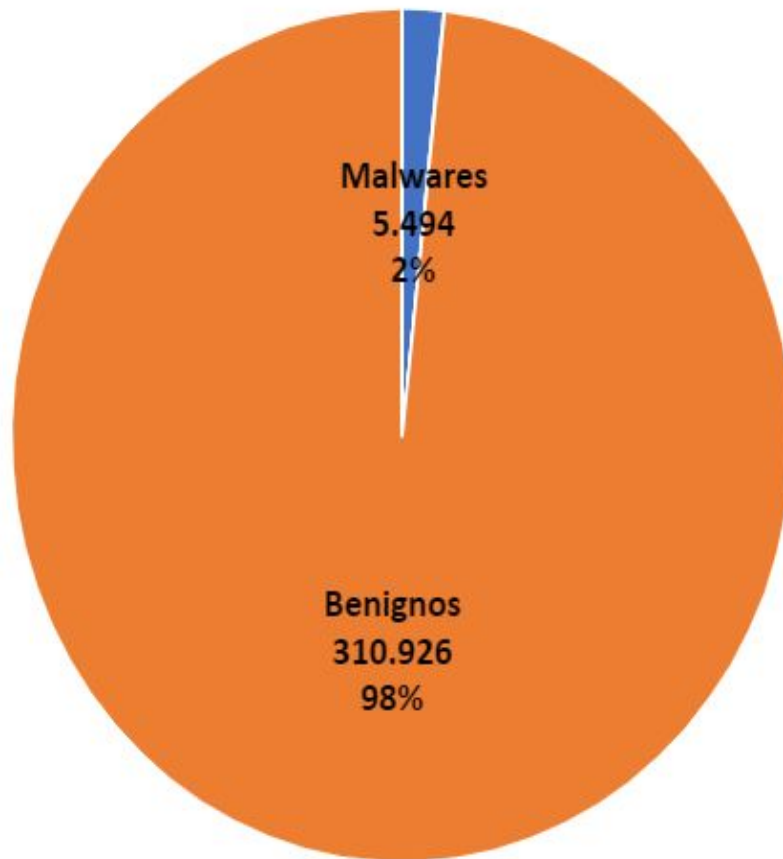




# Etapa 2: Análise da reprodutibilidade

# 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) ✓\*





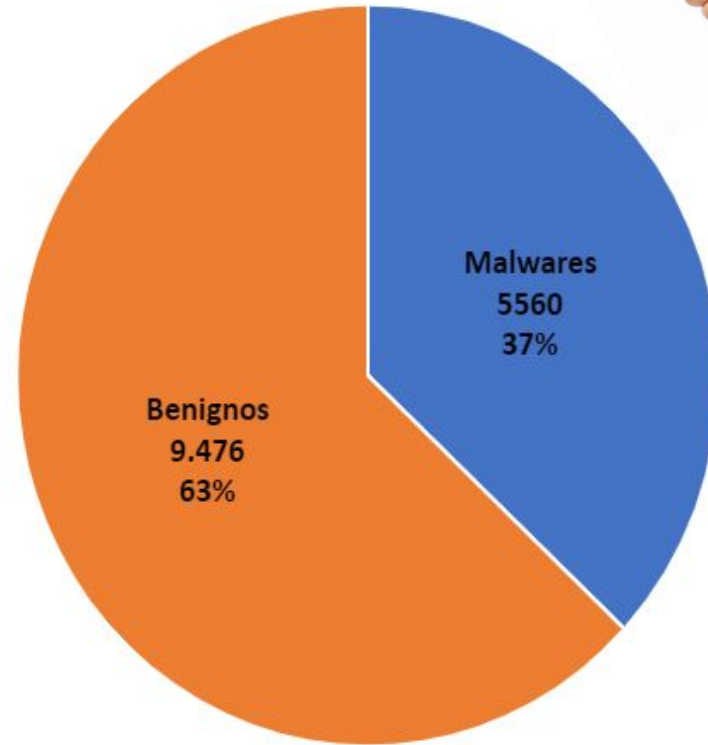
# Etapa 3: Reprodução do SigPID

# Dataset público escolhido (Drebin\_215)



- **Amostra**

- 5.560 Malwares
- 9.476 Benignos



[drebin-215-dataset-5560malware-9476-benign.csv](#)





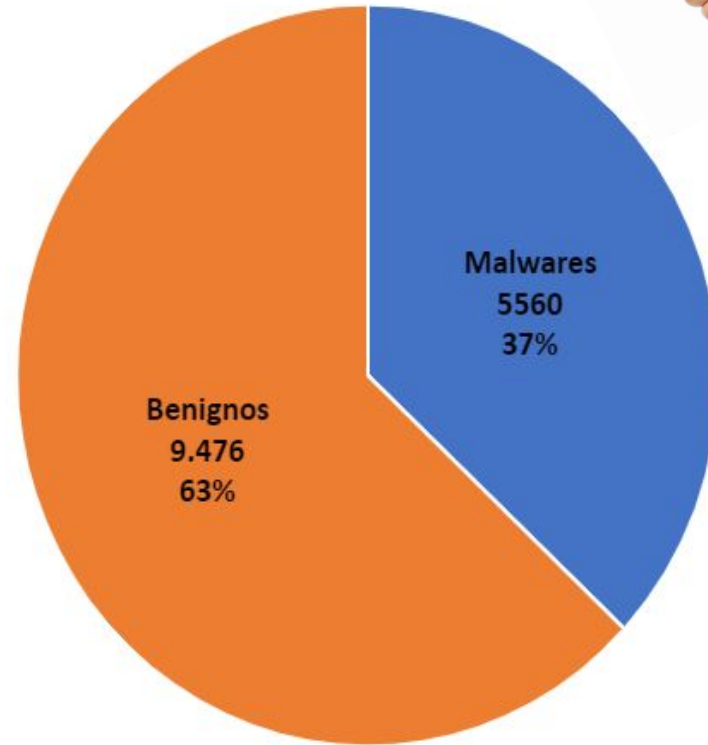
# Dataset público escolhido (Drebin\_215)



- **Amostra**

- 5.560 Malwares
- 9.476 Benignos

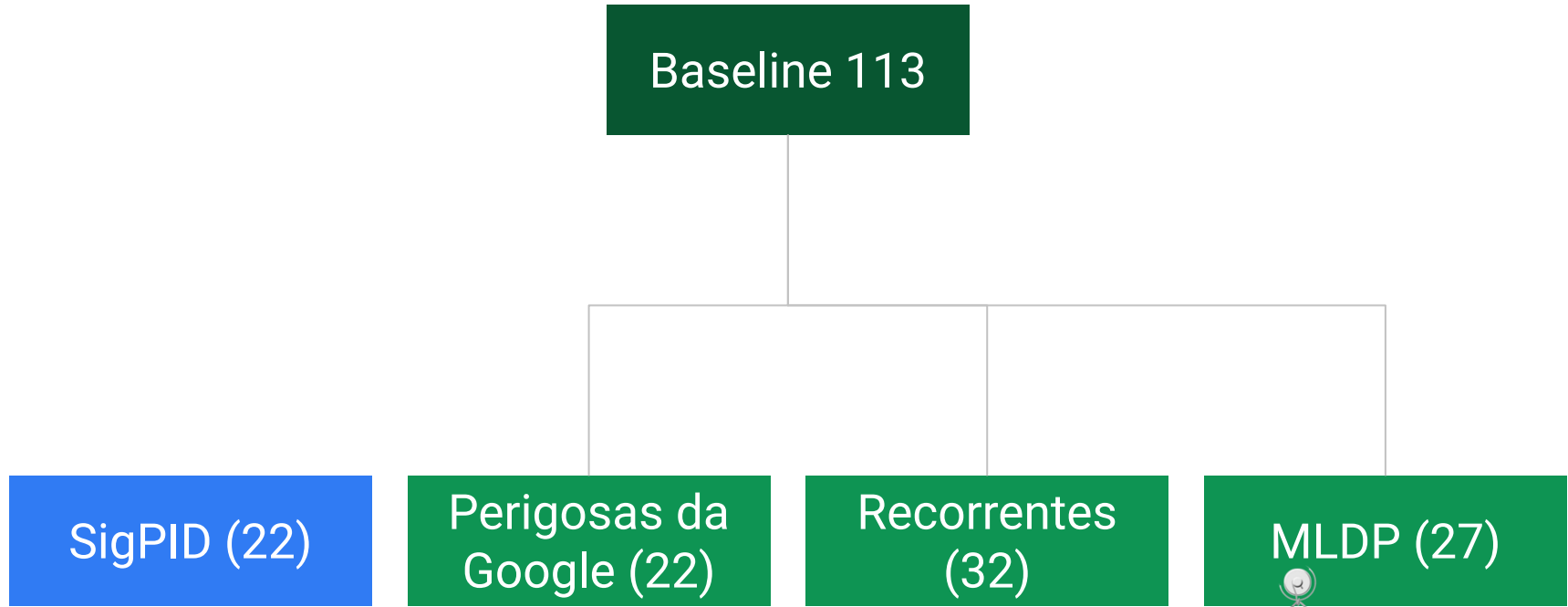
- **113 permissões**



[drebin-215-dataset-5560malware-9476-benign.csv](#)



# Datasets





# | Seleção de dados multinível (MLDP)

# Os 3 níveis de seleção do MLDP

PRNR

SPR

PMAR



# PRNR

Classificação de  
permissão  
baseada em  
suporte (Support  
Based Permission  
Ranking)



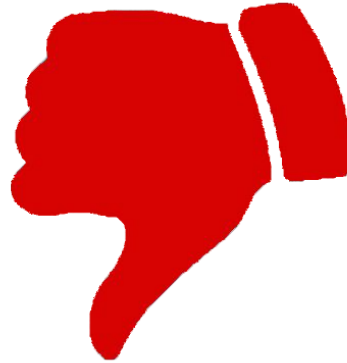
# Classificação de permissão com taxa negativa (PRNR)

Ranking = 1



Alto Risco

Ranking = -1



Baixo risco

Ranking = 0



Irrelevante



# Classificação de permissão com taxa negativa (PRNR)



Ordem  
crescente

| Permission    | R    |
|---------------|------|
| READ_CALENDAR | -1   |
| VIBRATE       | -1   |
| RECORD_AUDIO  | -0.3 |
| CAMERA        | 0    |
| INTERNET      | 0    |
| READ_SMS      | 1    |
| READ_LOGS     | 1    |
| WRITE_SMS     | 1    |

| Permission    | R    |
|---------------|------|
| WRITE_SMS     | 1    |
| READ_LOGS     | 1    |
| READ_SMS      | 1    |
| INTERNET      | 0    |
| CAMERA        | 0    |
| RECORD_AUDIO  | -0.3 |
| VIBRATE       | -1   |
| READ_CALENDAR | -1   |



Ordem  
decrescente





# Sistema incremental de permissão(PIS)

| Permission    | R    | Permission | R |
|---------------|------|------------|---|
| READ_CALENDAR | -1   | WRITE_SMS  | 1 |
| VIBRATE       | -1   | READ_LOGS  | 1 |
| RECORD_AUDIO  | -0.3 | READ_SMS   | 1 |

## SVM

Acurácia

Precisão

Recall

F1\_Score



# Sistema incremental de permissão(PIS)

| Permission    | R    | Permission   | R    |
|---------------|------|--------------|------|
| READ_CALENDAR | -1   | WRITE_SMS    | 1    |
| VIBRATE       | -1   | READ_LOGS    | 1    |
| RECORD_AUDIO  | -0.3 | READ_SMS     | 1    |
| CAMERA        | 0    | INTERNET     | 0    |
| INTERNET      | 0    | CAMERA       | 0    |
| READ_SMS      | 1    | RECORD_AUDIO | -0.3 |

## SVM

Acurácia

Precisão

Recall

F1\_Score



# Sistema incremental de permissão(PIS)

| Permission    | R    | Permission    | R    |
|---------------|------|---------------|------|
| READ_CALENDAR | -1   | WRITE_SMS     | 1    |
| VIBRATE       | -1   | READ_LOGS     | 1    |
| RECORD_AUDIO  | -0.3 | READ_SMS      | 1    |
| CAMERA        | 0    | INTERNET      | 0    |
| INTERNET      | 0    | CAMERA        | 0    |
| READ_SMS      | 1    | RECORD_AUDIO  | -0.3 |
| READ_LOGS     | 1    | VIBRATE       | -1   |
| WRITE_SMS     | 1    | READ_CALENDAR | -1   |

## SVM

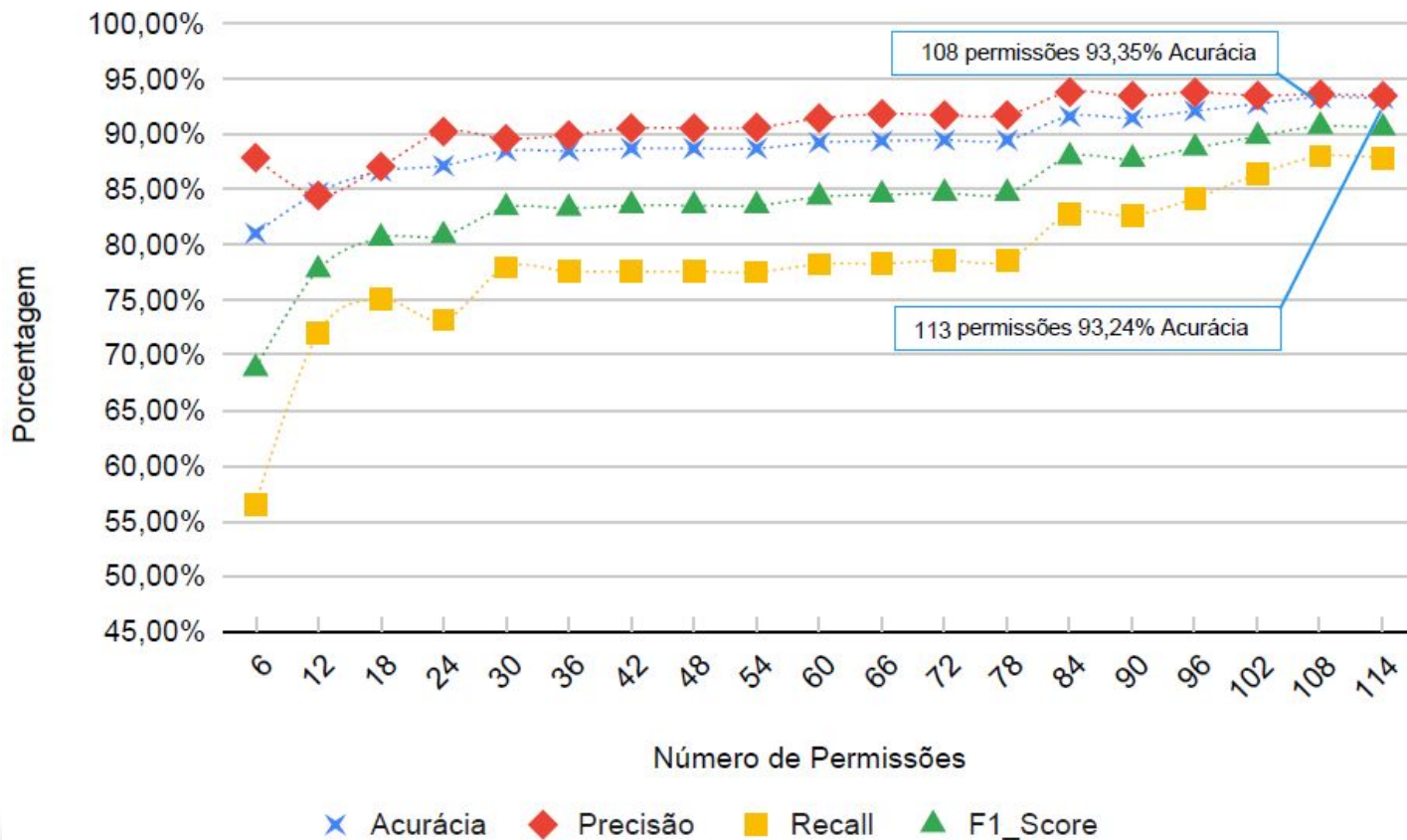
Acurácia

Precisão

Recall

F1\_Score



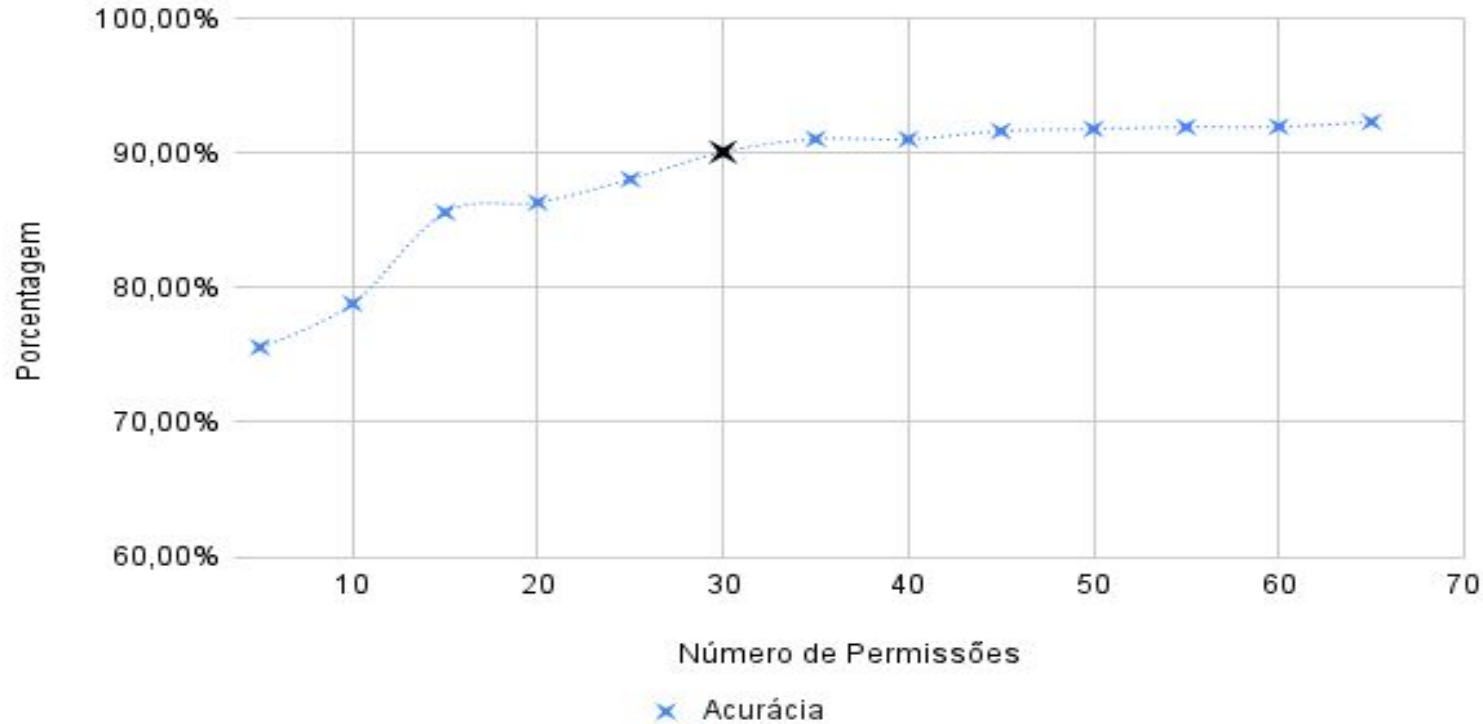


# SPR

Classificação de  
permissão  
baseada em  
suporte (Support  
Based Permission  
Ranking)



# Sistema incremental de permissão(PIS)



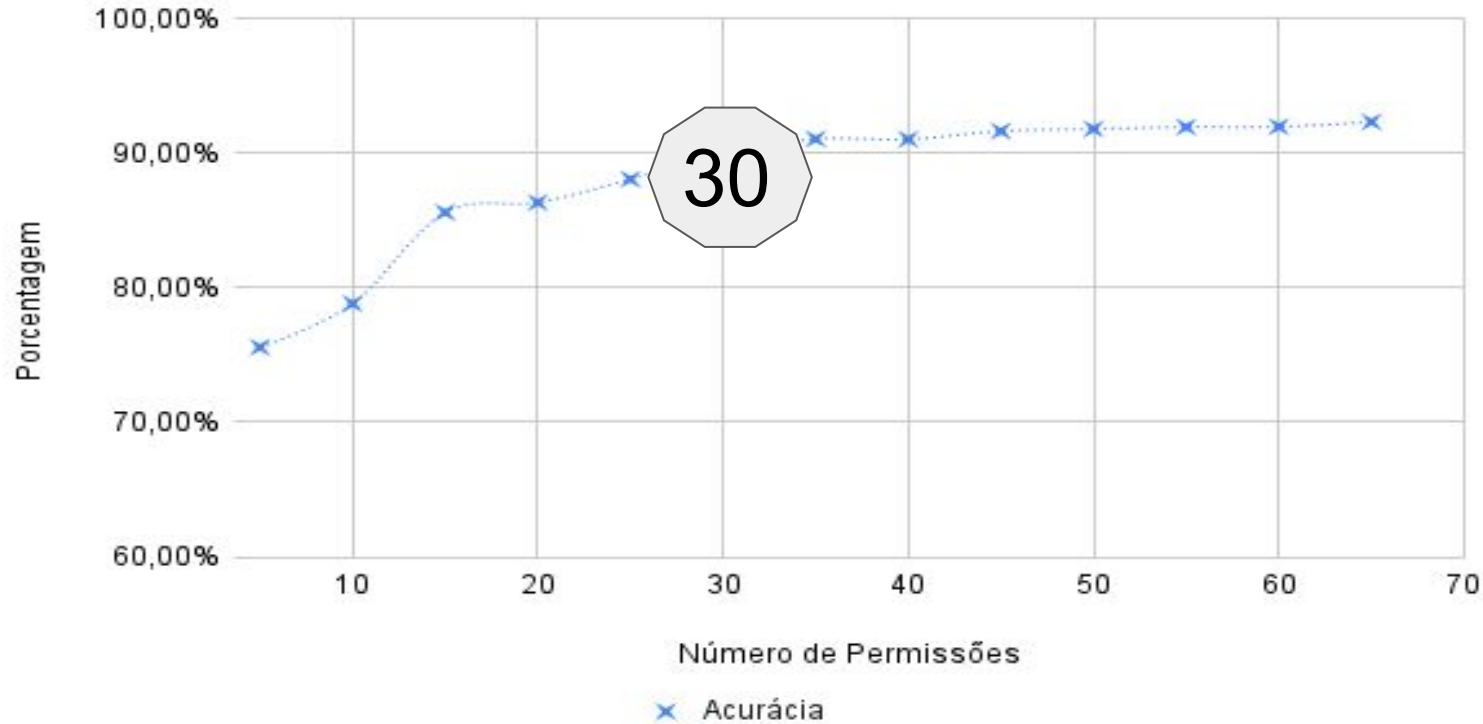
**SVM**

Acurácia

>90%



# Sistema incremental de permissão(PIS)



**SVM**

Acurácia

>90%





# PMAR

Mineração de  
permissões com  
regras de associação  
(Permission Mining  
with Association  
Rules)



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

| Antecedentes      | Consequentes      | Suporte  | Confiança | Lift |
|-------------------|-------------------|----------|-----------|------|
| CHANGE_WIFI_STATE | ACCESS_WIFI_STATE | 0.160758 | 0.993016  | 2.28 |
| MANAGE_ACCOUNTS   | GET_ACCOUNTS      | 0.103359 | 0.992971  | 3.32 |
| WRITE_SMS         | READ_SMS          | 0.111407 | 0.984136  | 5.26 |





# Treino e Teste



**matplotlib**

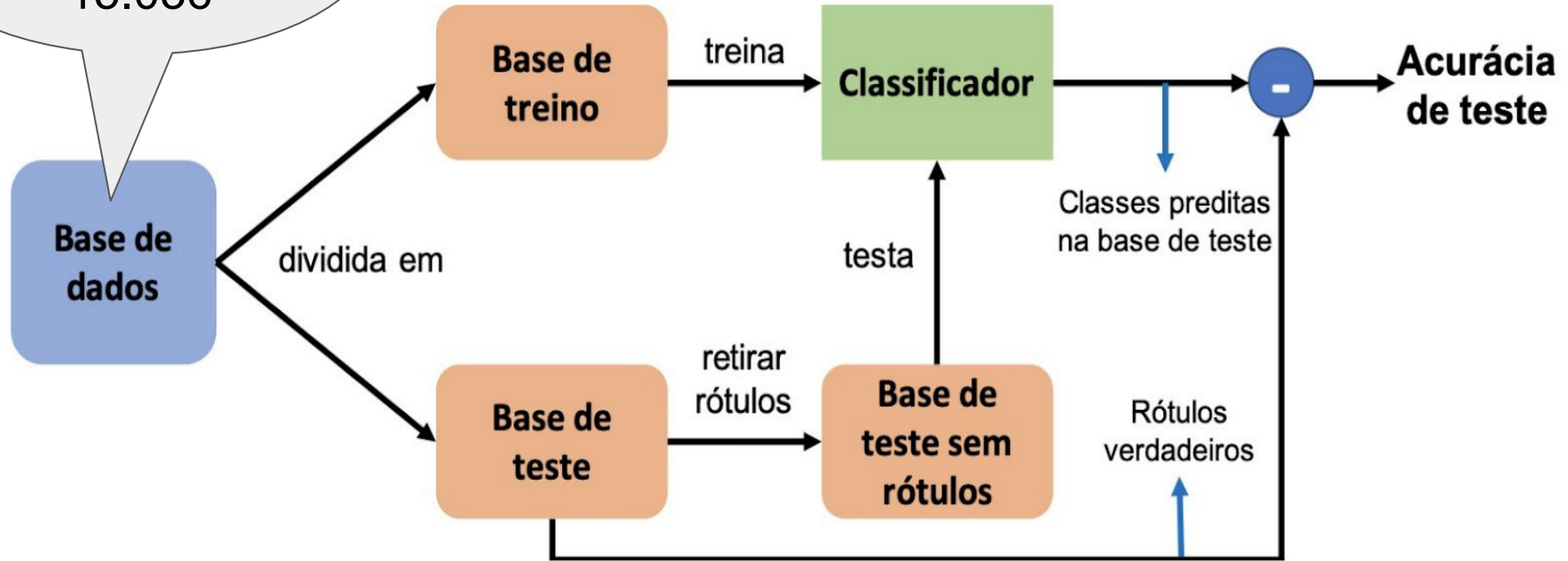


**NumPy**



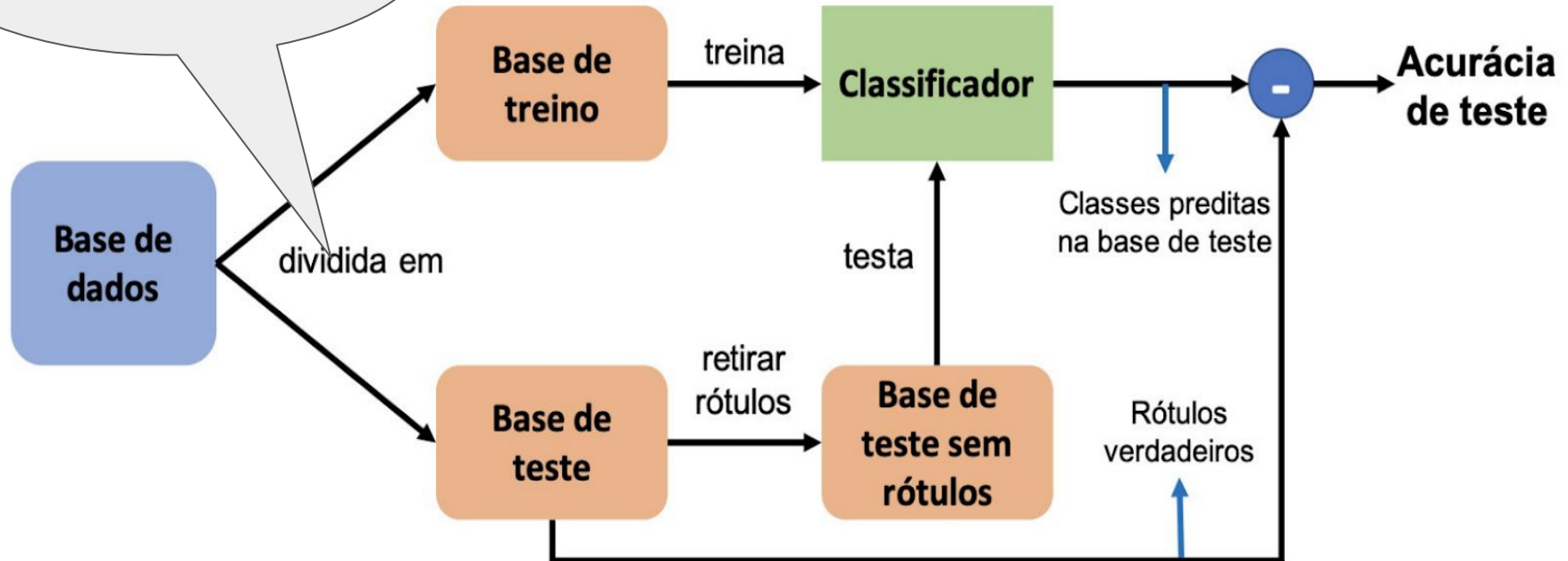
# Treino e teste dos conjuntos de dados

Drebin\_215  
Tamanho  
15.036



# Treino e teste dos conjuntos de dados

70% - Treino  
30% - Teste





# Resultados

# SVM

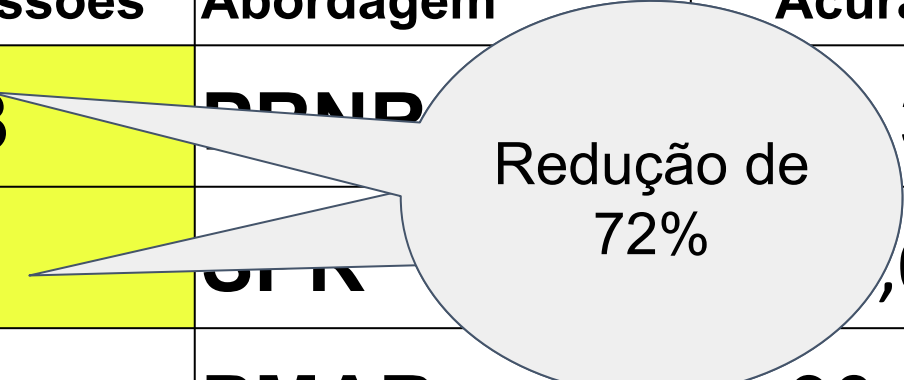
| Nº de Permissões | Abordagem   | Acurácia | Tempo Execução |
|------------------|-------------|----------|----------------|
| 108              | <b>PRNR</b> | 93,35    | 5,44           |
| 30               | <b>SPR</b>  | 90,07    | 2,41           |
| 27               | <b>PMAR</b> | 90,05    | 2,26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





# 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 'DDND' row, and the second arrow points to the value '30' in the 'OTR' row, highlighting the reduction from 108 to 30 permissions.



# SVM

| Nº de Permissões | Ordagem     | Acurácia | Tempo Execução |
|------------------|-------------|----------|----------------|
| 108              | <b>PMAR</b> | 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              | <b>PRNF</b> | 90,07    | 5,44s          |
| 30               | <b>SPR</b>  | 90,07    | 2,41s          |
| 27               | <b>PMAR</b> | 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



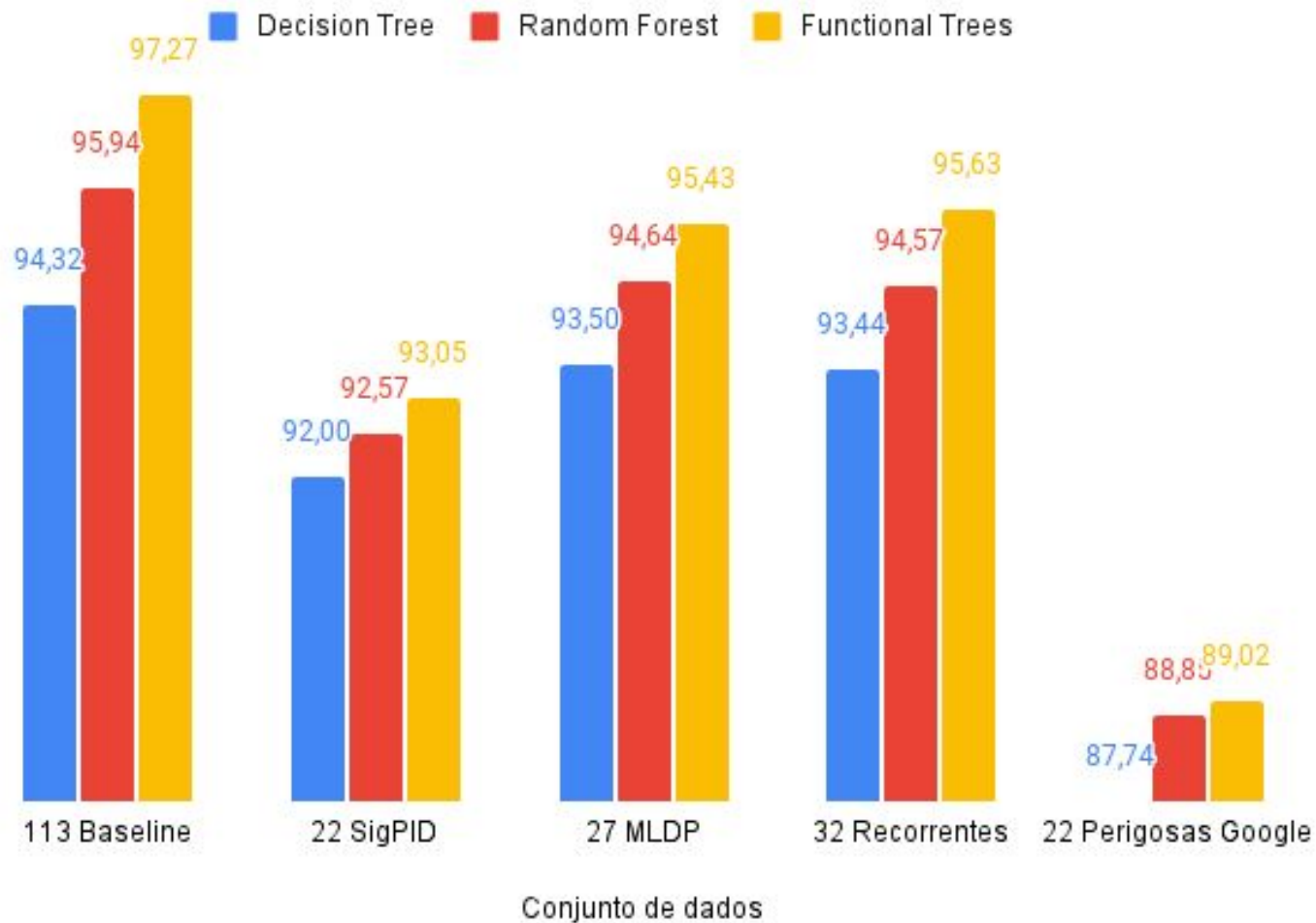
Decision  
Tree

Random  
Forest

Functional  
Trees

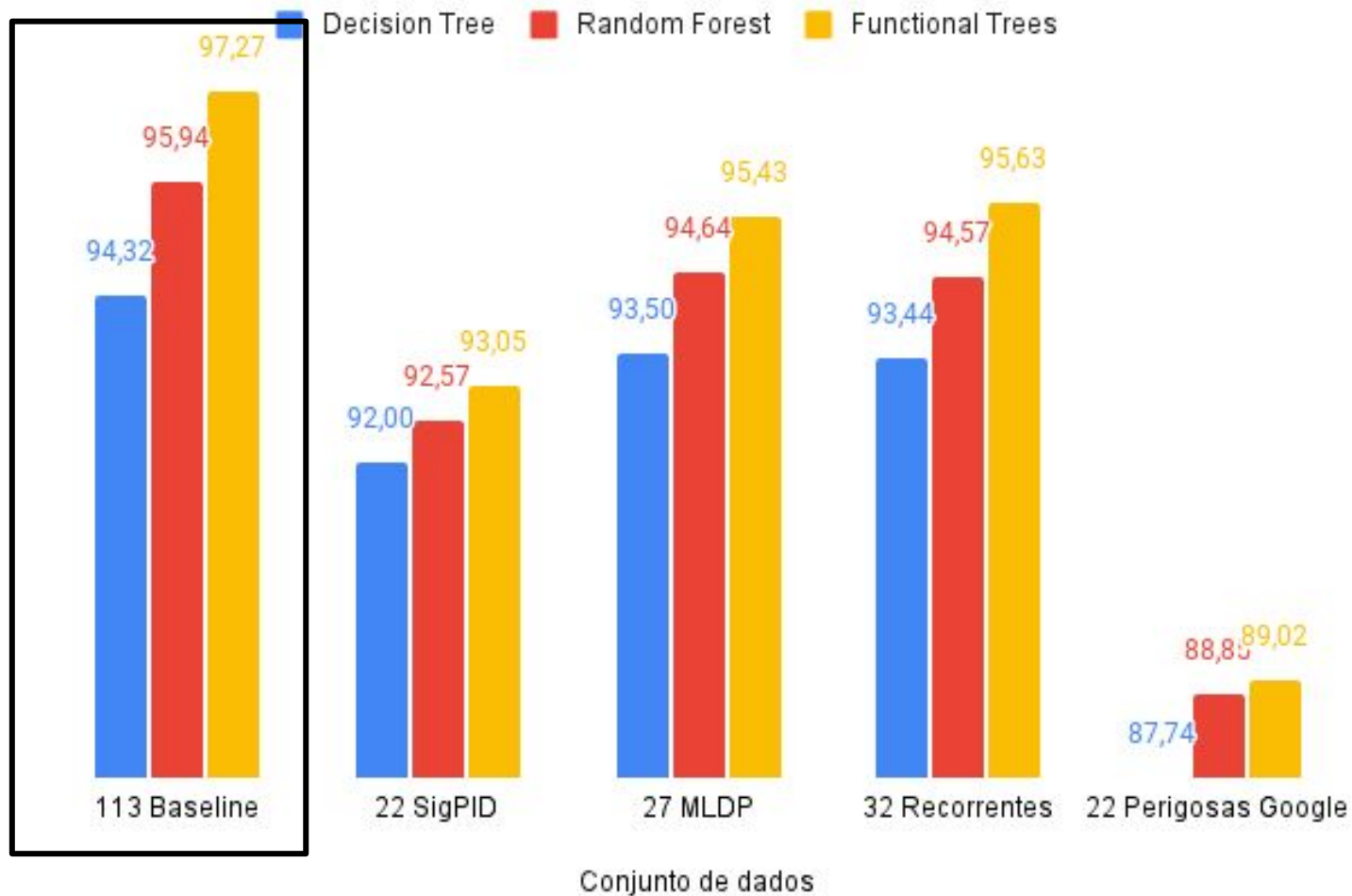


# Acurácia

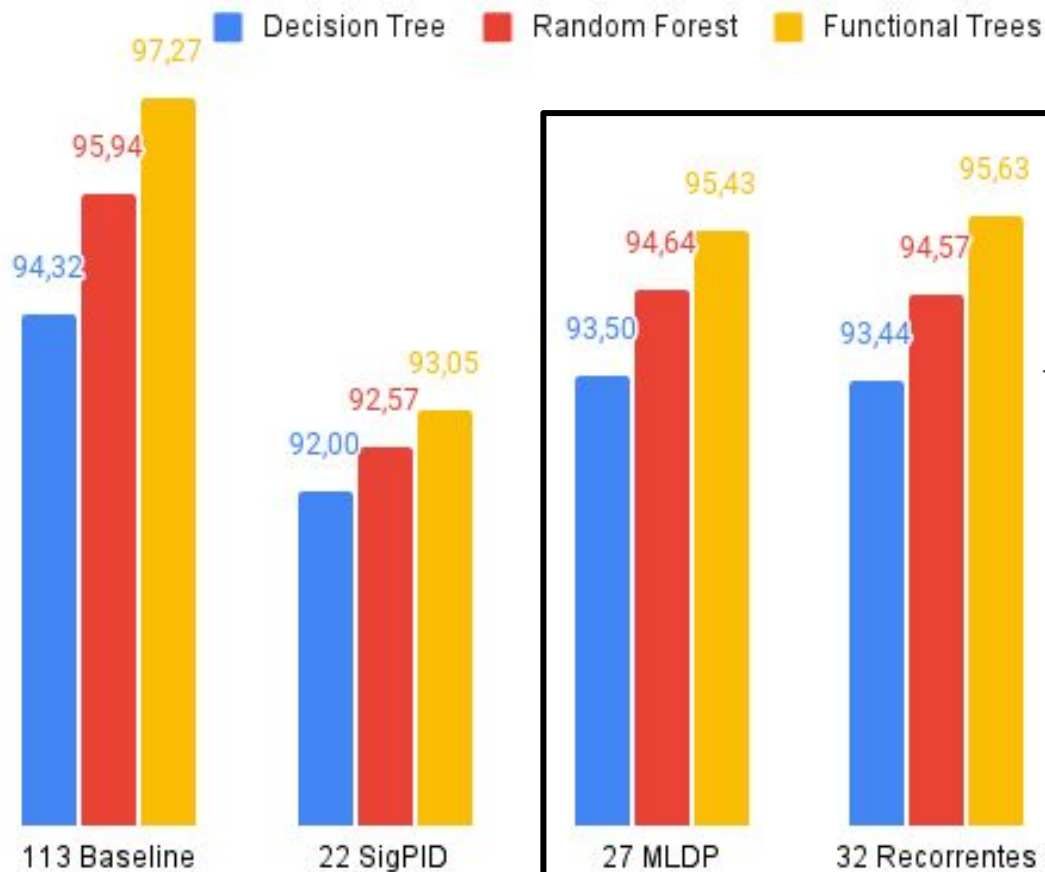




# Acurácia



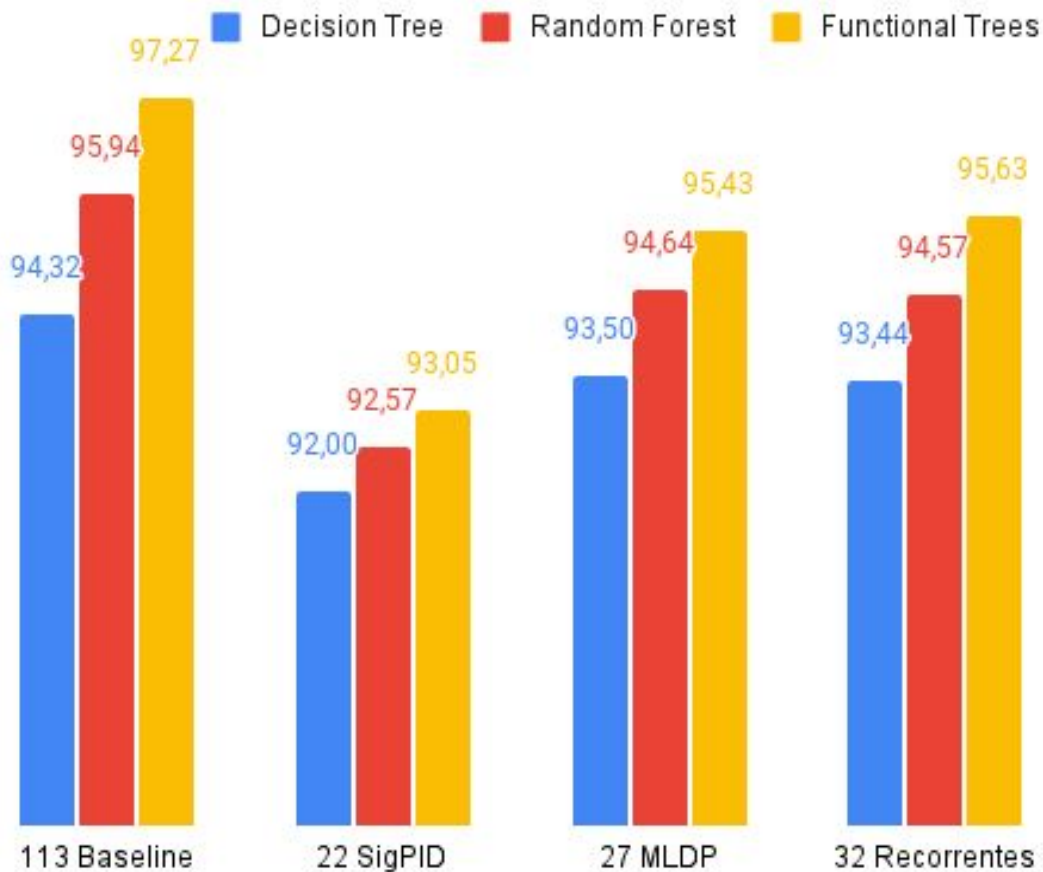
# Acurácia



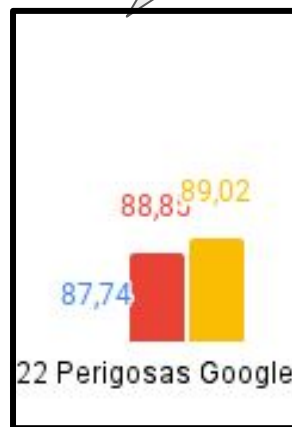
Equivalentes

Conjunto de dados

# Acurácia

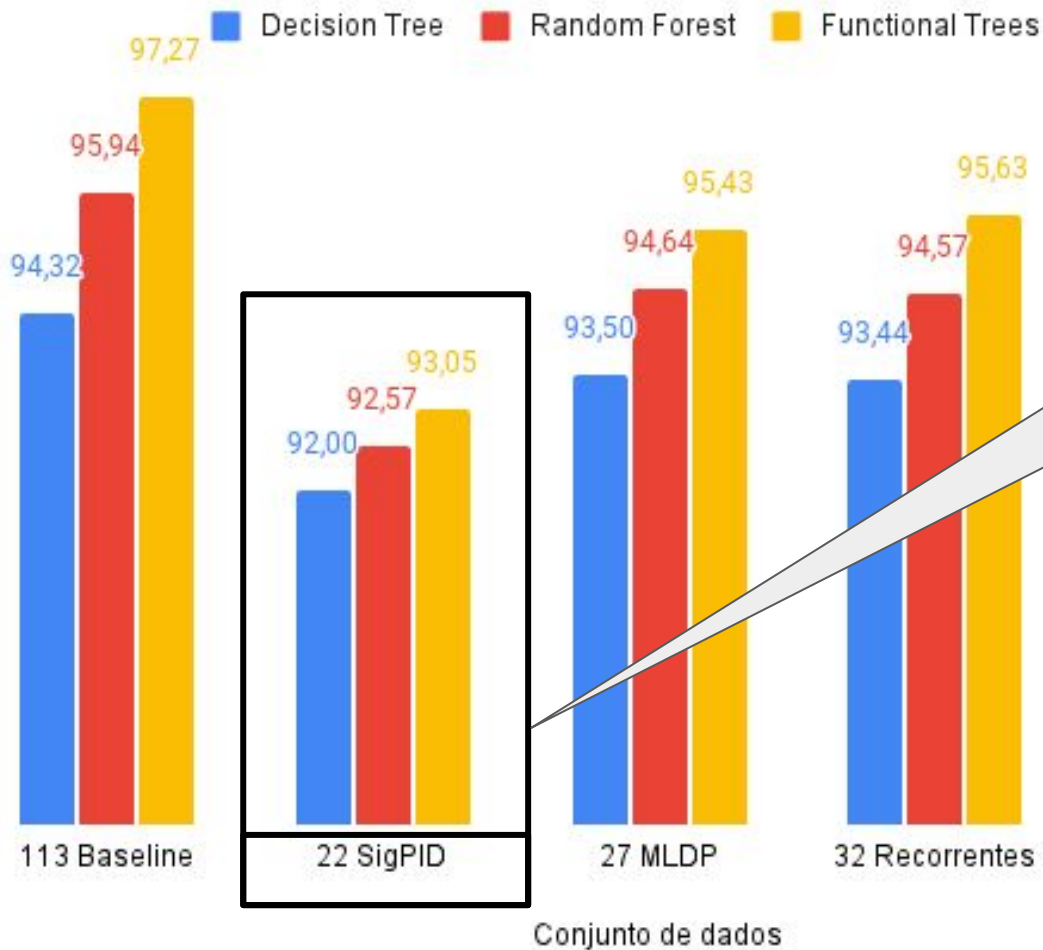


Pior desempenho



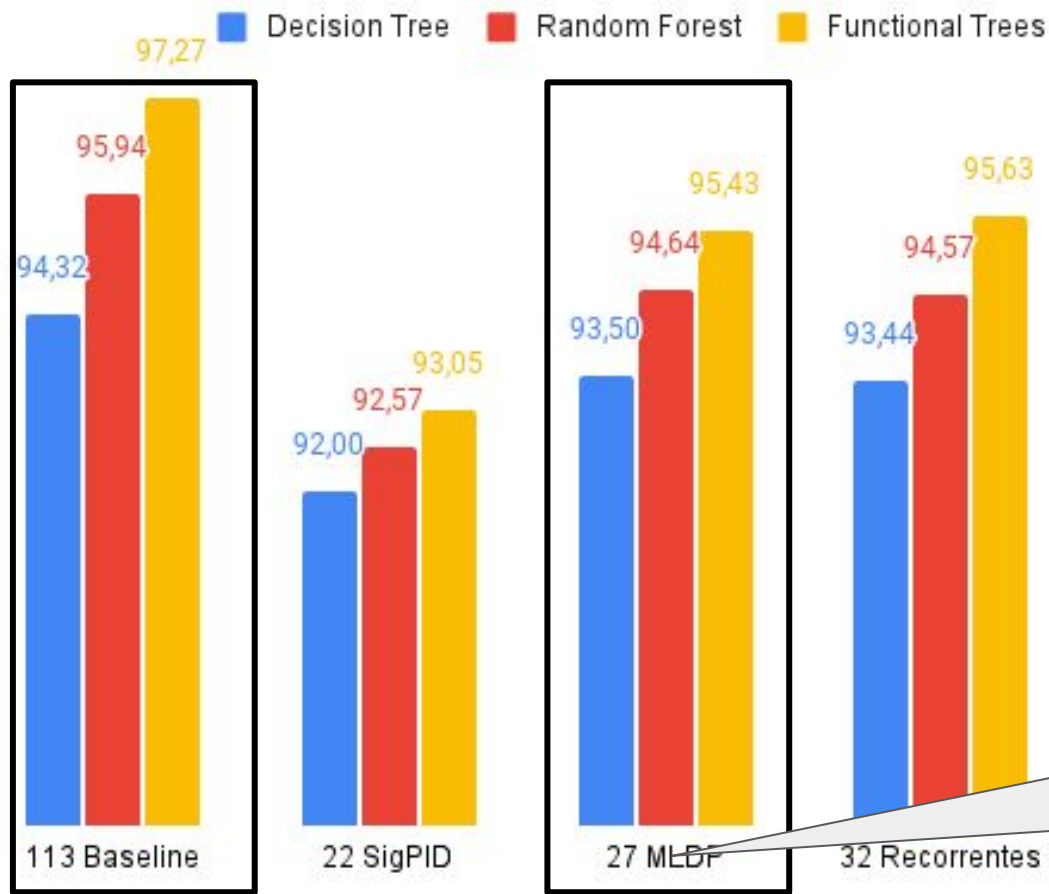
Conjunto de dados

# Acurácia



Mesmo N°  
de  
permissões  
porém  
diferentes

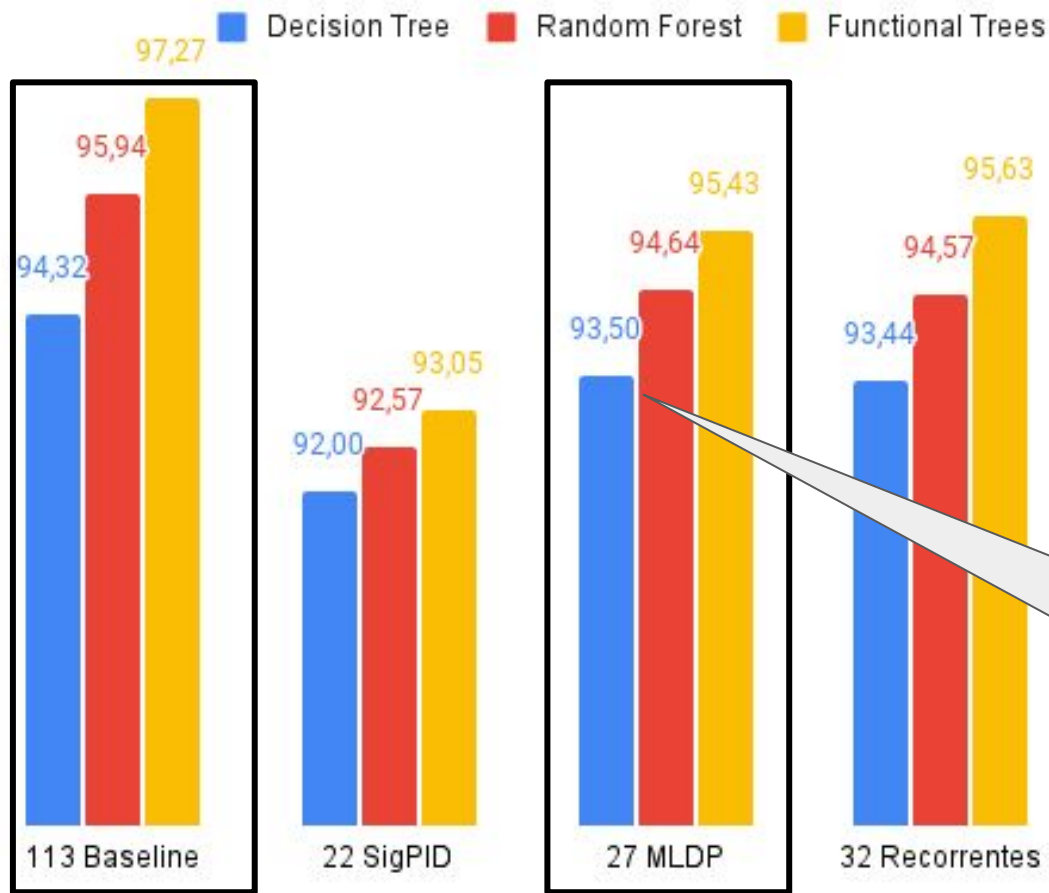
# Acurácia



Redução de  
76%  
(de 113 para 27)

Conjunto de dados

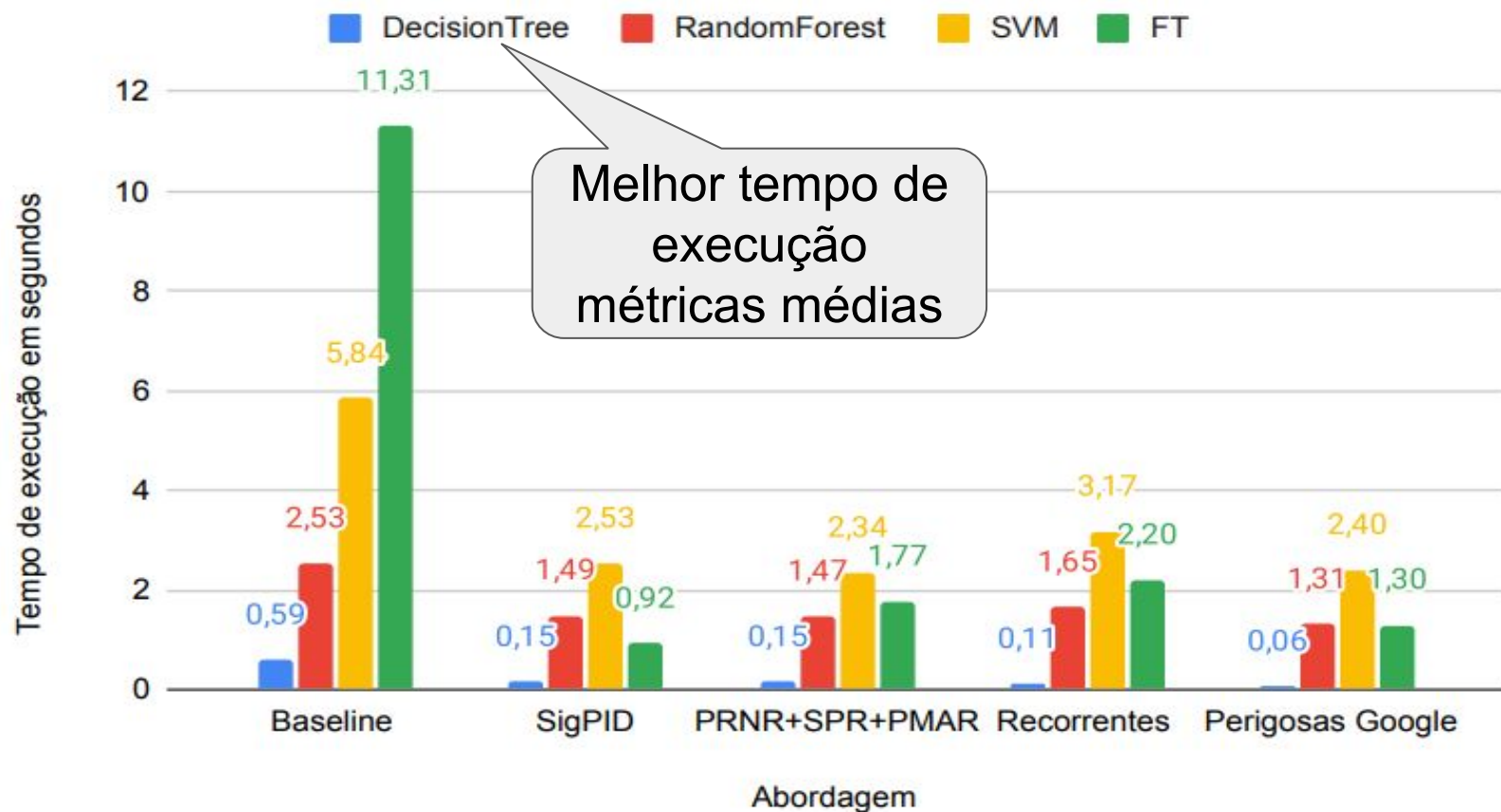
# Acurácia



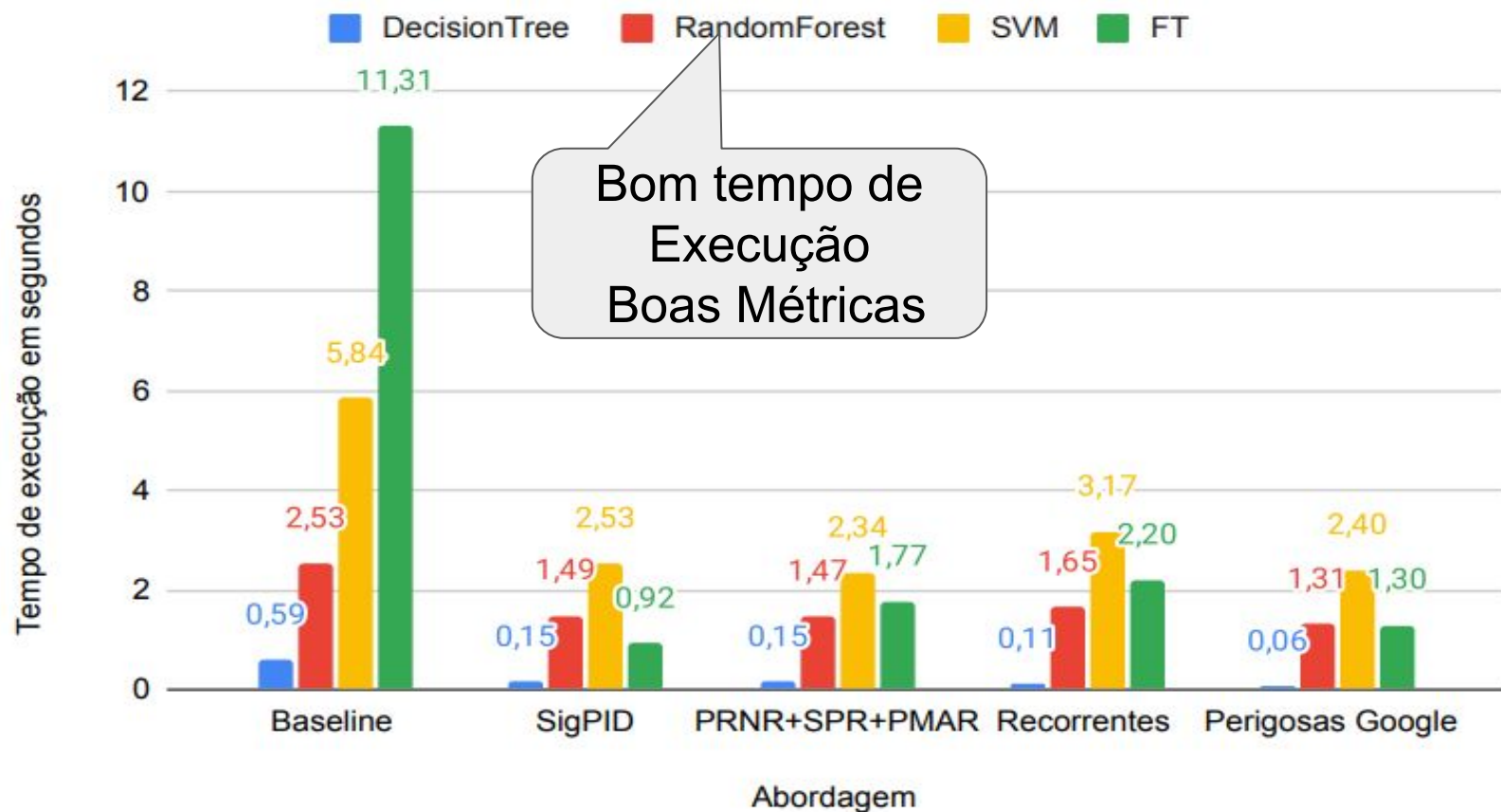
Boa Acurácia  
>90%

Conjunto de dados

# Tempo de execução

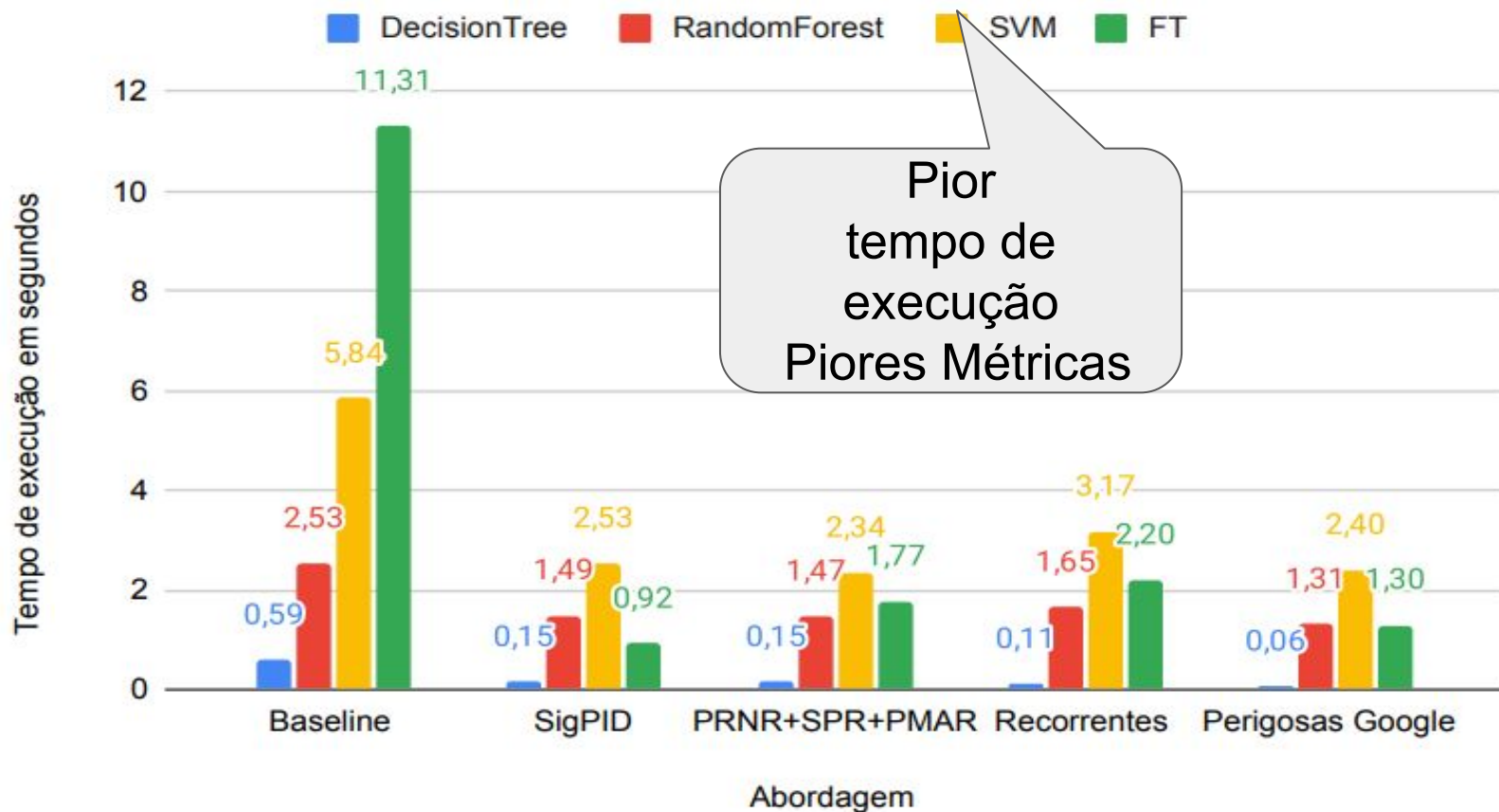


# Tempo de execução

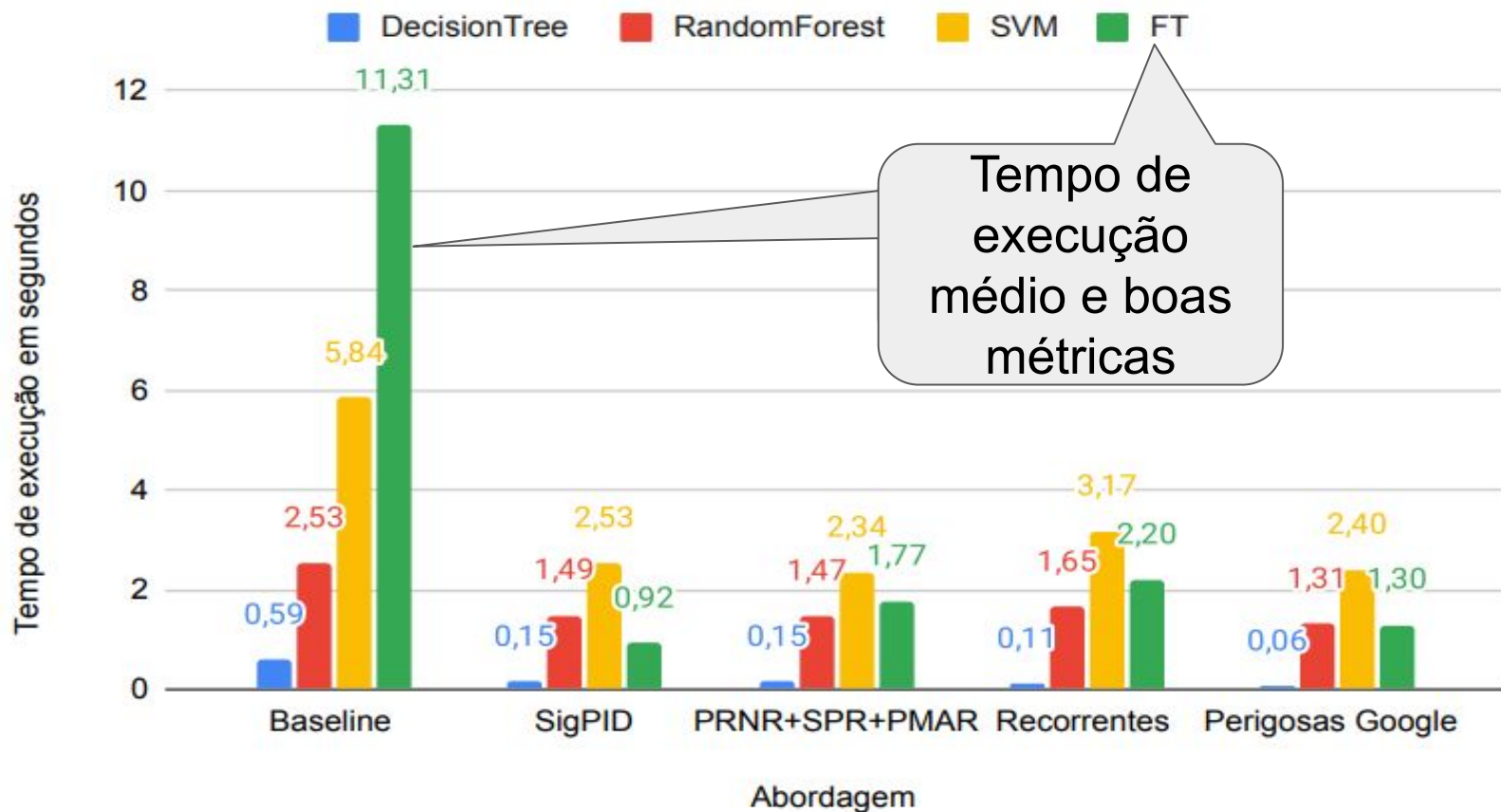





# Tempo de execução



# Tempo de execução





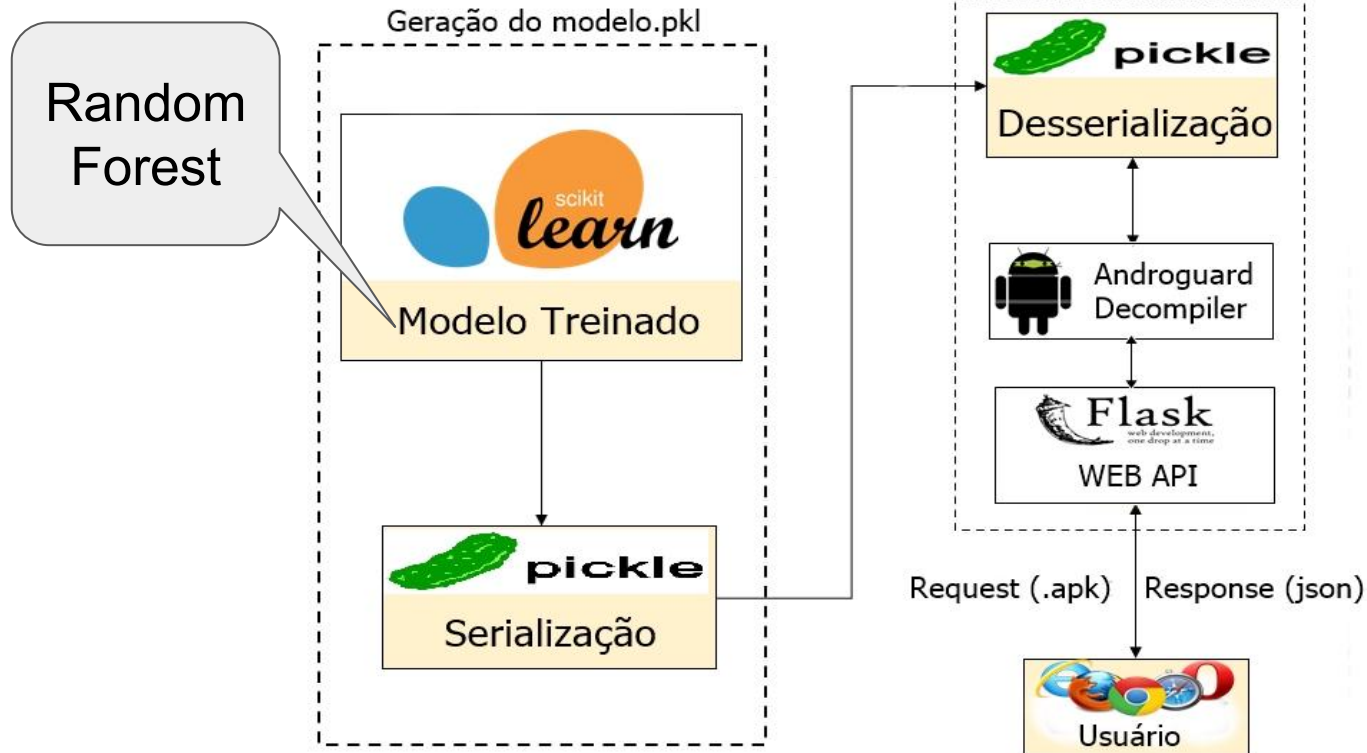
# Etapa 4: Criação da API Web

# Qual o propósito da ferramenta?



- **Criar uma interface simples para o usuário final**
- **Construir um repositório de aplicativos**
- **Alimentar o dataset para treinamento contínuo do modelo**

# API Web do modelo



# Interface Web

A screenshot of a web browser window. The address bar shows 'localhost:5000'. The browser's tab bar includes a tab titled 'JEMS: ERRC 2021: H...' and two bookmarked items: 'Outros favoritos' and 'Lista de leitura'. The main content area displays a web application titled 'Malware-Hunter'. This application features a text input field with the placeholder text 'Selecione um apk', a 'Browse' button to its right, and a large 'Analisar' button below the input field.

← → ↻ ⓘ localhost:5000 ☆ 🔴 ⚙️ 👤 ⋮

JEMS: ERRC 2021: H... » | 📁 Outros favoritos | 📖 Lista de leitura

### Malware-Hunter

# Interface Web



```
← → ↻ ⓘ localhost:5000/predict ☆ 🔴 ⚙️ 👤 ⋮
JEMS: ERRC 2021: H... >> | 🟡 Outros favoritos | 📖 Lista de leitura

[
  "Malware",
  "covid-19",
  "21",
  "9",
  [
    "android.permission.WRITE_SETTINGS",
    "android.permission.WRITE_SMS",
    "android.permission.CAMERA",
    "android.permission.CALL_PHONE",
    "android.permission.QUICKBOOT_POWERON",
    "android.permission.GET_ACCOUNTS",
    "android.permission.ACCESS_NETWORK_STATE",
    "android.permission.WRITE_EXTERNAL_STORAGE",
    "com.android.browser.permission.READ_HISTORY_BOOKMARKS",
    "android.permission.PROCESS_OUTGOING_CALLS",
    "android.permission.RECEIVE_SMS",
    "android.permission.READ_CONTACTS",
    "android.permission.RECEIVE_BOOT_COMPLETED",
    "android.permission.RECORD_AUDIO",
    "android.permission.READ_SMS",
    "android.permission.ACCESS_FINE_LOCATION",
    "android.permission.READ_PHONE_STATE",
    "android.permission.SEND_SMS",
    "android.permission.WAKE_LOCK",
    "android.permission.GET_TASKS",
    "android.permission.INTERNET"
  ]
]
```

# Avaliação da API Web



| Benignos |            | SigPID<br>API Web | Virus<br>Total | Qtd  |
|----------|------------|-------------------|----------------|------|
| 1        | Instagram  | ✓                 | ✓              | 0/63 |
| 2        | RealmCraft | ✓                 | ✓              | 0/62 |
| 3        | CartolaFC  | ✓                 | ✓              | 0/63 |
| 4        | Spotify    | ✓                 | ✗              | 1/59 |
| 5        | WhatsApp   | ✓                 | ✓              | 0/59 |



# Avaliação da API Web



| Malwares |                          | SigPID<br>API Web | Virus<br>Total | Qtd   |
|----------|--------------------------|-------------------|----------------|-------|
| 1        | Hudway                   | ×                 | ✓              | 19/63 |
| 2        | covid-19                 | ✓                 | ✓              | 33/63 |
| 3        | Adobe Flash<br>Player    | ✓                 | ✓              | 28/61 |
| 4        | Corona-libya             | ×                 | ✓              | 19/63 |
| 5        | AndroidSecu<br>reProduct | ✓                 | ✓              | 10/62 |

# Conclusão



- **Tempo de execução é o mais relevante?**
- **Quantidade de dados impacta o tempo de execução**
- **É possível reduzir o número de permissões e manter 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 features**
- **Otimização do modelo**
- **Testar os modelos em smartphones modernos**
- **Criar um dataset com dados atualizados**

# Obrigado!



[jonerassolin.aluno@unipampa.edu.br](mailto:jonerassolin.aluno@unipampa.edu.br)

## Dúvidas ou sugestões?



"A ciência trabalha na fronteira entre conhecimento e ignorância.

Não temos medo de admitir o que não sabemos,  
não há nenhuma vergonha nisso. A única  
vergonha é fingir que temos todas as respostas."

(Neil deGrasse Tyson)