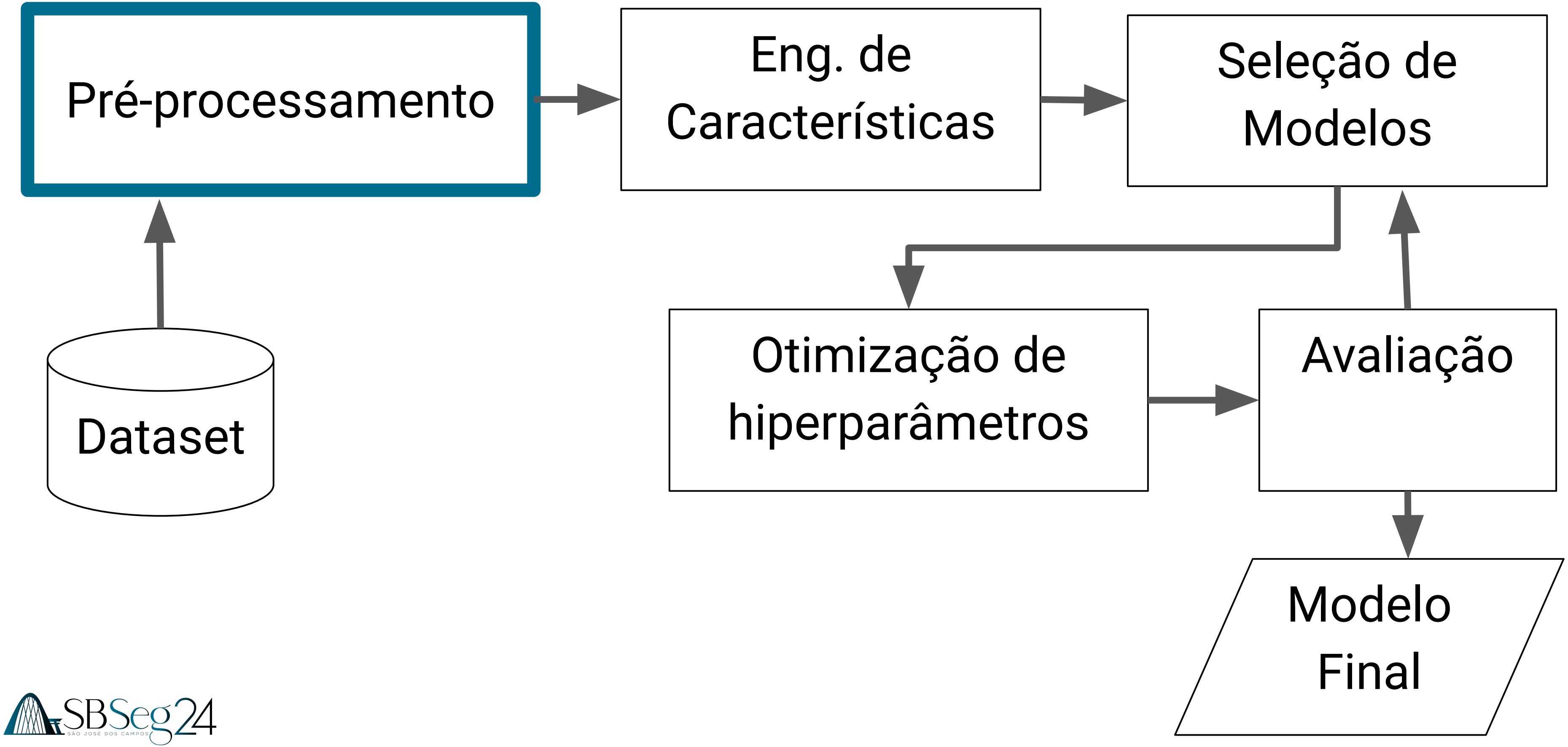
MH-AutoML: Transparência, Interpretabilidade e Desempenho na Detecção de Malware Android

Joner Assolin[1,], Gabriel Canto[1,], Diego Kreutz[2], Eduardo Feitosa[1,]

Universidade federal do Amazonas[1], Universidade Federal do Pampa[2]

Motivação



Pré-processamento

Eng. de Características

Seleção de Modelos

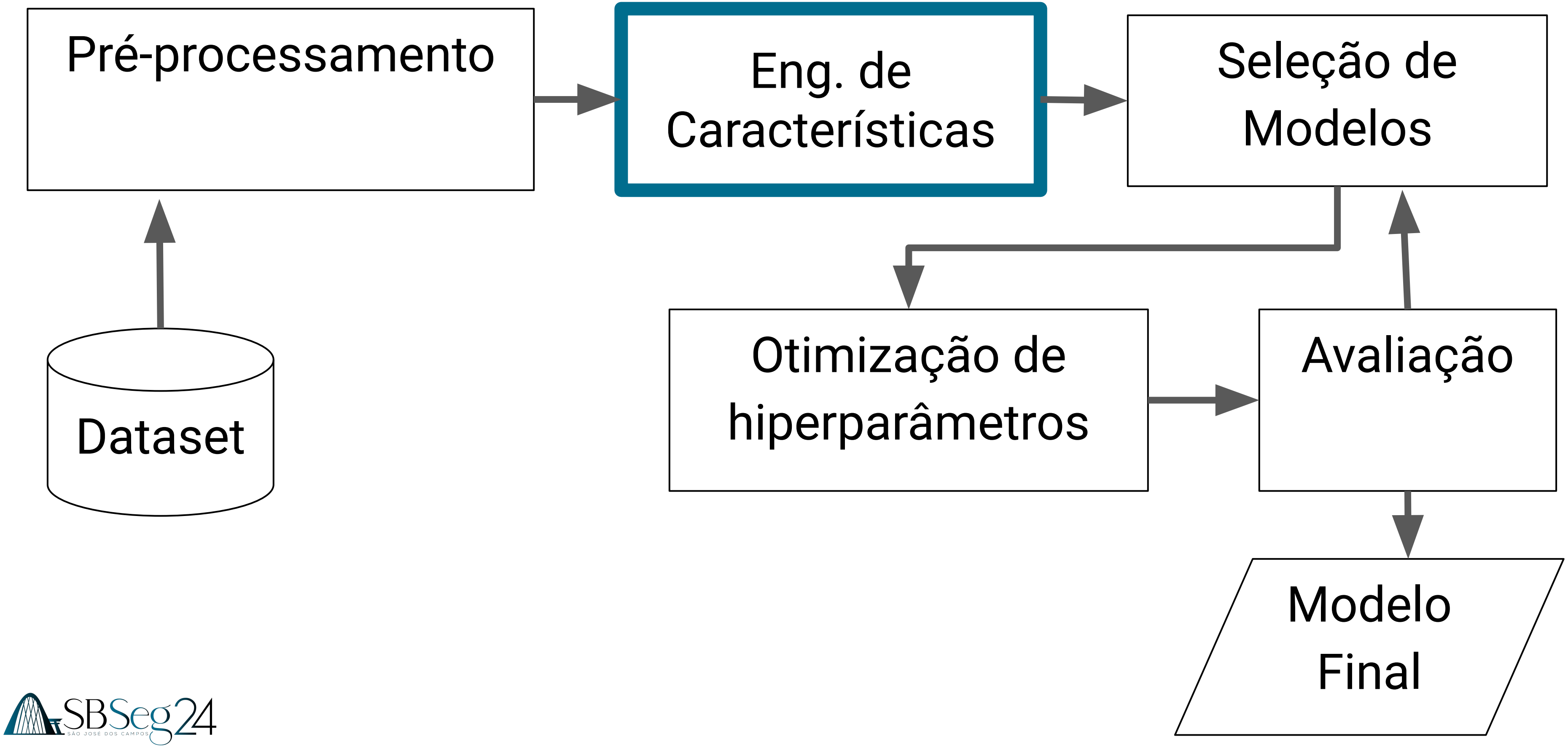
Dataset

Otimização de hiperparâmetros

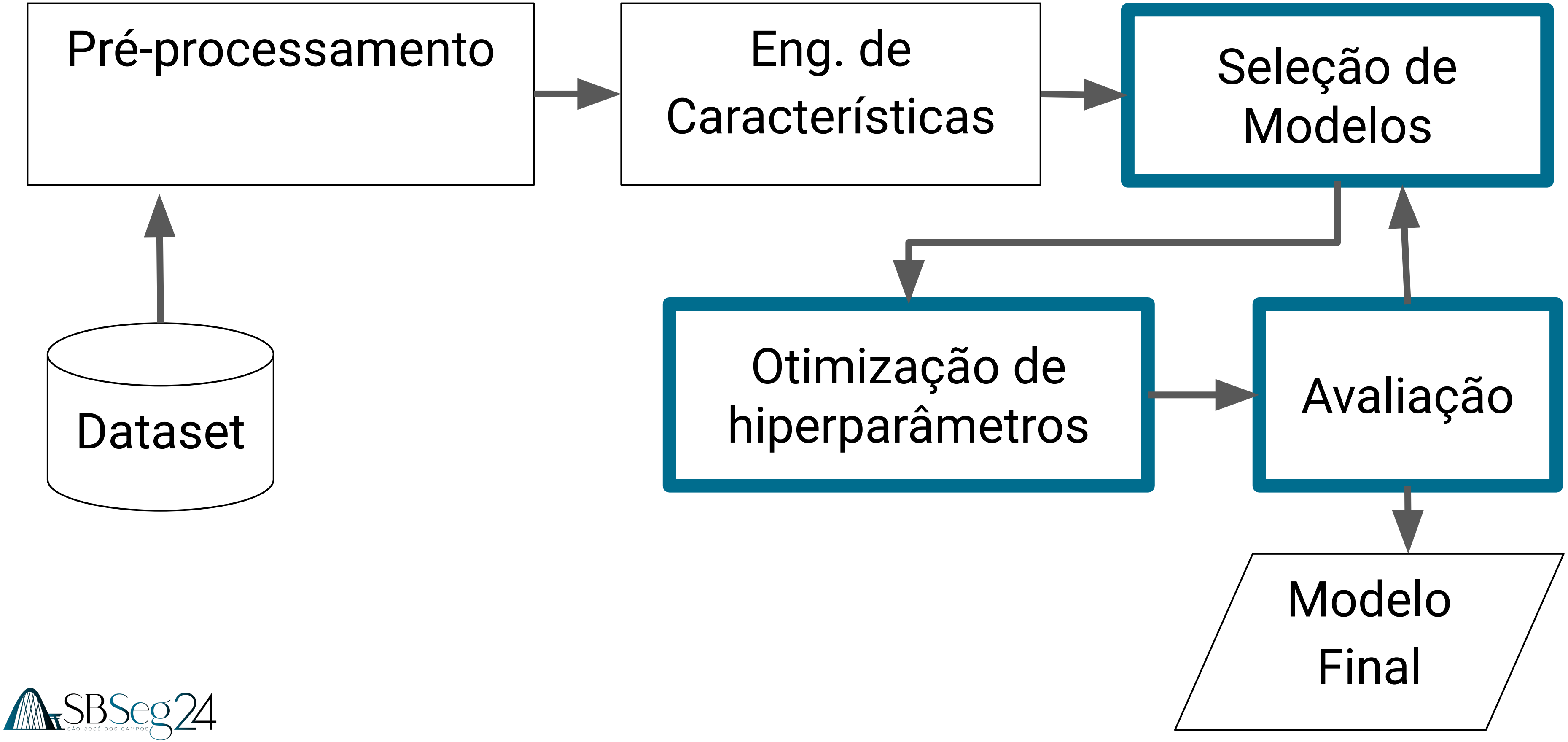
Avaliação

Modelo Final

Motivação


Pré-processamento
Eng. de Características
Seleção de Modelos
Dataset
Otimização de hiperparâmetros
Avaliação
Modelo Final

Motivação


Pré-processamento

Eng. de Características

Seleção de Modelos

Dataset

Otimização de hiperparâmetros

Avaliação

Modelo Final

SBSeg24 SÃO JOSÉ DOS CAMPOS

Desafio(s)

1 Transparência

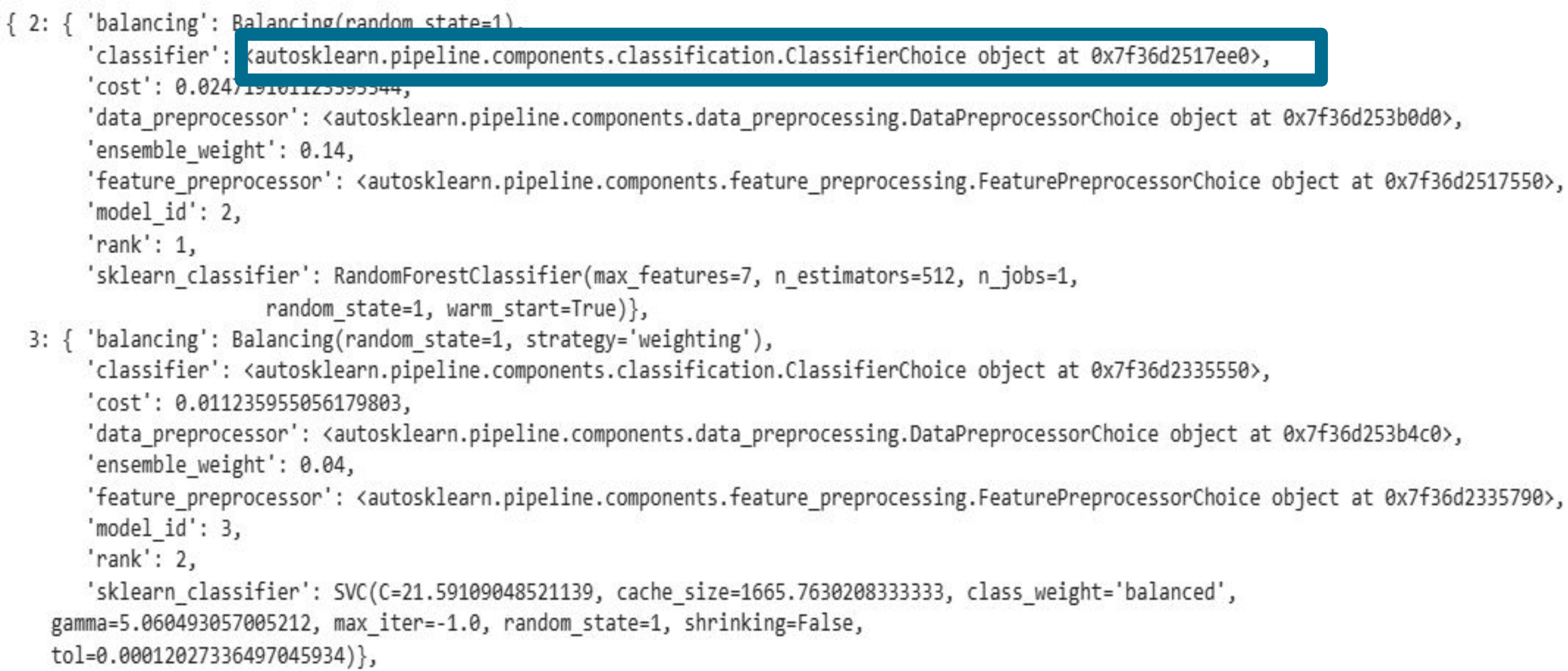
- Qual etapa está sendo executada?
- Qual método de seleção de características foi utilizado?
- Quais foram as características selecionadas?
- Quais hiperparâmetros foram otimizados?

Desafio(s)

2 Interpretabilidade

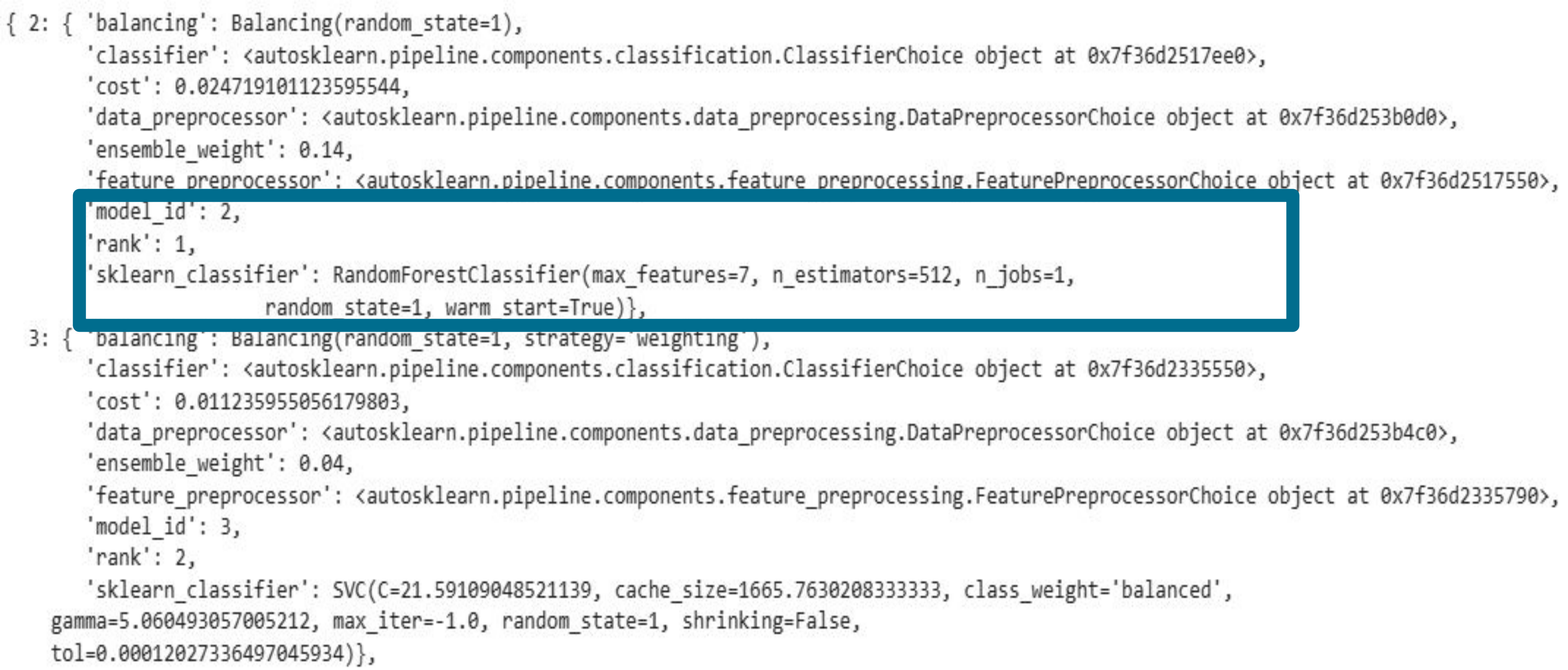
- Qual classe tem mais impacto na predição?
- Quais características mais contribuem para predição?

Auto-Sklearn



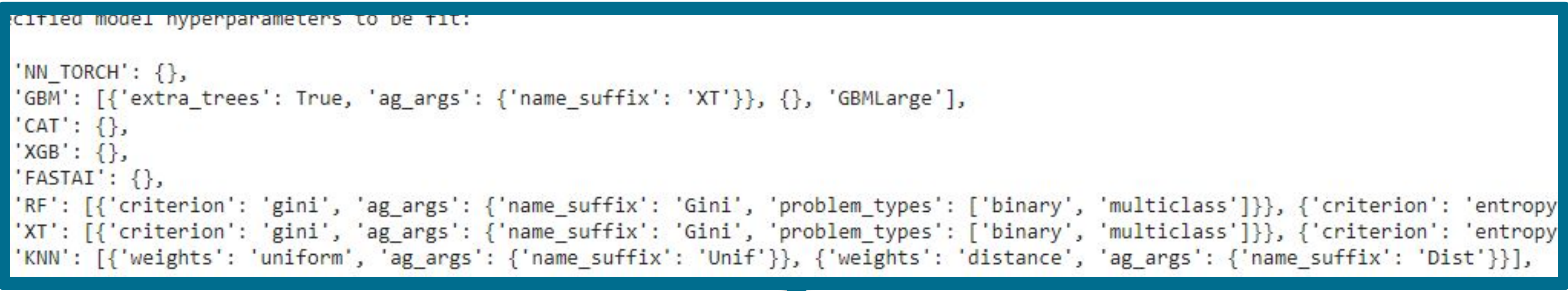
```
{ 2: { 'balancing': Balancing(random_state=1),
       'classifier': <autosklearn.pipeline.components.classification.ClassifierChoice object at 0x7f36d2517ee0>,
       'cost': 0.024719101123595544,
       'data_preprocessor': <autosklearn.pipeline.components.data_preprocessing.DataPreprocessorChoice object at 0x7f36d253b0d0>,
       'ensemble_weight': 0.14,
       'feature_preprocessor': <autosklearn.pipeline.components.feature_preprocessing.FeaturePreprocessorChoice object at 0x7f36d2517550>,
       'model_id': 2,
       'rank': 1,
       'sklearn_classifier': RandomForestClassifier(max_features=7, n_estimators=512, n_jobs=1,
                       random_state=1, warm_start=True)},
  3: { 'balancing': Balancing(random_state=1, strategy='weighting'),
       'classifier': <autosklearn.pipeline.components.classification.ClassifierChoice object at 0x7f36d2335550>,
       'cost': 0.011235955056179803,
       'data_preprocessor': <autosklearn.pipeline.components.data_preprocessing.DataPreprocessorChoice object at 0x7f36d253b4c0>,
       'ensemble_weight': 0.04,
       'feature_preprocessor': <autosklearn.pipeline.components.feature_preprocessing.FeaturePreprocessorChoice object at 0x7f36d2335790>,
       'model_id': 3,
       'rank': 2,
       'sklearn_classifier': SVC(C=21.59109048521139, cache_size=1665.7630208333333, class_weight='balanced',
    gamma=5.060493057005212, max_iter=-1.0, random_state=1, shrinking=False,
    tol=0.00012027336497045934)},
```

Auto-Sklearn



```
{ 2: { 'balancing': Balancing(random_state=1),
       'classifier': <autosklearn.pipeline.components.classification.ClassifierChoice object at 0x7f36d2517ee0>,
       'cost': 0.024719101123595544,
       'data_preprocessor': <autosklearn.pipeline.components.data_preprocessing.DataPreprocessorChoice object at 0x7f36d253b0d0>,
       'ensemble_weight': 0.14,
       'feature_preprocessor': <autosklearn.pipeline.components.feature_preprocessing.FeaturePreprocessorChoice object at 0x7f36d2517550>,
       'model_id': 2,
       'rank': 1,
       'sklearn_classifier': RandomForestClassifier(max_features=7, n_estimators=512, n_jobs=1,
                       random_state=1, warm_start=True)},
  3: { 'balancing': Balancing(random_state=1, strategy='weighting'),
       'classifier': <autosklearn.pipeline.components.classification.ClassifierChoice object at 0x7f36d2335550>,
       'cost': 0.011235955056179803,
       'data_preprocessor': <autosklearn.pipeline.components.data_preprocessing.DataPreprocessorChoice object at 0x7f36d253b4c0>,
       'ensemble_weight': 0.04,
       'feature_preprocessor': <autosklearn.pipeline.components.feature_preprocessing.FeaturePreprocessorChoice object at 0x7f36d2335790>,
       'model_id': 3,
       'rank': 2,
       'sklearn_classifier': SVC(C=21.59109048521139, cache_size=1665.7630208333333, class_weight='balanced',
    gamma=5.060493057005212, max_iter=-1.0, random_state=1, shrinking=False,
    tol=0.00012027336497045934)},
```

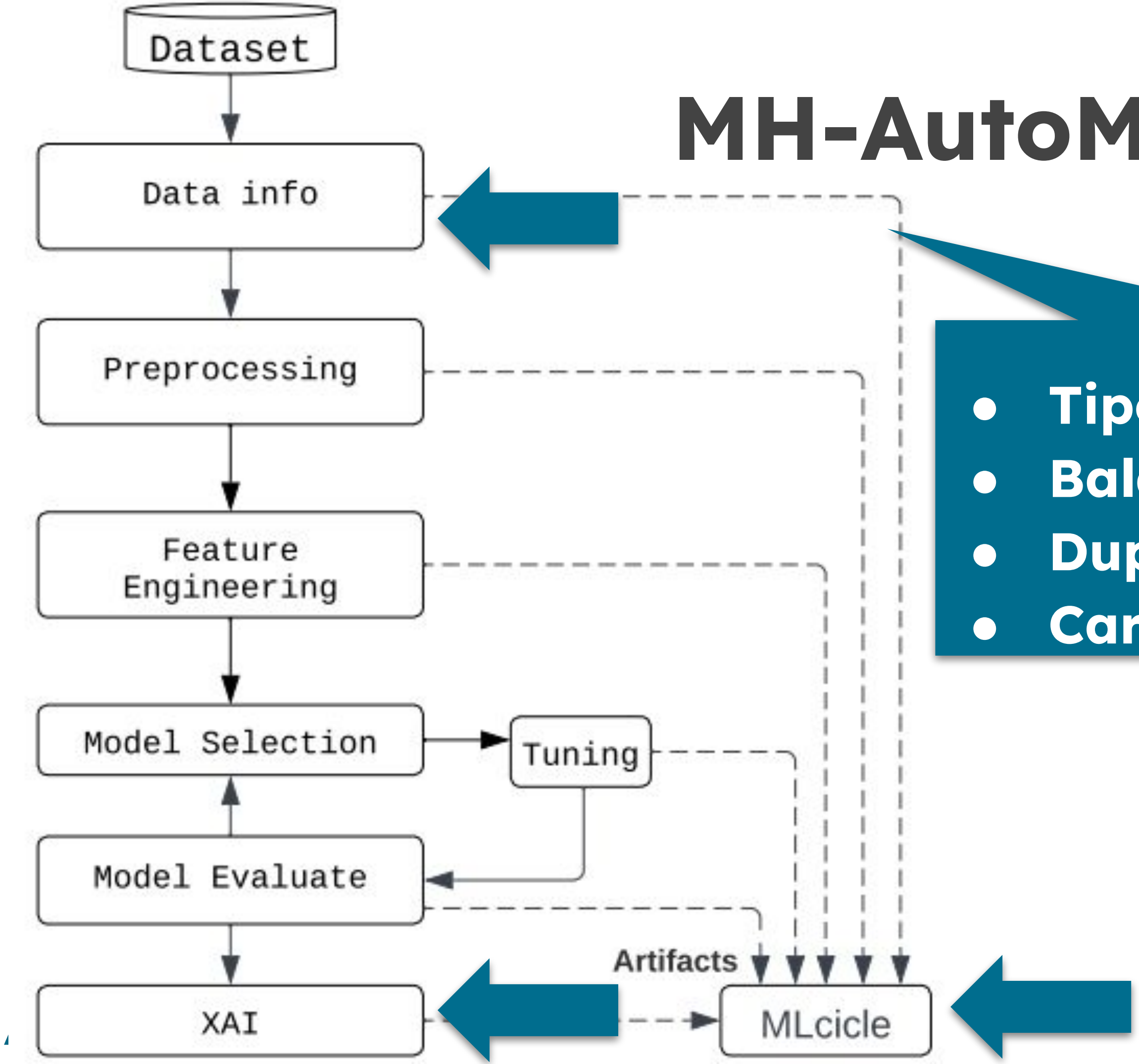
AutoGluon



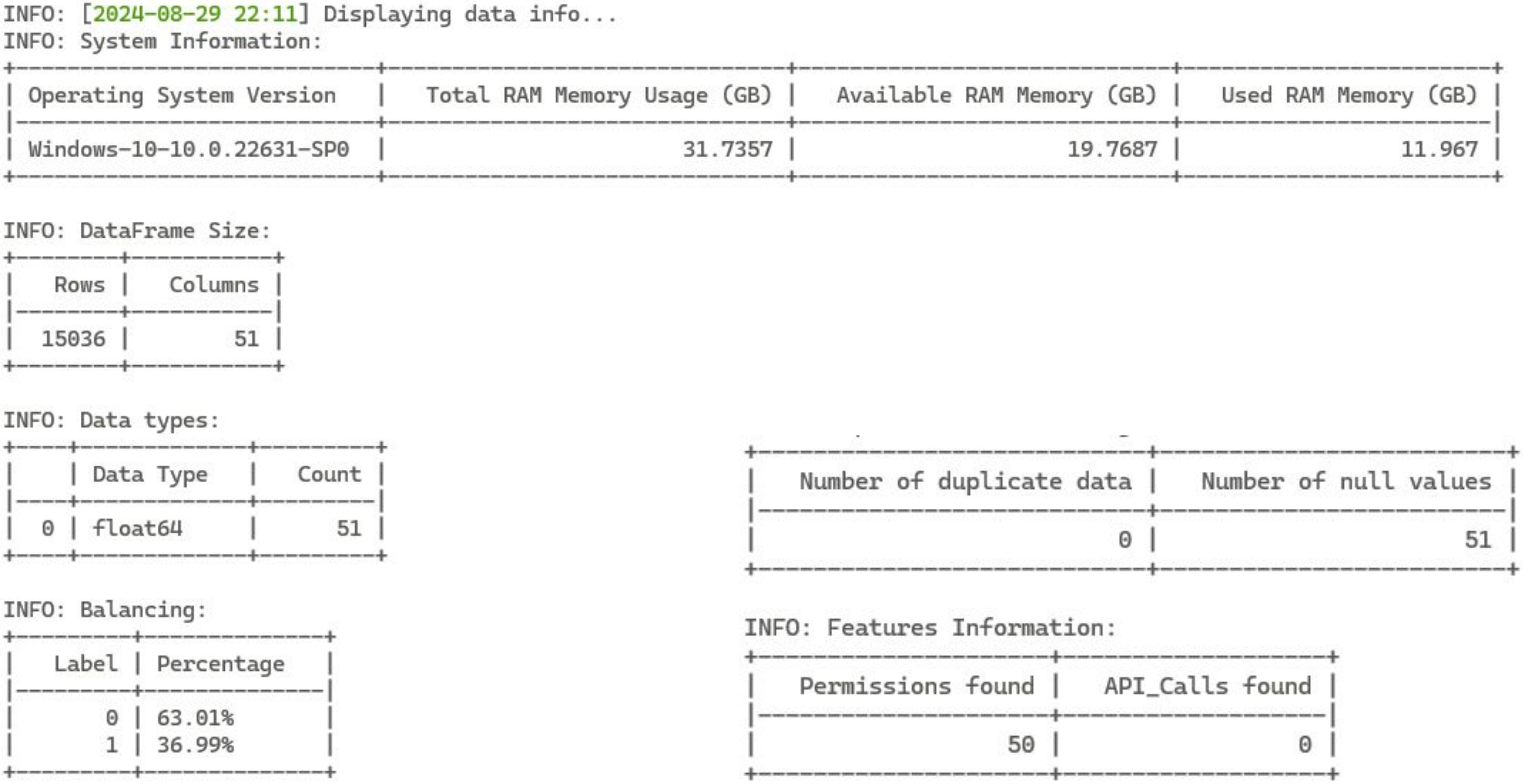
```
cified model hyperparameters to be fit:

'NN_TORCH': {},
'GBM': [{'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}}, {}, 'GBMLarge'],
'CAT': {},
'XGB': {},
'FASTAI': {},
'RF': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy
'XT': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'entropy
'KNN': [{'weights': 'uniform', 'ag_args': {'name_suffix': 'Unif'}}, {'weights': 'distance', 'ag_args': {'name_suffix': 'Dist'}}],
```

- **Valores pré-definidos**
- **N-neighbourd ?**
- **N_estimator ?**

MH-AutoML



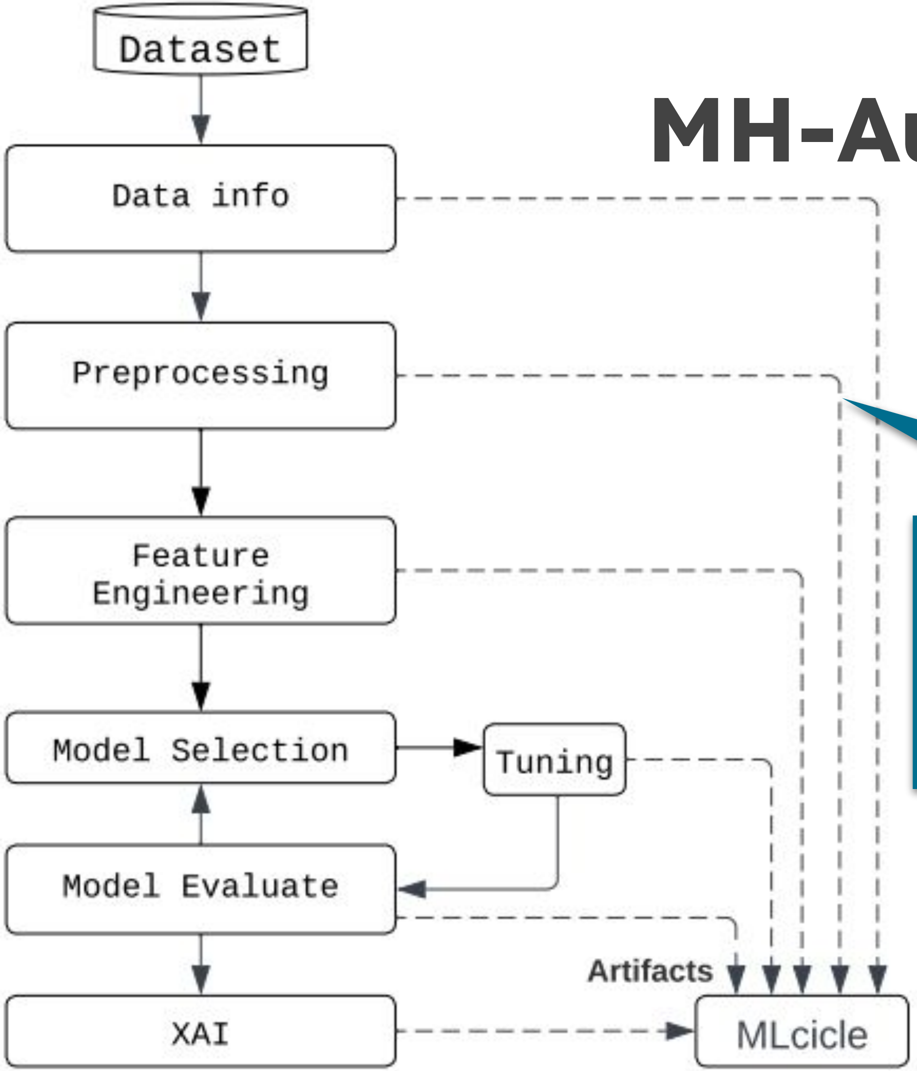
- Tipos de dados do dataset
- Balanceamento dos dados
- Duplicidade de dados
- Caracteristicas



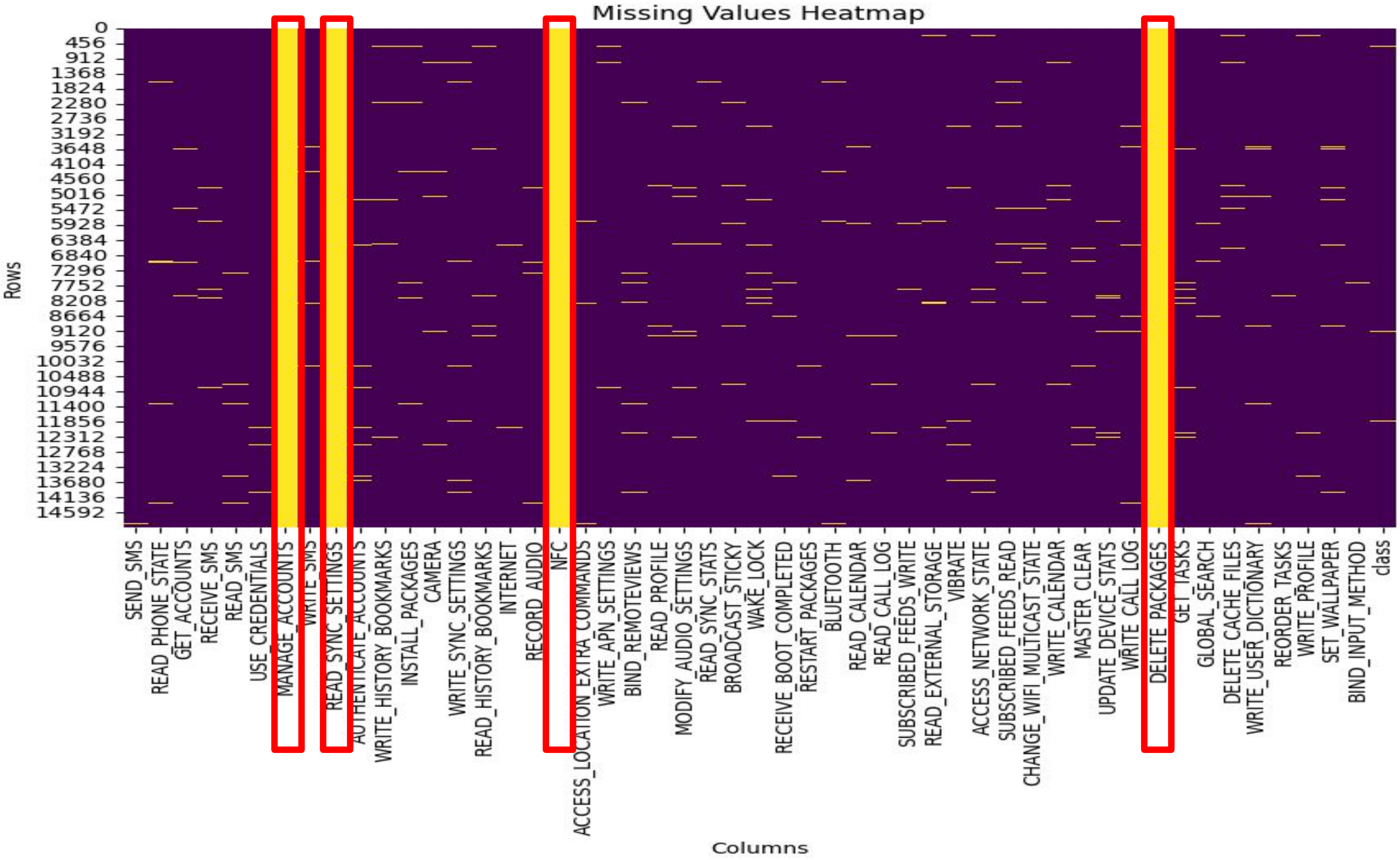
```
INFO: [2024-08-29 22:11] Displaying data info...
INFO: System Information:
+----------------------------+---------------------------------+------------------------------+---------------------------+
| Operating System Version   |   Total RAM Memory Usage (GB) |   Available RAM Memory (GB) |   Used RAM Memory (GB) |
|----------------------------+---------------------------------+------------------------------+---------------------------|
| Windows-10-10.0.22631-SP0  |                       31.7357 |                     19.7687 |                  11.967 |
+----------------------------+---------------------------------+------------------------------+---------------------------+

INFO: DataFrame Size:
+---------+-----------+
|    Rows |   Columns |
|---------+-----------|
|   15036 |        51 |
+---------+-----------+

INFO: Data types:
+----+--------------+---------+
|    | Data Type    |   Count |
|----+--------------+---------|
|  0 | float64      |      51 |
+----+--------------+---------+

INFO: Balancing:
+---------+--------------+
|   Label | Percentage   |
|---------+--------------|
|       0 | 63.01%       |
|       1 | 36.99%       |
+---------+--------------+
```

```
+------------------------------+-----------------------------+
|   Number of duplicate data |   Number of null values |
|------------------------------+-----------------------------|
|                          0 |                      51 |
+------------------------------+-----------------------------+

INFO: Features Information:
+-----------------------+----------------------+
|   Permissions found |   API_Calls found |
|-----------------------+----------------------|
|                  50 |                 0 |
+-----------------------+----------------------+
```

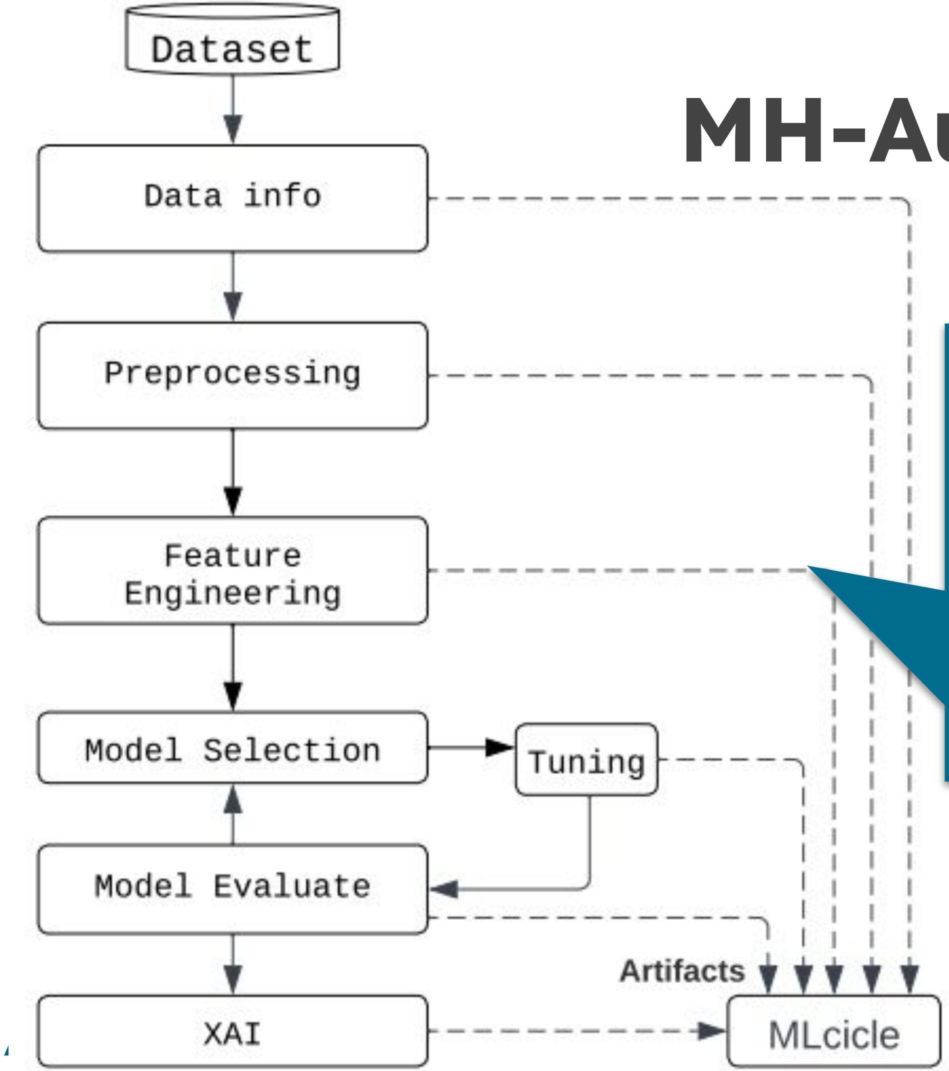
MH-AutoML



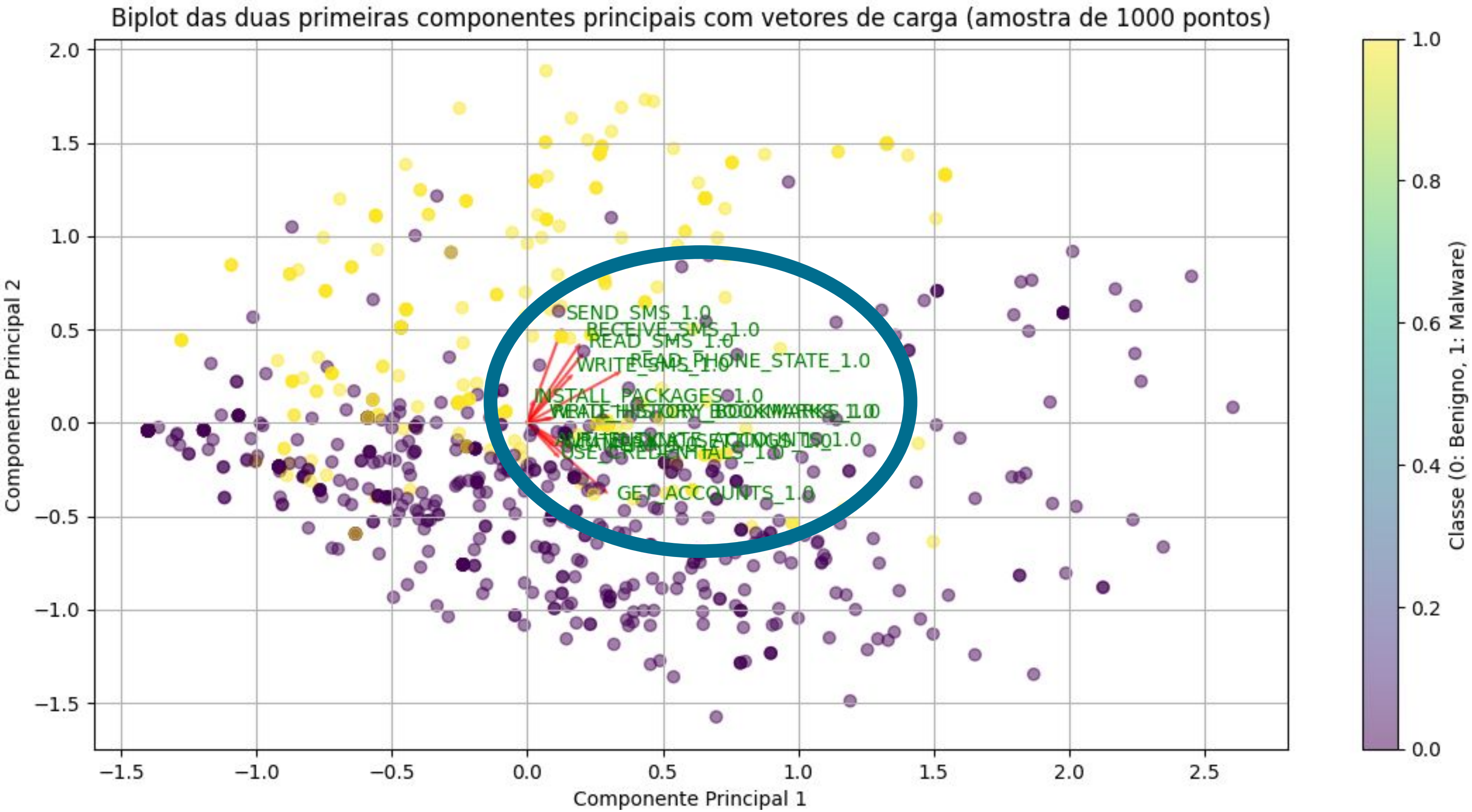
- Transformação de dados
- Reconhecimento de assinatura criptográfica


Missing Values Heatmap
Rows
0
456
912
1368
1824
2280
2736
3192
3648
4104
4560
5016
5472
5928
6384
6840
7296
7752
8208
8664
9120
9576
10032
10488
10944
11400
11856
12312
12768
13224
13680
14136
14592
SEND_SMS
READ_PHONE_STATE
GET_ACCOUNTS
RECEIVE_SMS
READ_SMS
USE_CREDENTIALS
MANAGE_ACCOUNTS
WRITE_SMS
READ_SYNC_SETTINGS
AUTHENTICATE_ACCOUNTS
WRITE_HISTORY_BOOKMARKS
INSTALL_PACKAGES
CAMERA
WRITE_SYNC_SETTINGS
READ_HISTORY_BOOKMARKS
INTERNET
RECORD_AUDIO
NFC
ACCESS_LOCATION_EXTRA_COMMANDS
WRITE_APN_SETTINGS
BIND_REMOTEVIEWS
READ_PROFILE
MODIFY_AUDIO_SETTINGS
READ_SYNC_STATS
BROADCAST_STICKY
WAKE_LOCK
RECEIVE_BOOT_COMPLETED
RESTART_PACKAGES
BLUETOOTH
READ_CALENDAR
READ_CALL_LOG
SUBSCRIBED_FEEDS_WRITE
READ_EXTERNAL_STORAGE
VIBRATE
ACCESS_NETWORK_STATE
SUBSCRIBED_FEEDS_READ
CHANGE_WIFI_MULTICAST_STATE
WRITE_CALENDAR
MASTER_CLEAR
UPDATE_DEVICE_STATS
WRITE_CALL_LOG
DELETE_PACKAGES
GET_TASKS
GLOBAL_SEARCH
DELETE_CACHE_FILES
WRITE_USER_DICTIONARY
REORDER_TASKS
WRITE_PROFILE
SET_WALLPAPER
BIND_INPUT_METHOD
class
Columns

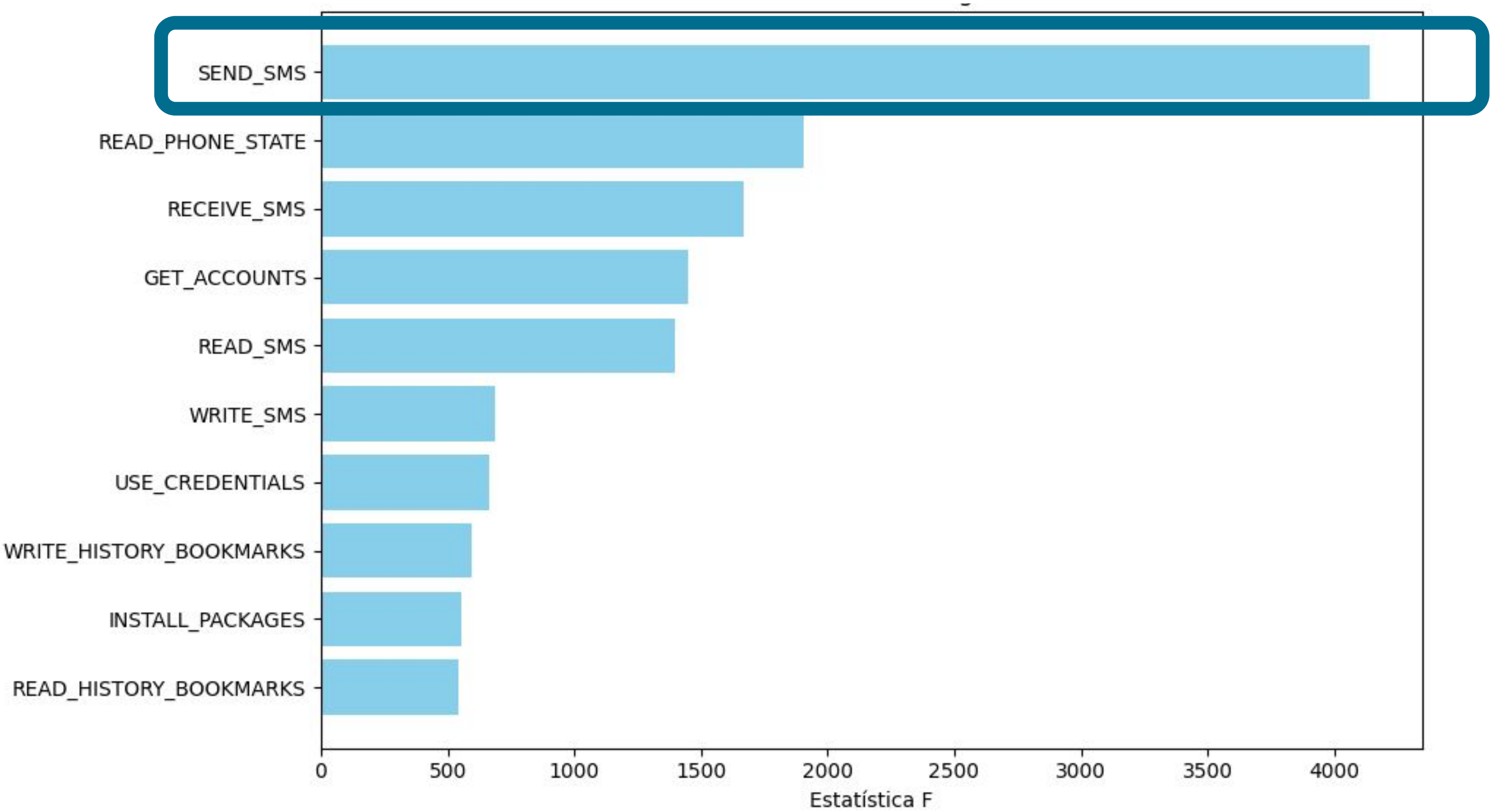
MH-AutoML

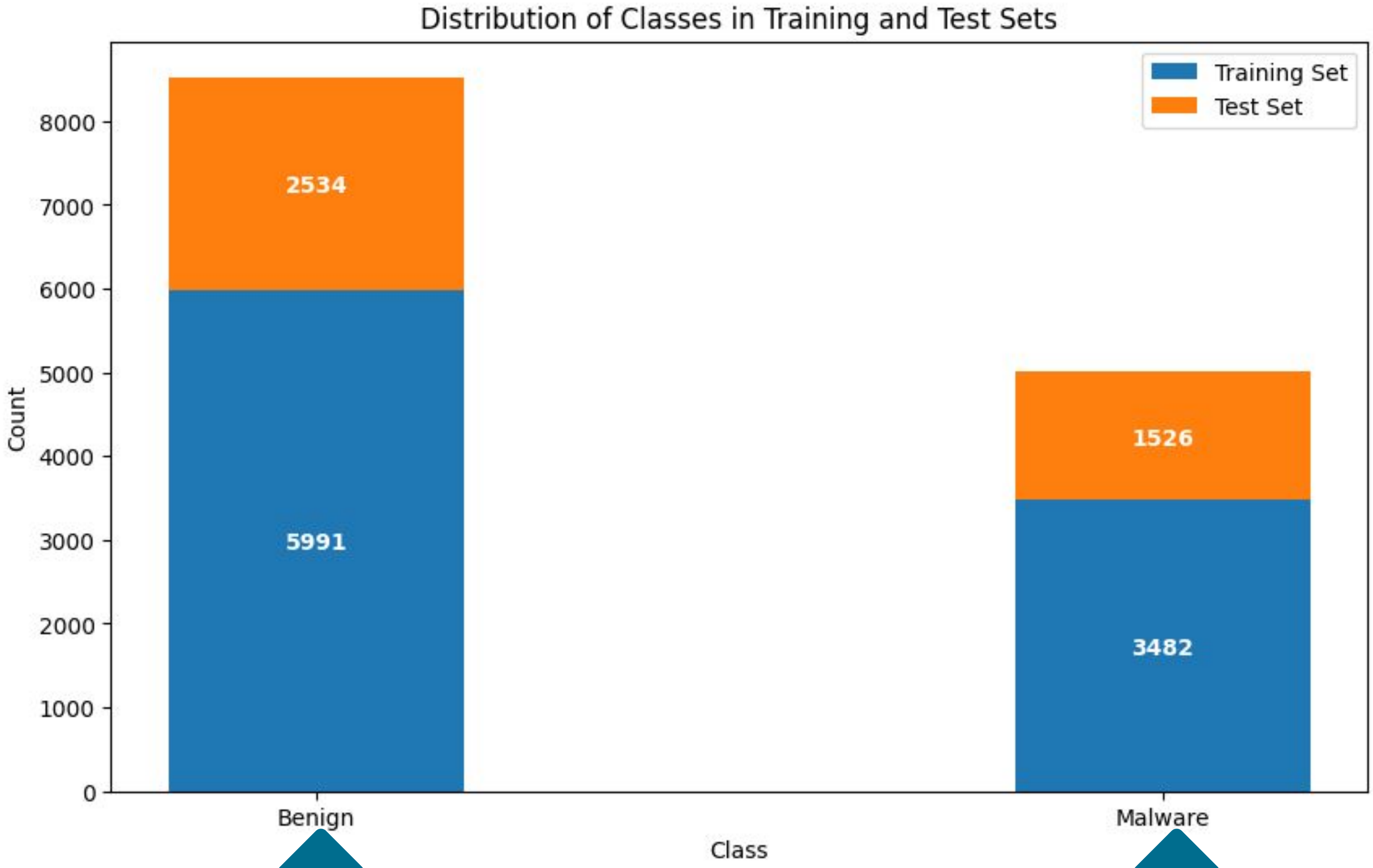


- Subset da seleção
- Subset dos dados de treino
- Gráfico da importância das características
- Gráfico de divisão treino teste por classe



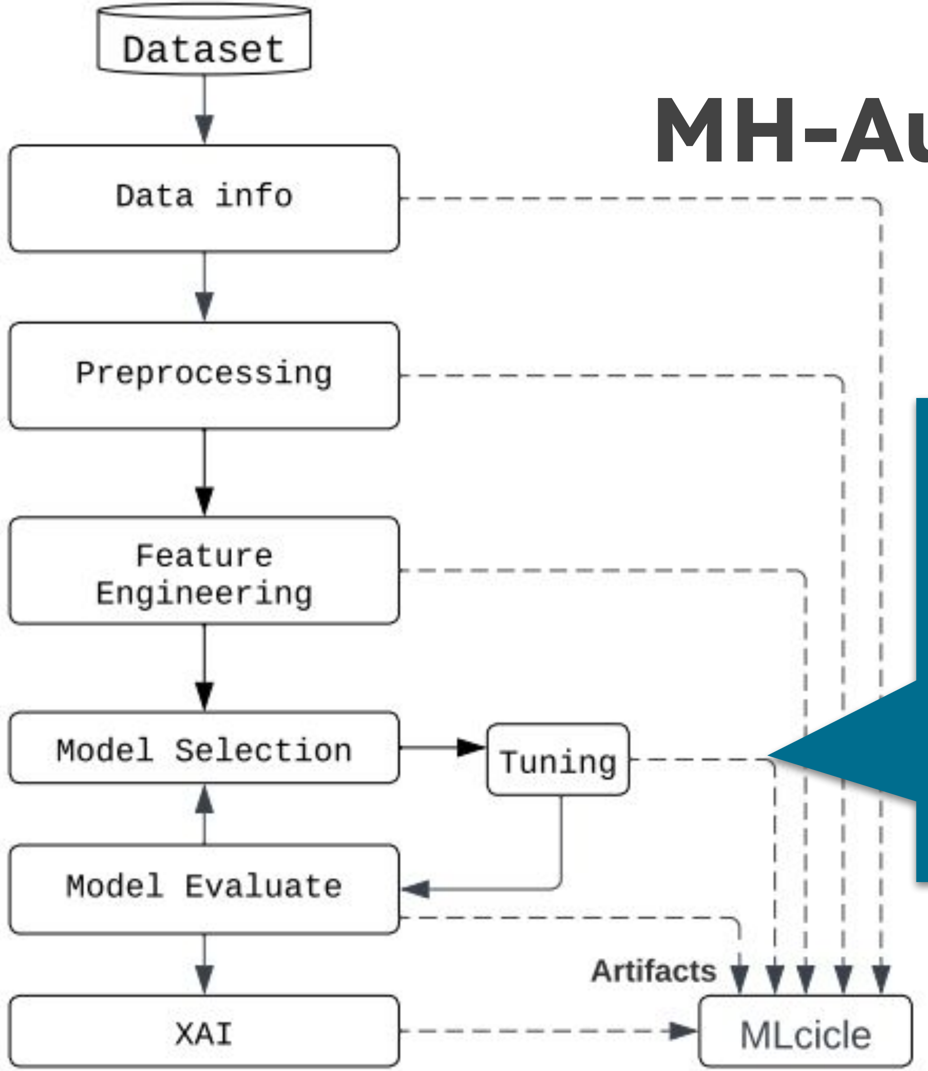
Biplot das duas primeiras componentes principais com vetores de carga (amostra de 1000 pontos)


SEND_SMS
READ_PHONE_STATE
RECEIVE_SMS
GET_ACCOUNTS
READ_SMS
WRITE_SMS
USE_CREDENTIALS
WRITE_HISTORY_BOOKMARKS
INSTALL_PACKAGES
READ_HISTORY_BOOKMARKS
0
500
1000
1500
2000
2500
3000
3500
4000
Estatística F

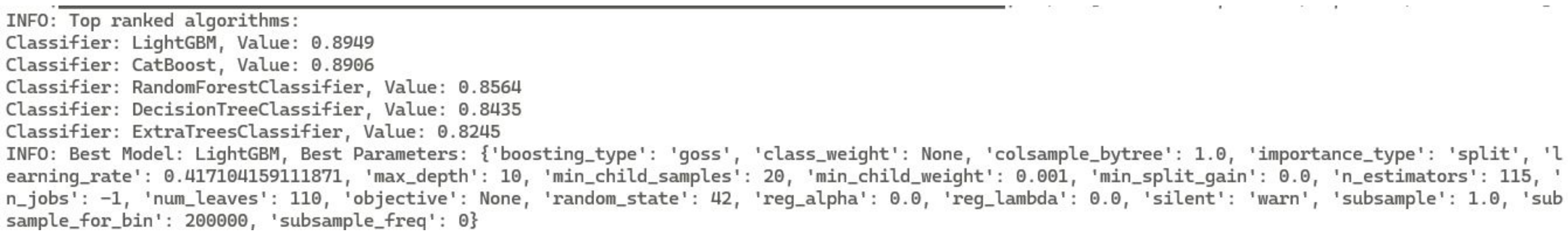


Distribution of Classes in Training and Test Sets

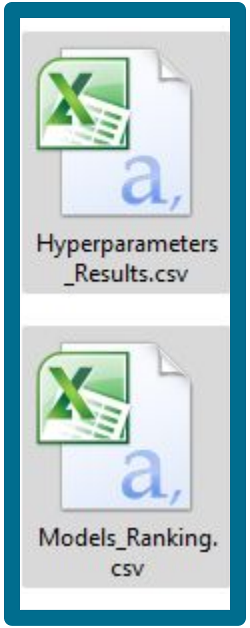
MH-AutoML

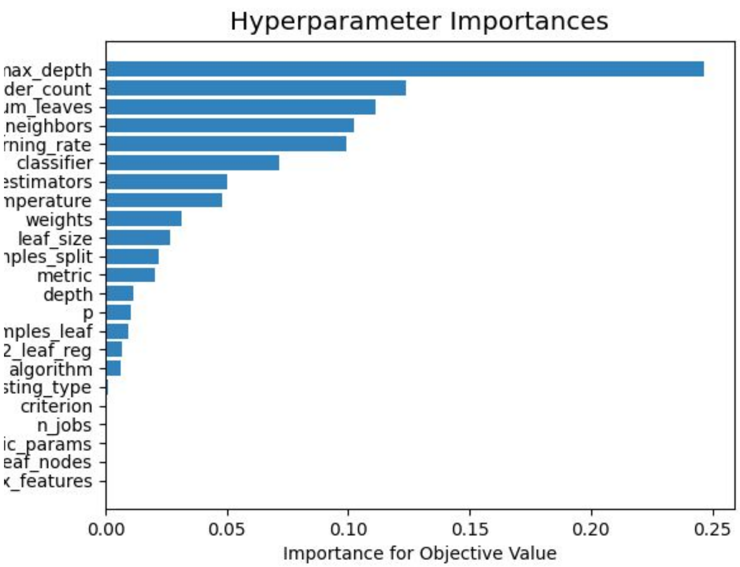


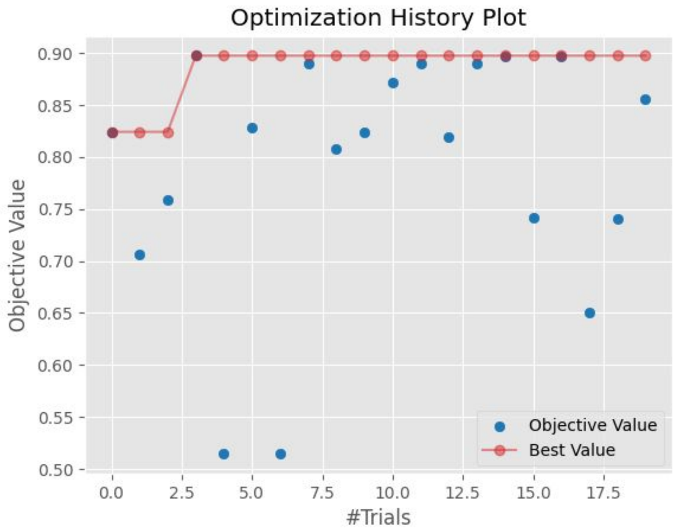
- Hiperparâmetros e modelos
- Ranking dos modelos
- Gráfico da importância de hiperparâmetros
- Gráfico de estudos dos parâmetros

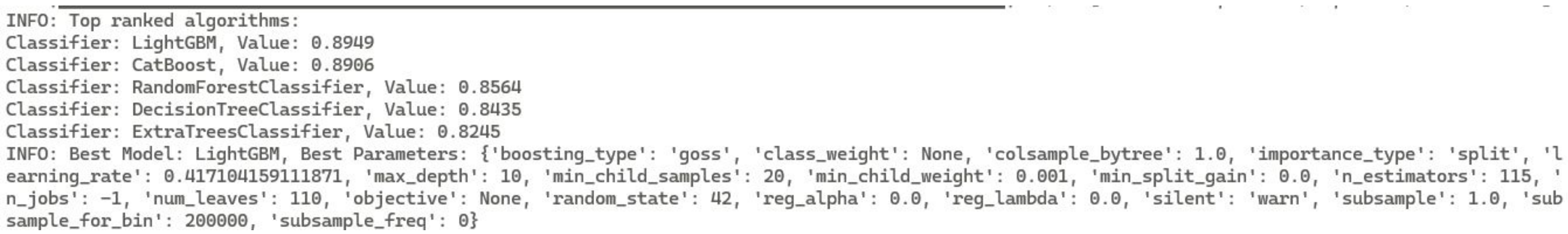


```
INFO: Top ranked algorithms:
Classifier: LightGBM, Value: 0.8949
Classifier: CatBoost, Value: 0.8906
Classifier: RandomForestClassifier, Value: 0.8564
Classifier: DecisionTreeClassifier, Value: 0.8435
Classifier: ExtraTreesClassifier, Value: 0.8245
INFO: Best Model: LightGBM, Best Parameters: {'boosting_type': 'goss', 'class_weight': None, 'colsample_bytree': 1.0, 'importance_type': 'split', 'l
earning_rate': 0.417104159111871, 'max_depth': 10, 'min_child_samples': 20, 'min_child_weight': 0.001, 'min_split_gain': 0.0, 'n_estimators': 115, '
n_jobs': -1, 'num_leaves': 110, 'objective': None, 'random_state': 42, 'reg_alpha': 0.0, 'reg_lambda': 0.0, 'silent': 'warn', 'subsample': 1.0, 'sub
sample_for_bin': 200000, 'subsample_freq': 0}
```





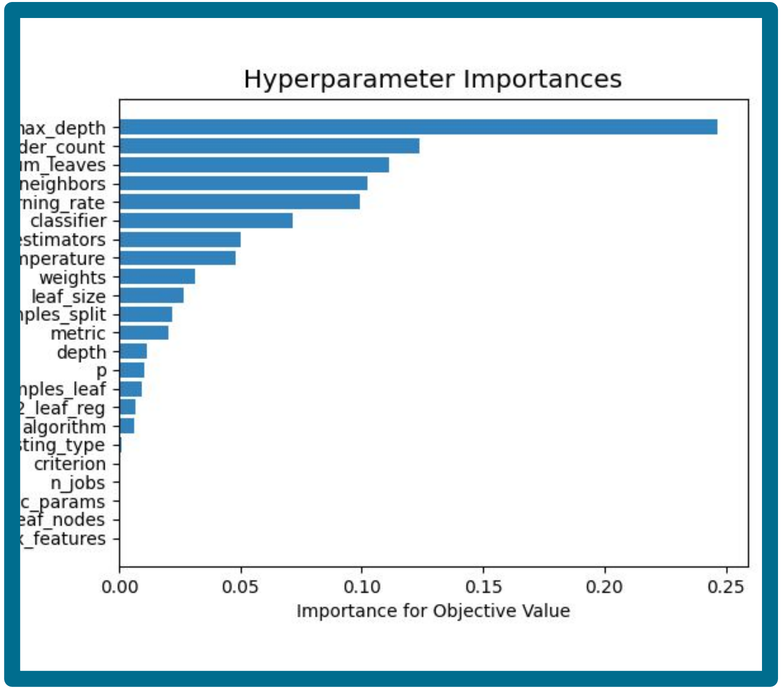


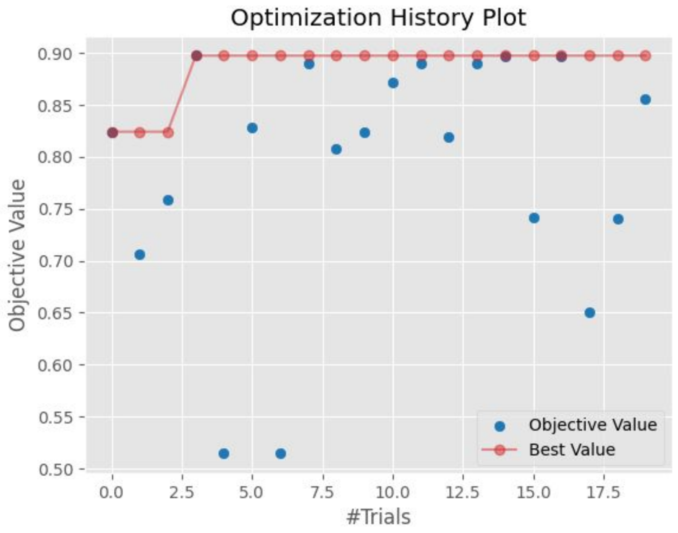


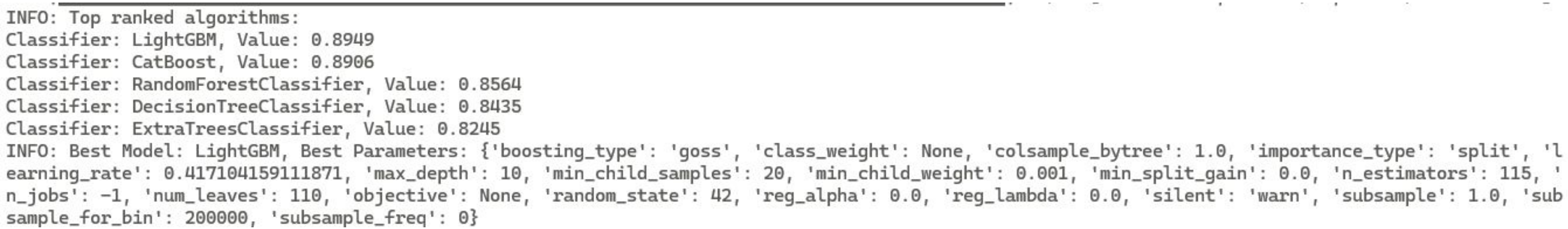
```
INFO: Top ranked algorithms:
Classifier: LightGBM, Value: 0.8949
Classifier: CatBoost, Value: 0.8906
Classifier: RandomForestClassifier, Value: 0.8564
Classifier: DecisionTreeClassifier, Value: 0.8435
Classifier: ExtraTreesClassifier, Value: 0.8245
INFO: Best Model: LightGBM, Best Parameters: {'boosting_type': 'goss', 'class_weight': None, 'colsample_bytree': 1.0, 'importance_type': 'split', 'l
earning_rate': 0.41710415911871, 'max_depth': 10, 'min_child_samples': 20, 'min_child_weight': 0.001, 'min_split_gain': 0.0, 'n_estimators': 115, '
n_jobs': -1, 'num_leaves': 110, 'objective': None, 'random_state': 42, 'reg_alpha': 0.0, 'reg_lambda': 0.0, 'silent': 'warn', 'subsample': 1.0, 'sub
sample_for_bin': 200000, 'subsample_freq': 0}
```

Hyperparameters_Results.csv

Models_Ranking.csv



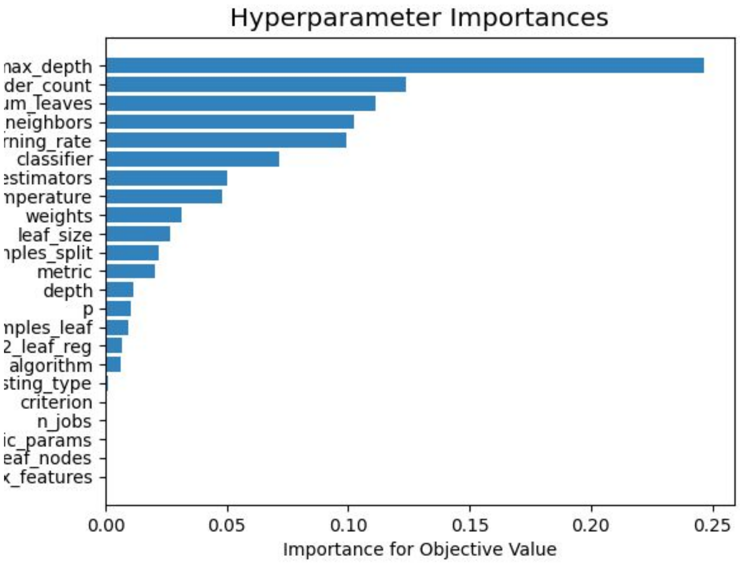


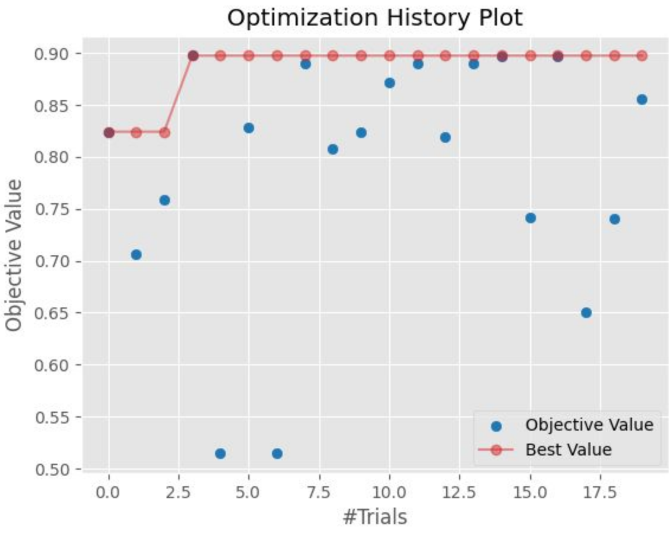


```
INFO: Top ranked algorithms:
Classifier: LightGBM, Value: 0.8949
Classifier: CatBoost, Value: 0.8906
Classifier: RandomForestClassifier, Value: 0.8564
Classifier: DecisionTreeClassifier, Value: 0.8435
Classifier: ExtraTreesClassifier, Value: 0.8245
INFO: Best Model: LightGBM, Best Parameters: {'boosting_type': 'goss', 'class_weight': None, 'colsample_bytree': 1.0, 'importance_type': 'split', 'l
earning_rate': 0.417104159111871, 'max_depth': 10, 'min_child_samples': 20, 'min_child_weight': 0.001, 'min_split_gain': 0.0, 'n_estimators': 115, '
n_jobs': -1, 'num_leaves': 110, 'objective': None, 'random_state': 42, 'reg_alpha': 0.0, 'reg_lambda': 0.0, 'silent': 'warn', 'subsample': 1.0, 'sub
sample_for_bin': 200000, 'subsample_freq': 0}
```

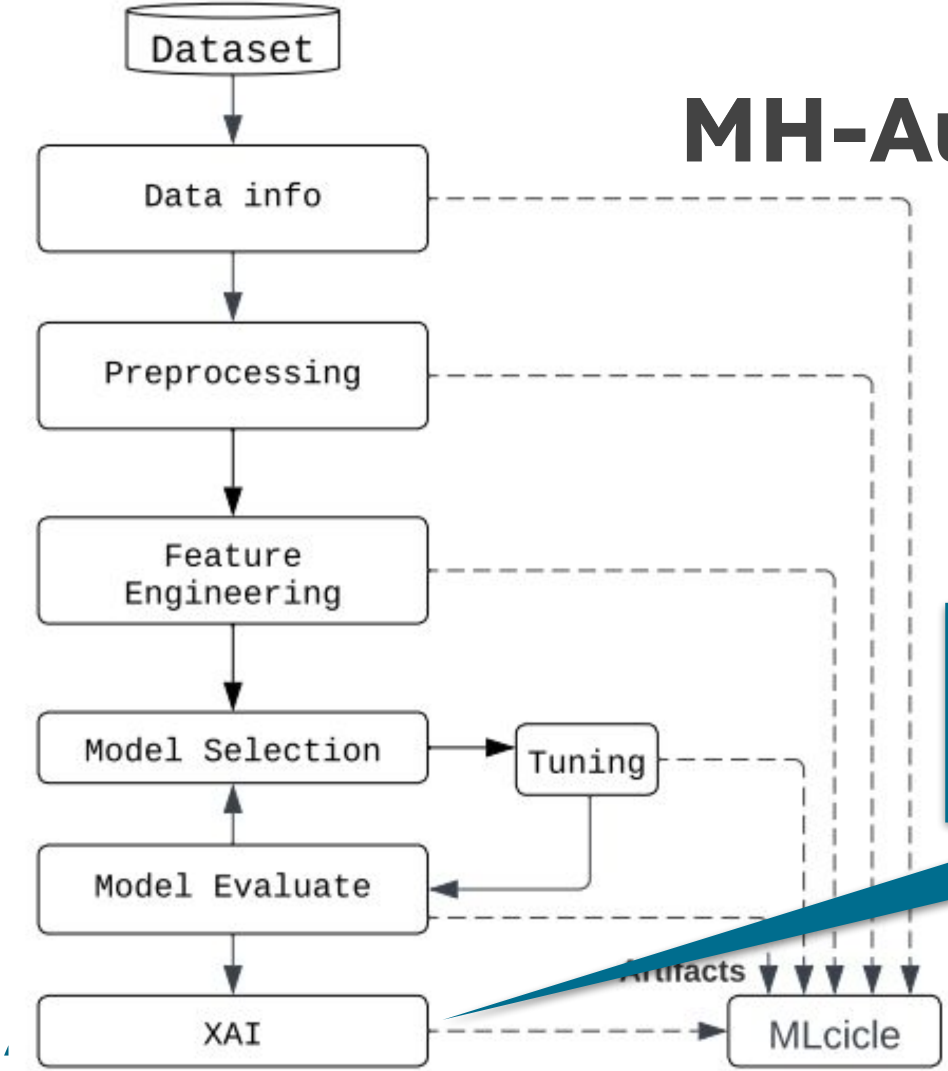
Hyperparameters_Results.csv

Models_Ranking.csv

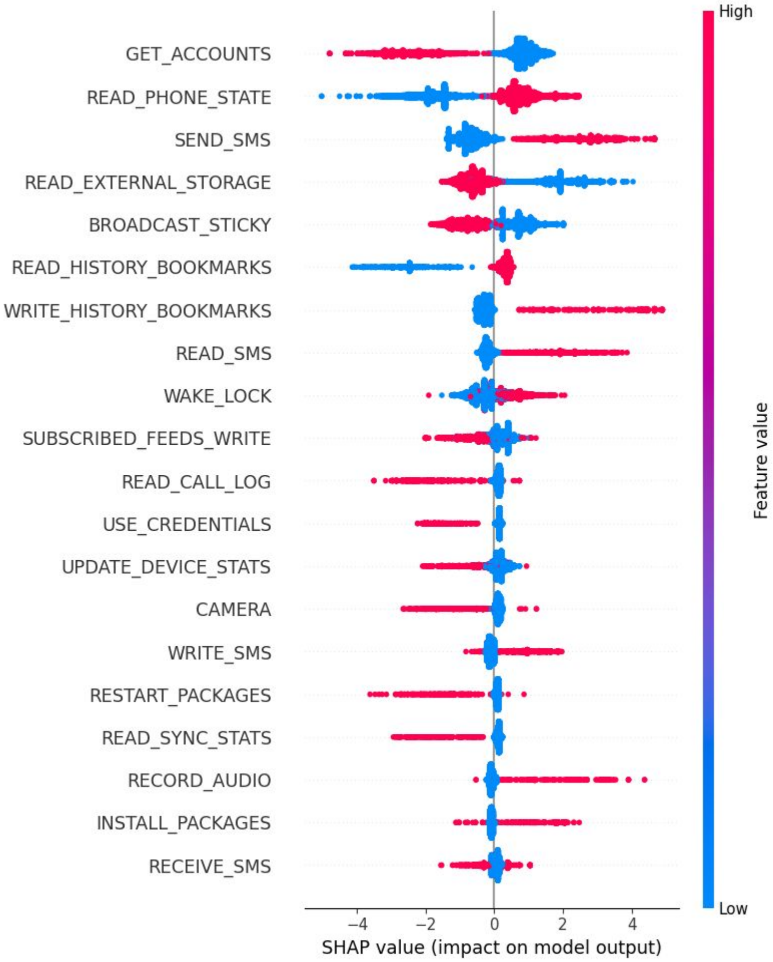


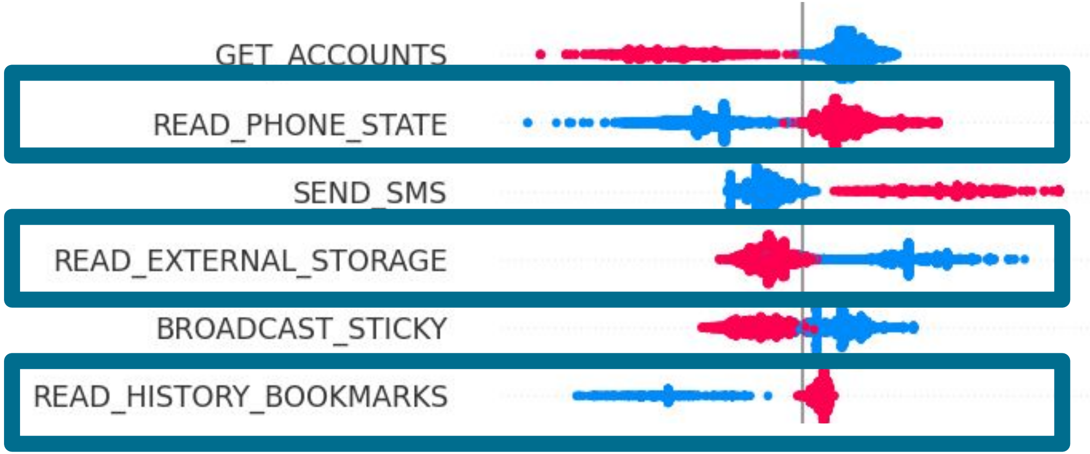


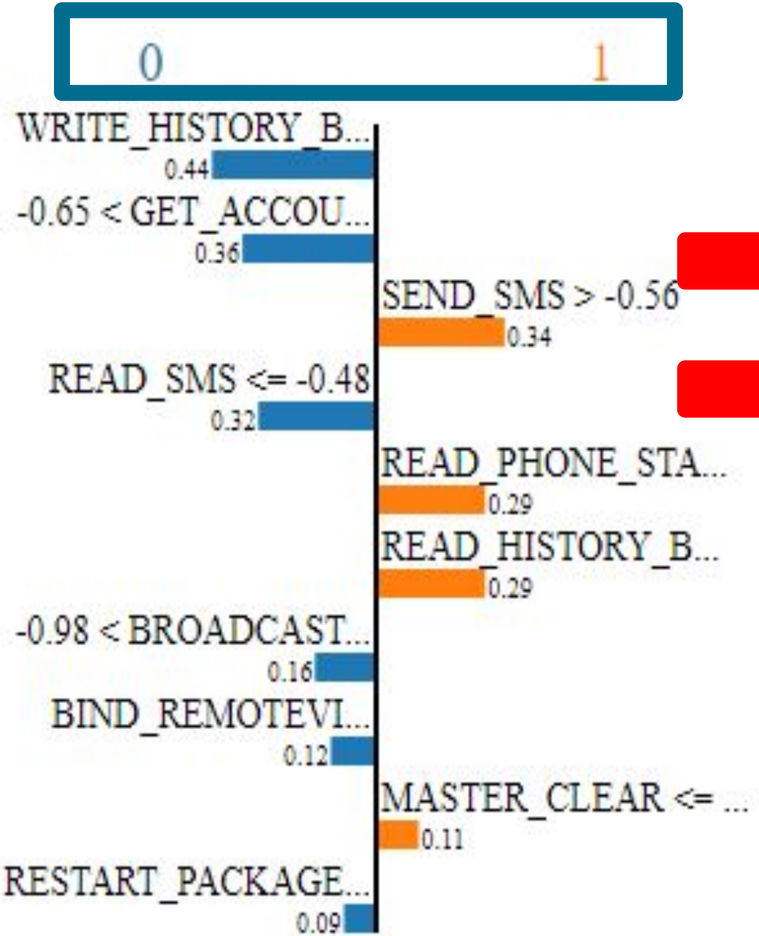
MH-AutoML

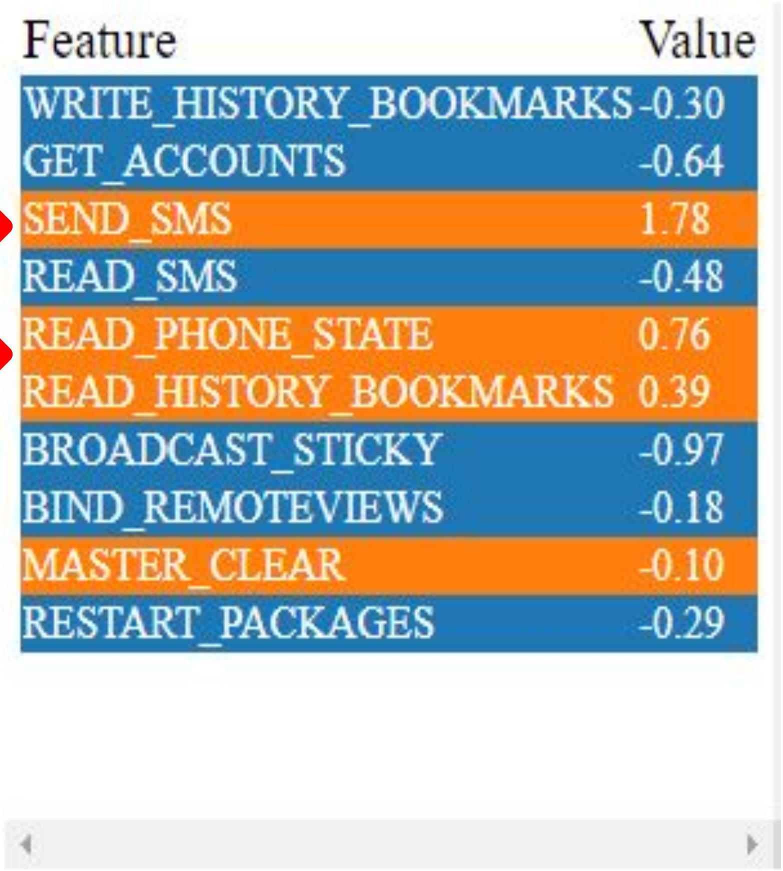


- **Interpretabilidade com SHAP e LIME**

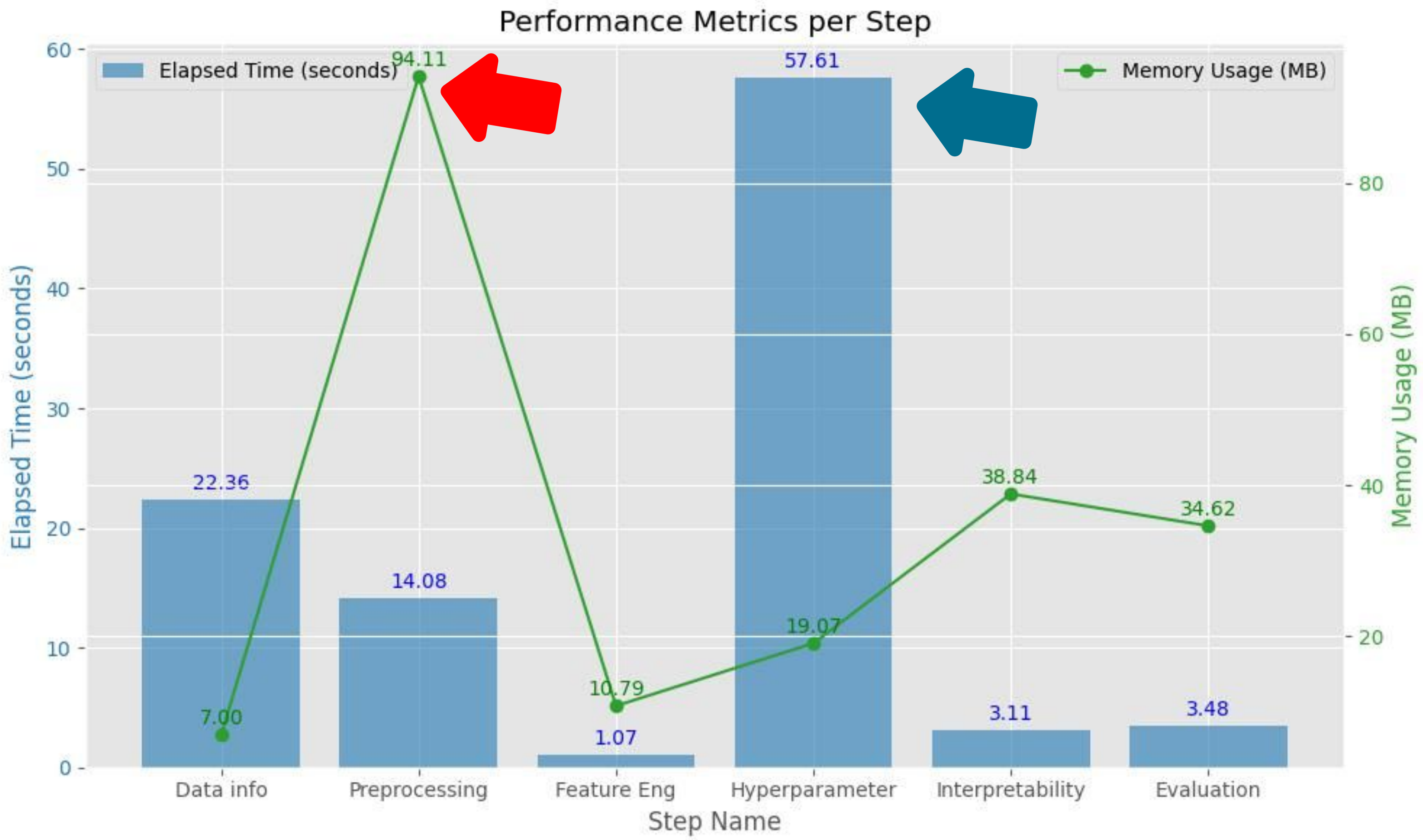

GET_ACCOUNTS
READ_PHONE_STATE
SEND_SMS
READ_EXTERNAL_STORAGE
BROADCAST_STICKY
READ_HISTORY_BOOKMARKS
WRITE_HISTORY_BOOKMARKS
READ_SMS
WAKE_LOCK
SUBSCRIBED_FEEDS_WRITE
READ_CALL_LOG
USE_CREDENTIALS
UPDATE_DEVICE_STATS
CAMERA
WRITE_SMS
RESTART_PACKAGES
READ_SYNC_STATS
RECORD_AUDIO
INSTALL_PACKAGES
RECEIVE_SMS
−4
−2
0
2
4
SHAP value (impact on model output)
High
Low
Feature value


GET_ACCOUNTS
READ_PHONE_STATE
SEND_SMS
READ_EXTERNAL_STORAGE
BROADCAST_STICKY
READ_HISTORY_BOOKMARKS

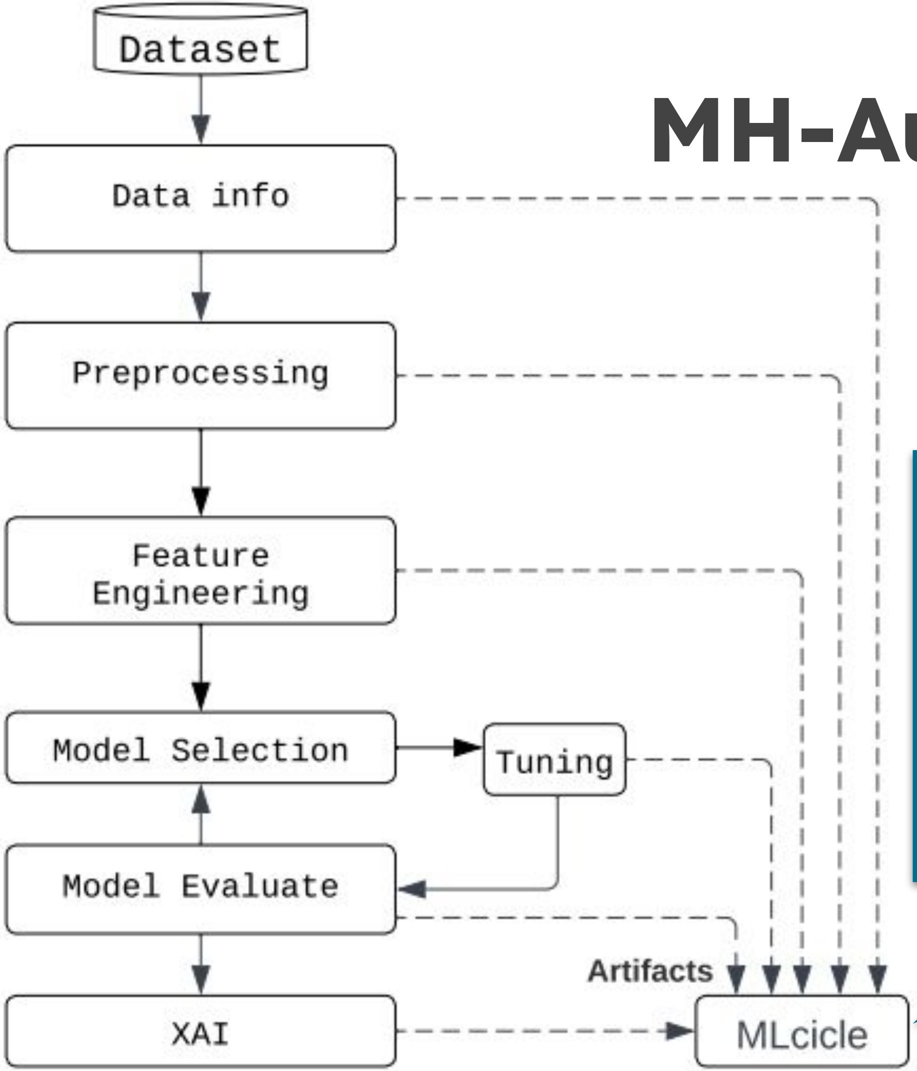




Feature	Value
WRITE_HISTORY_BOOKMARKS	-0.30
GET_ACCOUNTS	-0.64
SEND_SMS	1.78
READ_SMS	-0.48
READ_PHONE_STATE	0.76
READ_HISTORY_BOOKMARKS	0.39
BROADCAST_STICKY	-0.97
BIND_REMOTEVIEWS	-0.18
MASTER_CLEAR	-0.10
RESTART_PACKAGES	-0.29

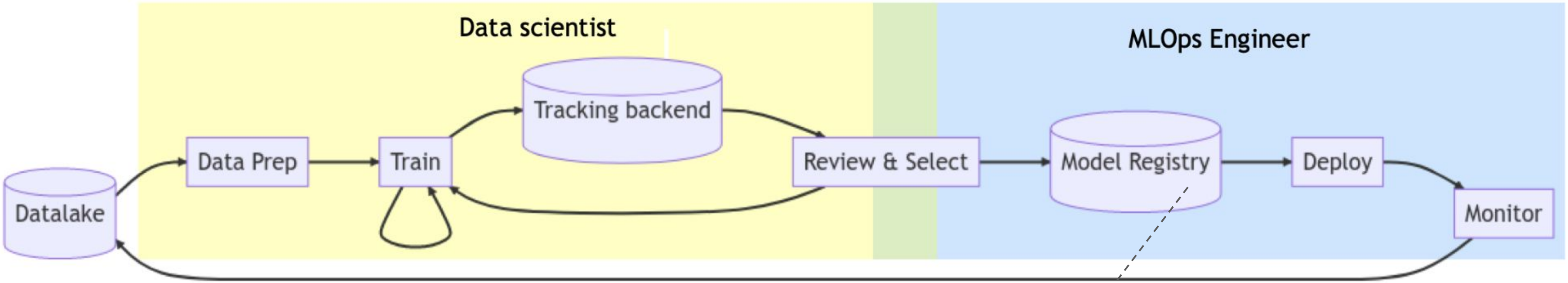

Performance Metrics per Step
Elapsed Time (seconds)
Memory Usage (MB)
Elapsed Time (seconds)
Memory Usage (MB)
Step Name
0
10
20
30
40
50
60
20
40
60
80
Data info
Preprocessing
Feature Eng
Hyperparameter
Interpretability
Evaluation
22.36
14.08
1.07
57.61
3.11
3.48
7.00
94.11
10.79
19.07
38.84
34.62

MH-AutoML

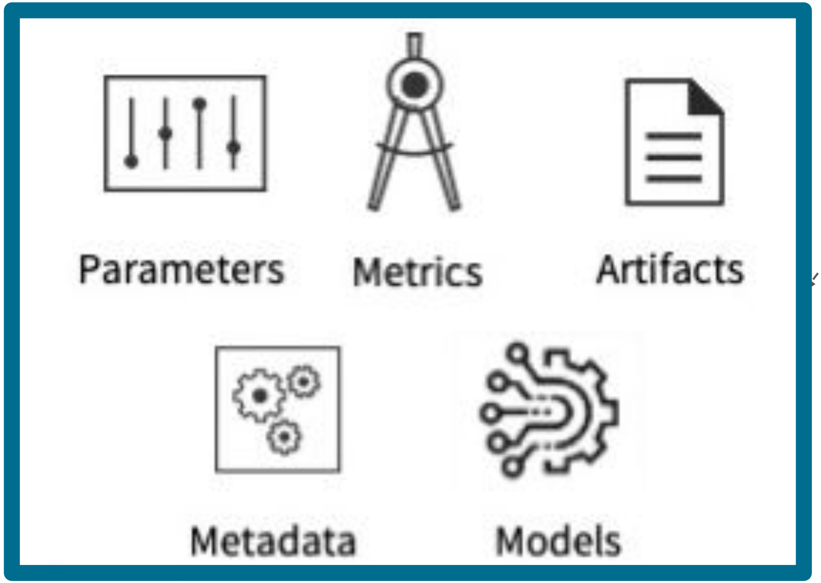


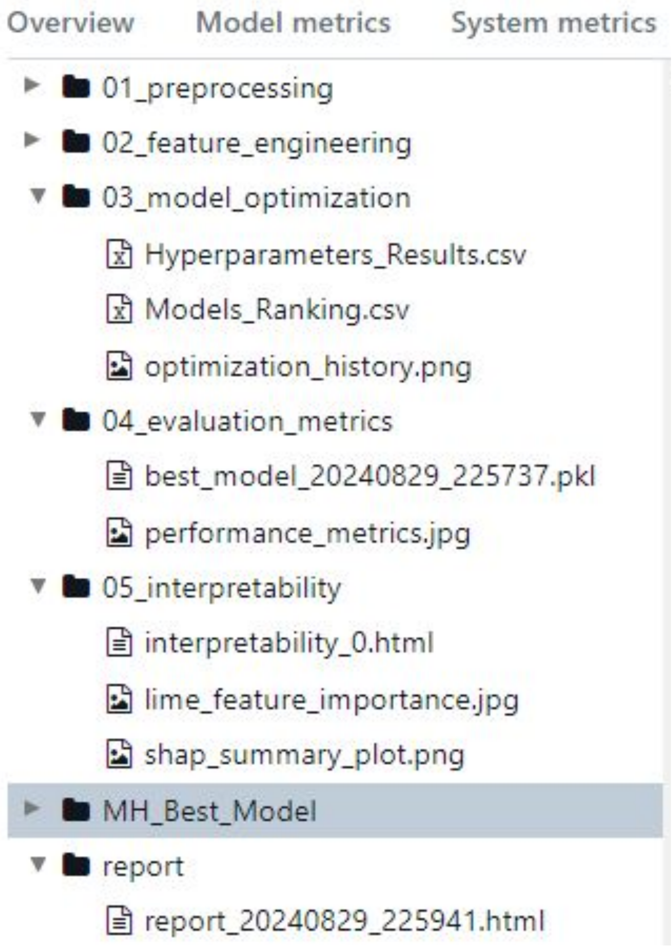
- Rastreamento de experimentos
- Registro de artefatos (códigos, imagens,datasets)
- Versionamento de modelos

MLflow Tracking



mlflow.org





```
▶ 01_preprocessing
▶ 02_feature_engineering
▼ 03_model_optimization
    Hyperparameters_Results.csv
    Models_Ranking.csv
    optimization_history.png
▼ 04_evaluation_metrics
    best_model_20240829_225737.pkl
    performance_metrics.jpg
▼ 05_interpretability
    interpretability_0.html
    lime_feature_importance.jpg
    shap_summary_plot.png
▶ MH_Best_Model
▼ report
    report_20240829_225941.html
```

- **Arquivos de dados (CSV)**
- **Arquivos de imagens(PNG...)**
- **Arquivos HTML**
- **Arquivos PKL**

Avaliação

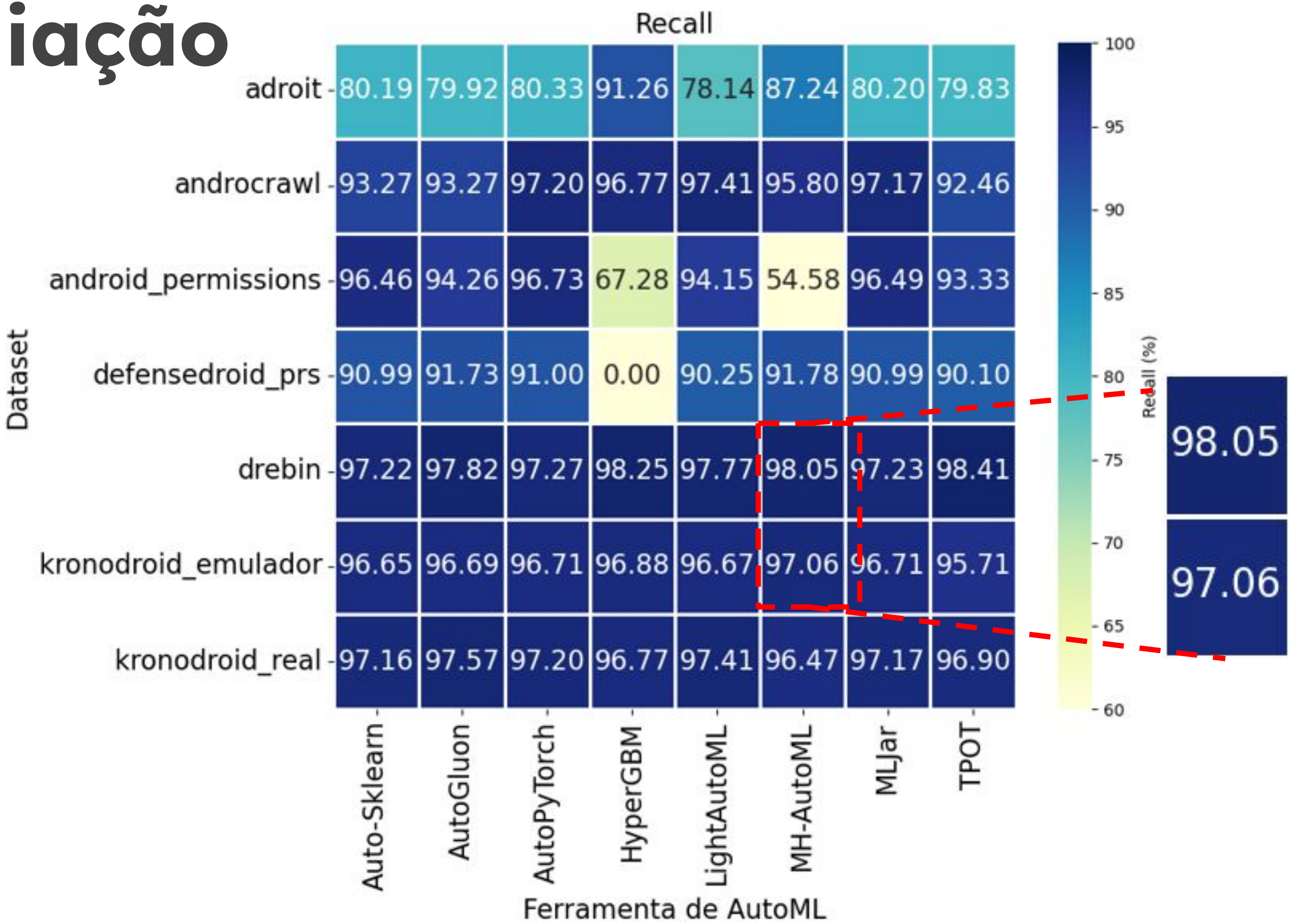
Dataset	Qntd.	Tipos	Total
adroit	166	P	11476
androcrawl	141	A(26), I(8), P(84), O(23)	96744
android_permissions	151	P	26864
defensedroid_prs	2877	P(1489), I(1388)	11975
drebin	215	A(73), P(113), S(6), I(23)	15031
kronodroid_emulador	276	P(145), A(123), O(8)	63991
kronodroid_real	286	P(146), A(100), O(40)	78137

[P] Permissões, [A] Chamadas de API, [I] Intenções, [S] Comandos do Sistema, [O] Outros

Avaliação

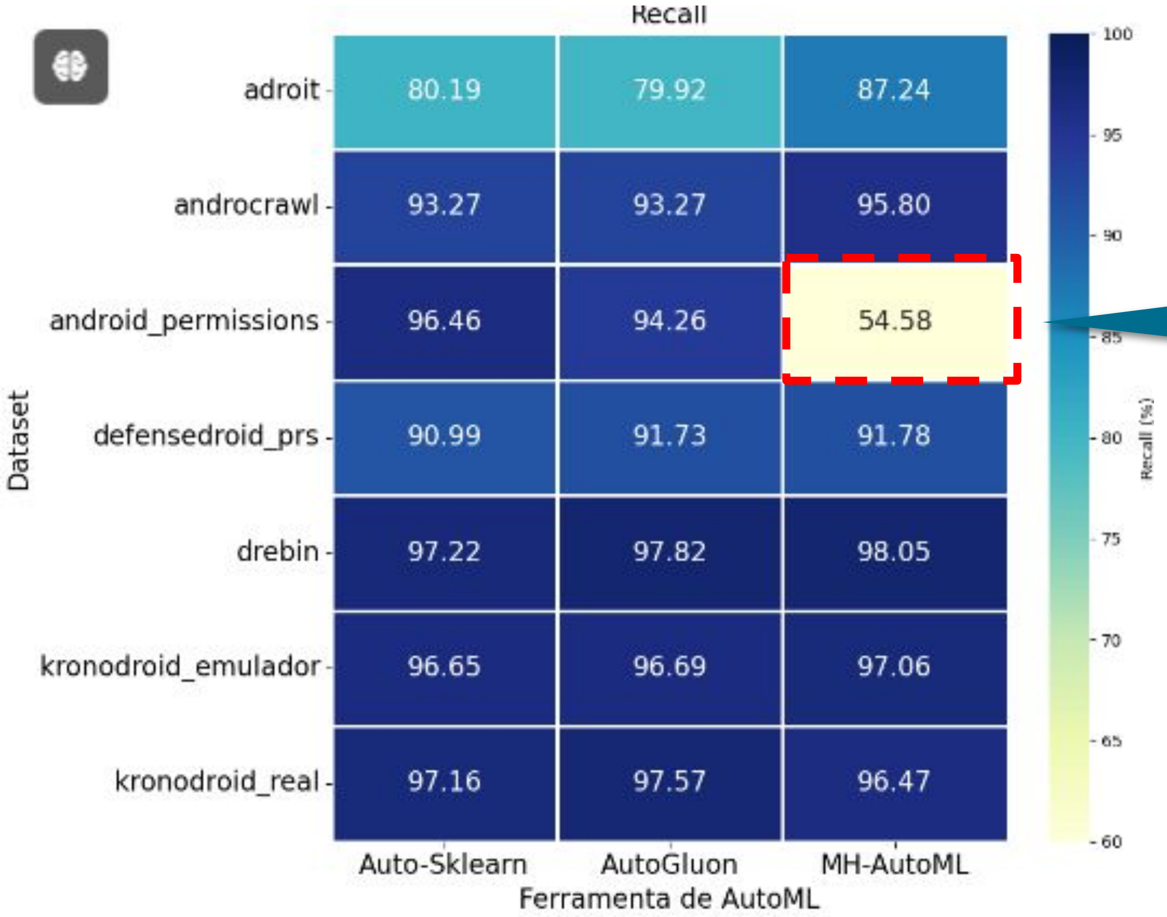
Ferramentas	Citação Google Scholar
Auto-Sklearn	2781
AutoGluon	682
TPOT	405
Lightautoml	38
Mljar	35
Auto-pytorch	5
HyperGBM	1

Avaliação



Dataset	Auto-Sklearn	AutoGluon	AutoPyTorch	HyperGBM	LightAutoML	MH-AutoML	MLJar	TPOT
adroit	80.19	79.92	80.33	91.26	78.14	87.24	80.20	79.83
androcrawl	93.27	93.27	97.20	96.77	97.41	95.80	97.17	92.46
android_permissions	96.46	94.26	96.73	67.28	94.15	54.58	96.49	93.33
defensedroid_prs	90.99	91.73	91.00	0.00	90.25	91.78	90.99	90.10
drebin	97.22	97.82	97.27	98.25	97.77	98.05	97.23	98.41
kronodroid_emulador	96.65	96.69	96.71	96.88	96.67	97.06	96.71	95.71
kronodroid_real	97.16	97.57	97.20	96.77	97.41	96.47	97.17	96.90

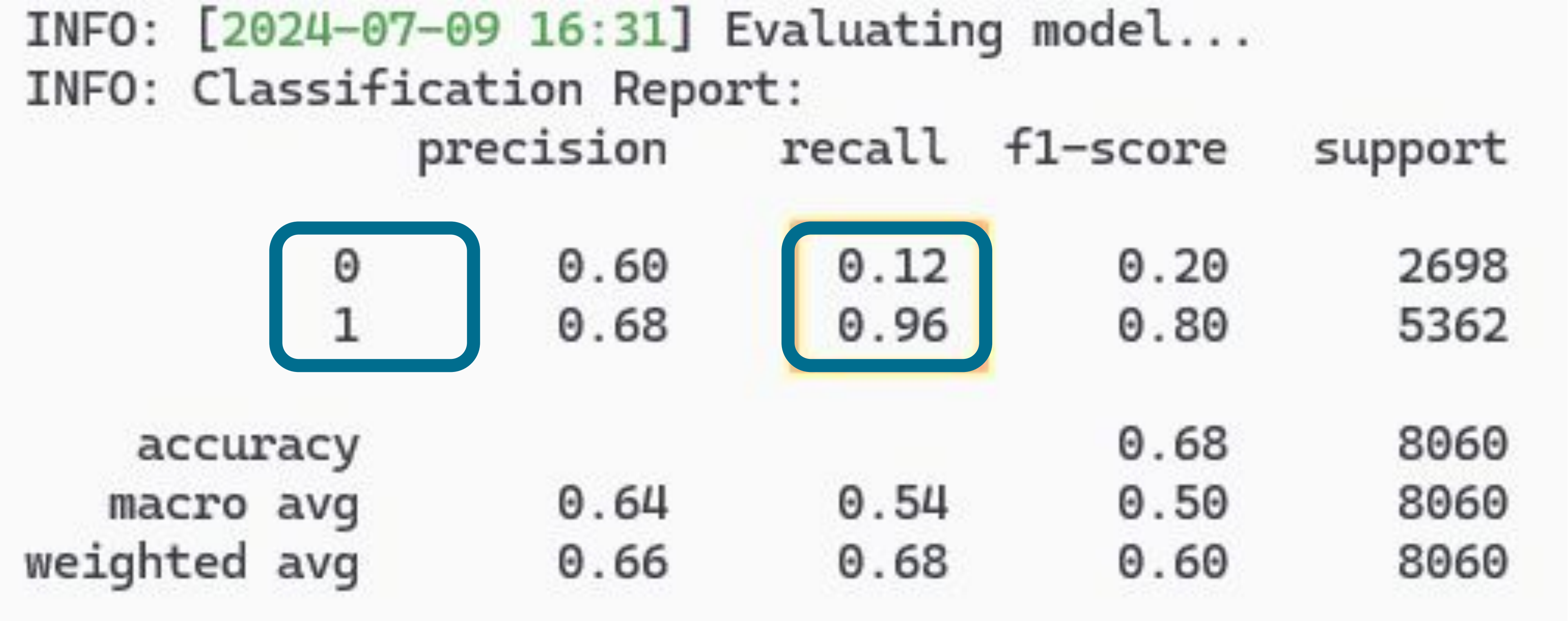
Avaliação



Dataset	Auto-Sklearn	AutoGluon	MH-AutoML
adroit	80.19	79.92	87.24
androcrawl	93.27	93.27	95.80
android_permissions	96.46	94.26	54.58
defensedroid_prs	90.99	91.73	91.78
drebin	97.22	97.82	98.05
kronodroid_emulador	96.65	96.69	97.06
kronodroid_real	97.16	97.57	96.47

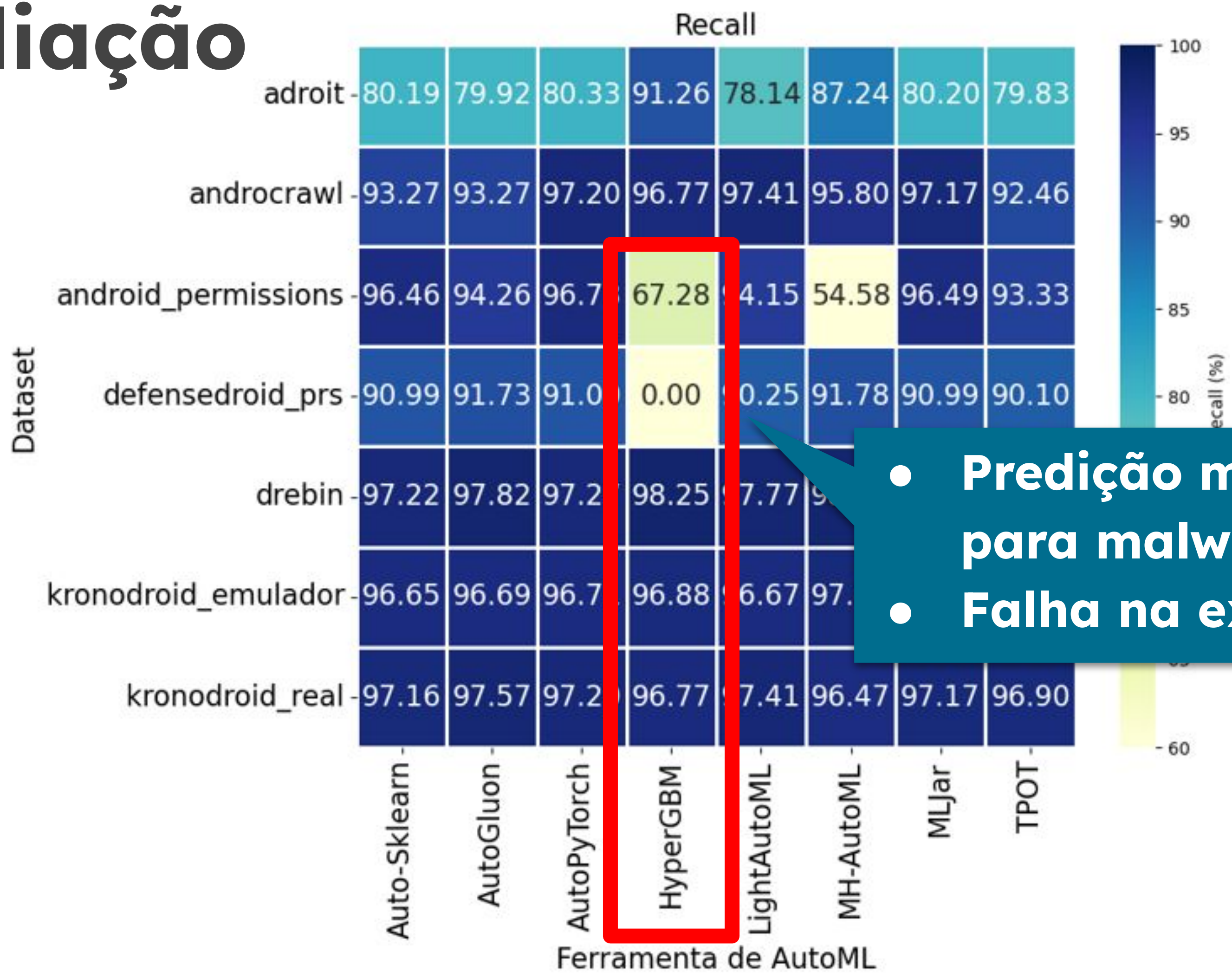
- **Benignos 12%**
- **Malwares 96%**

Avaliação



```
INFO: [2024-07-09 16:31] Evaluating model...
INFO: Classification Report:
              precision    recall  f1-score   support

           0       0.60      0.12      0.20      2698
           1       0.68      0.96      0.80      5362

    accuracy                           0.68      8060
   macro avg       0.64      0.54      0.50      8060
weighted avg       0.66      0.68      0.60      8060
```

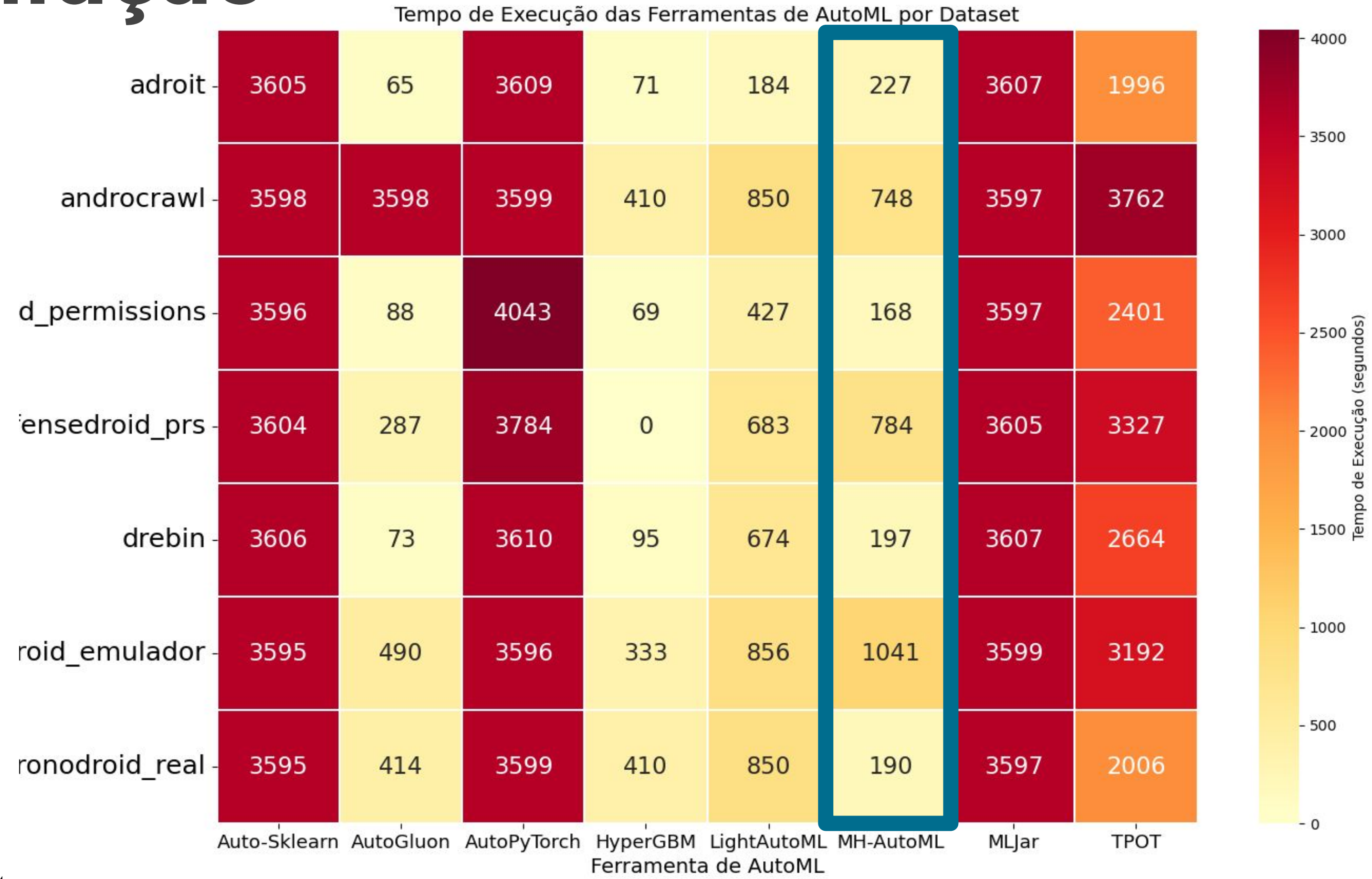
Avaliação



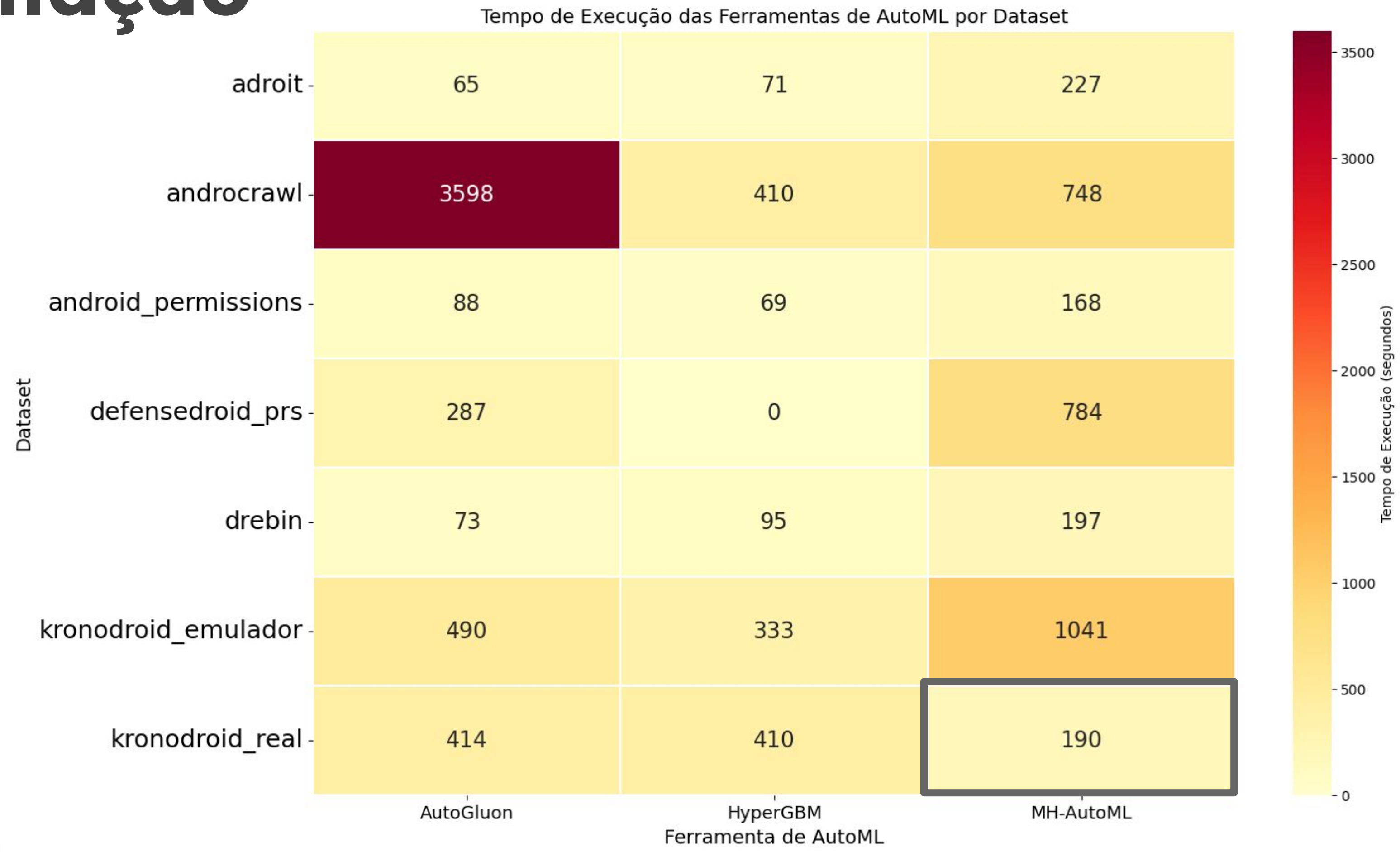
	Auto-Sklearn	AutoGluon	AutoPyTorch	HyperGBM	LightAutoML	MH-AutoML	MLJar	TPOT
adroit	80.19	79.92	80.33	91.26	78.14	87.24	80.20	79.83
androcrawl	93.27	93.27	97.20	96.77	97.41	95.80	97.17	92.46
android_permissions	96.46	94.26	96.7	67.28	4.15	54.58	96.49	93.33
defensedroid_prs	90.99	91.73	91.0	0.00	0.25	91.78	90.99	90.10
drebin	97.22	97.82	97.2	98.25	7.77	9	[illegible]	[illegible]
kronodroid_emulador	96.65	96.69	96.7	96.88	6.67	97.	[illegible]	[illegible]
kronodroid_real	97.16	97.57	97.2	96.77	7.41	96.47	97.17	96.90

- Predição maior para malwares
- Falha na execução

Avaliação


Tempo de Execução das Ferramentas de AutoML por Dataset
adroit
3605
65
3609
71
184
227
3607
1996
androcrawl
3598
3598
3599
410
850
748
3597
3762
d_permissions
3596
88
4043
69
427
168
3597
2401
ensedroid_prs
3604
287
3784
0
683
784
3605
3327
drebin
3606
73
3610
95
674
197
3607
2664
roid_emulador
3595
490
3596
333
856
1041
3599
3192
ronodroid_real
3595
414
3599
410
850
190
3597
2006
Auto-Sklearn
AutoGluon
AutoPyTorch
HyperGBM
LightAutoML
MH-AutoML
MLJar
TPOT
Ferramenta de AutoML
Tempo de Execução (segundos)
4000
3500
3000
2500
2000
1500
1000
500
0

Avaliação


Tempo de Execução das Ferramentas de AutoML por Dataset
Dataset
adroit
androcrawl
android_permissions
defensedroid_prs
drebin
kronodroid_emulador
kronodroid_real
65
71
227
3598
410
748
88
69
168
287
0
784
73
95
197
490
333
1041
414
410
190
AutoGluon
HyperGBM
MH-AutoML
Ferramenta de AutoML
Tempo de Execução (segundos)
3500
3000
2500
2000
1500
1000
500
0

Demostração

Considerações finais

- Bom desempenho
- Rastreável
- Versionavel
- Transparente
- Interpretável

Trabalhos futuros

- Melhorias de desempenho
- Disponibilizar como serviço web
- Explorar novas tecnicas de explicabilidade dos modelos

Obrigado!

MOTOROLA

UNIVERSIDADE FEDERAL DO AMAZONAS
IN UNIVERSA SCIENTIA VERITAS
UFAM

unipampa
Universidade Federal do Pampa

CAPES

iComp
Instituto de Computação

CNPq

FAPERGS