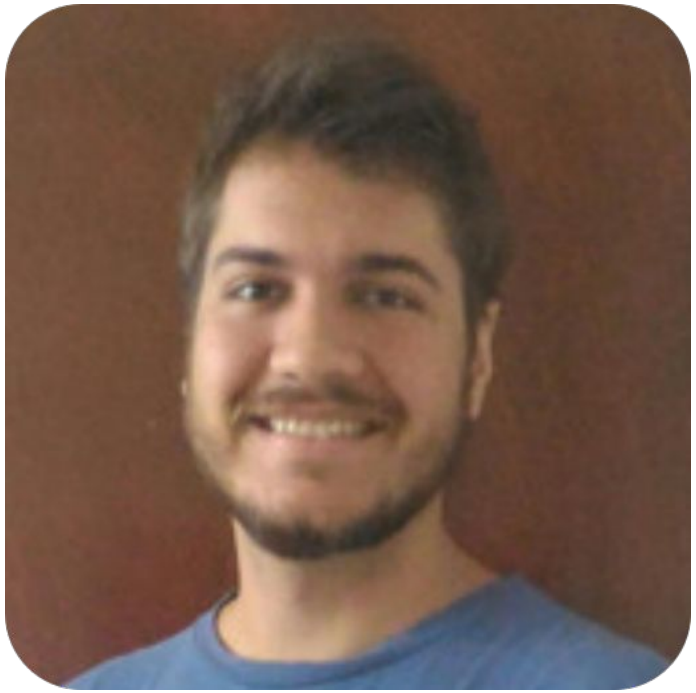
Gerenciando de Modelos de Machine Learning com MLflow



Arthur Lacerda

Graduado em Ciência da Computação - UFG

Pós Graduando em Machine Learning e Big Data - FASAM

Desenvolvedor Full-stack e Cientista de Dados - CQuantt

Machine Learning

Extração de padrões nos dados para que esses padrões possam ser utilizados na resolução de problemas.

Machine Learning

Classificação

- Detecção de Fraude
- Reconhecimento Facial

Regressão

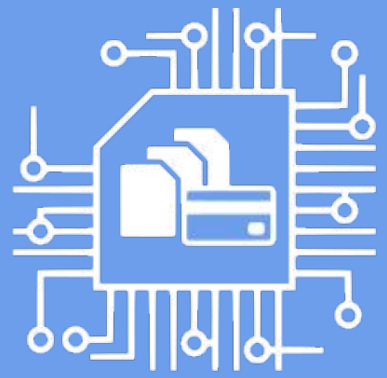
- Previsão de Demanda
- Estimativa de Aceitação

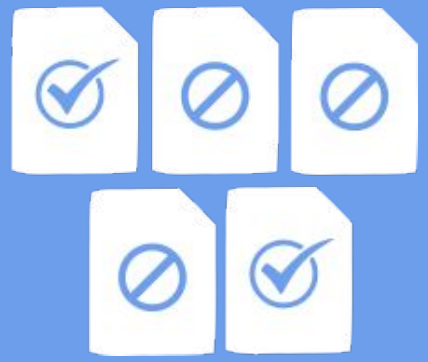
Clusterização

- Segmentação de Cliente
- Modelagem de Tópico

Machine Learning

Dados → ? → Machine Learning → ? → $$$





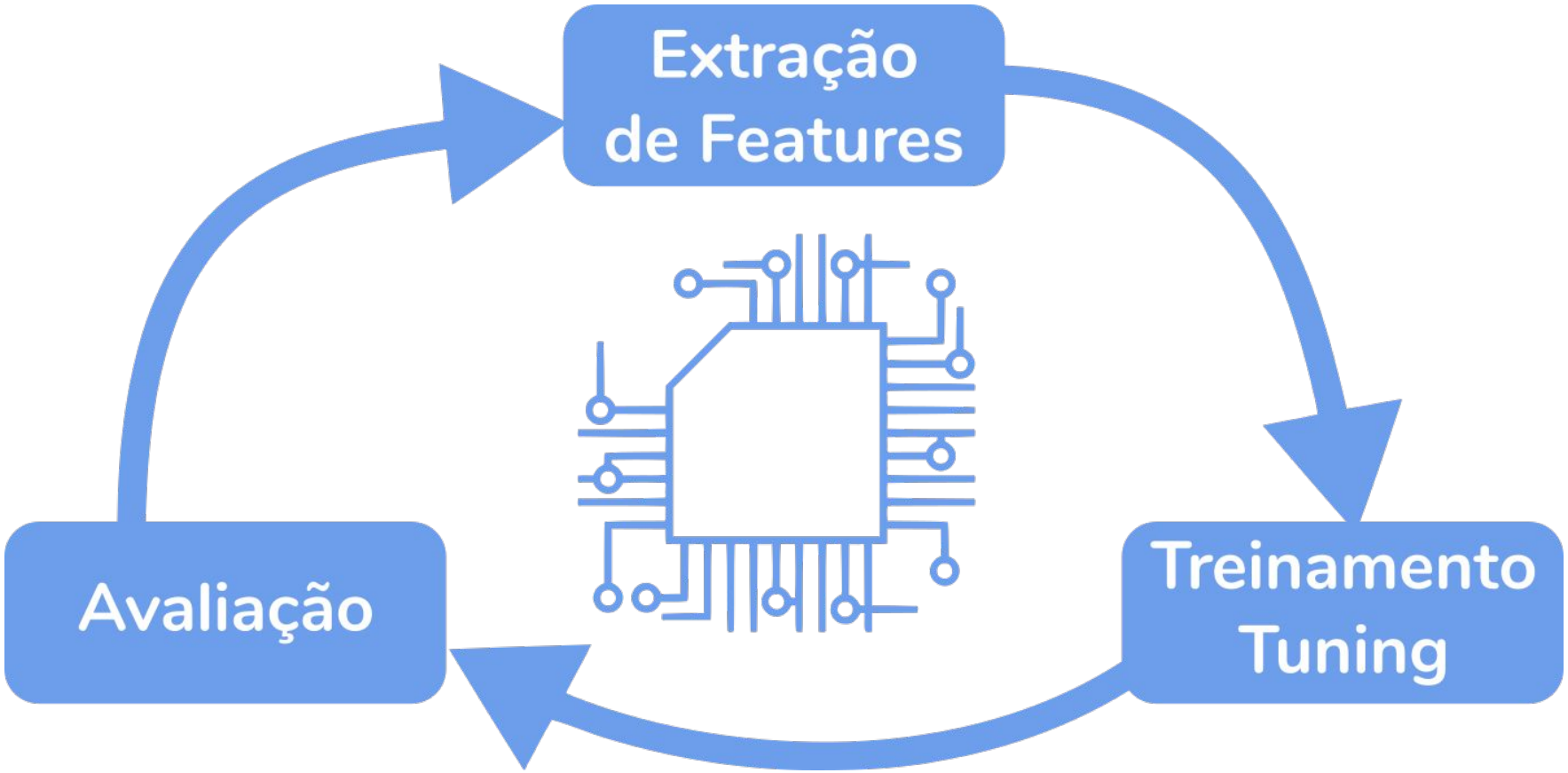
Pipeline de ML

Pipeline de ML

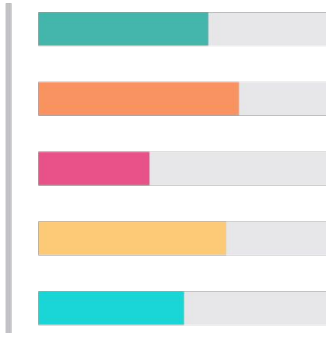
- Visualizações
- Transformação & Extração de Features
- Compartilhamento de Análises

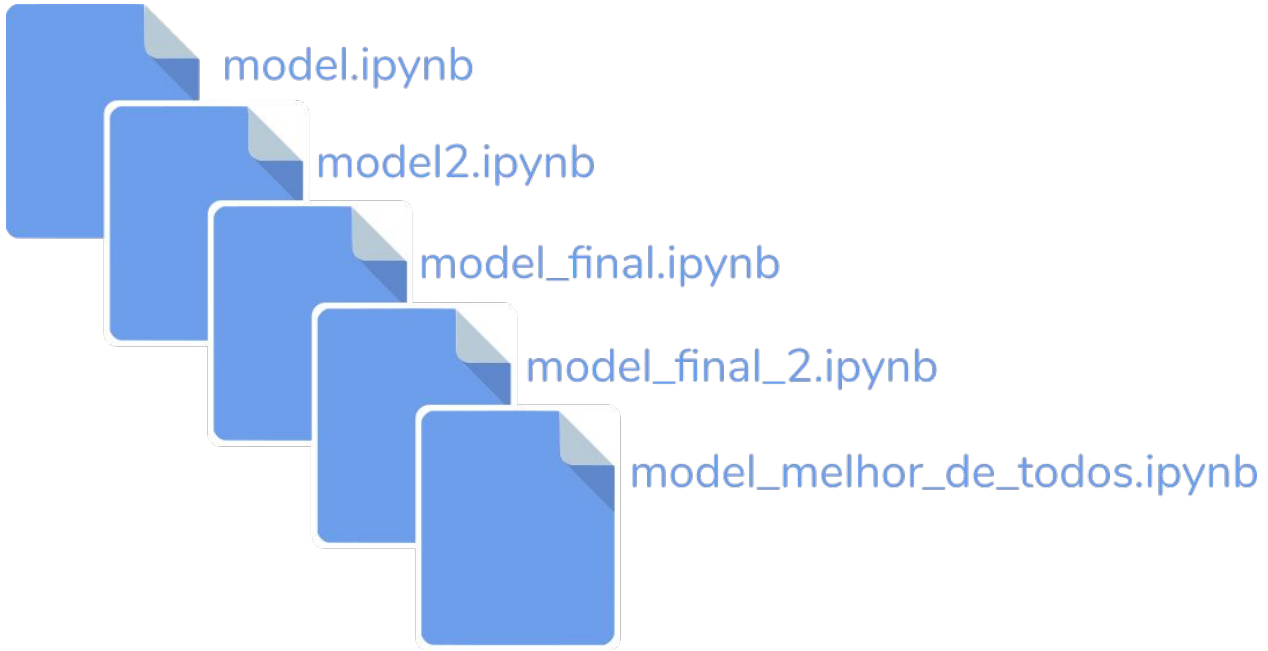
- Escolha do Melhor modelo
- Tunning
- Avaliação
- Versionamento

Pipeline de ML

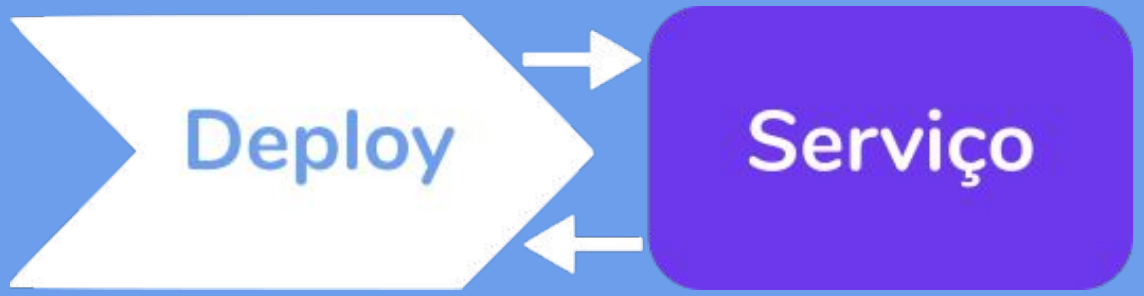


Pipeline de ML





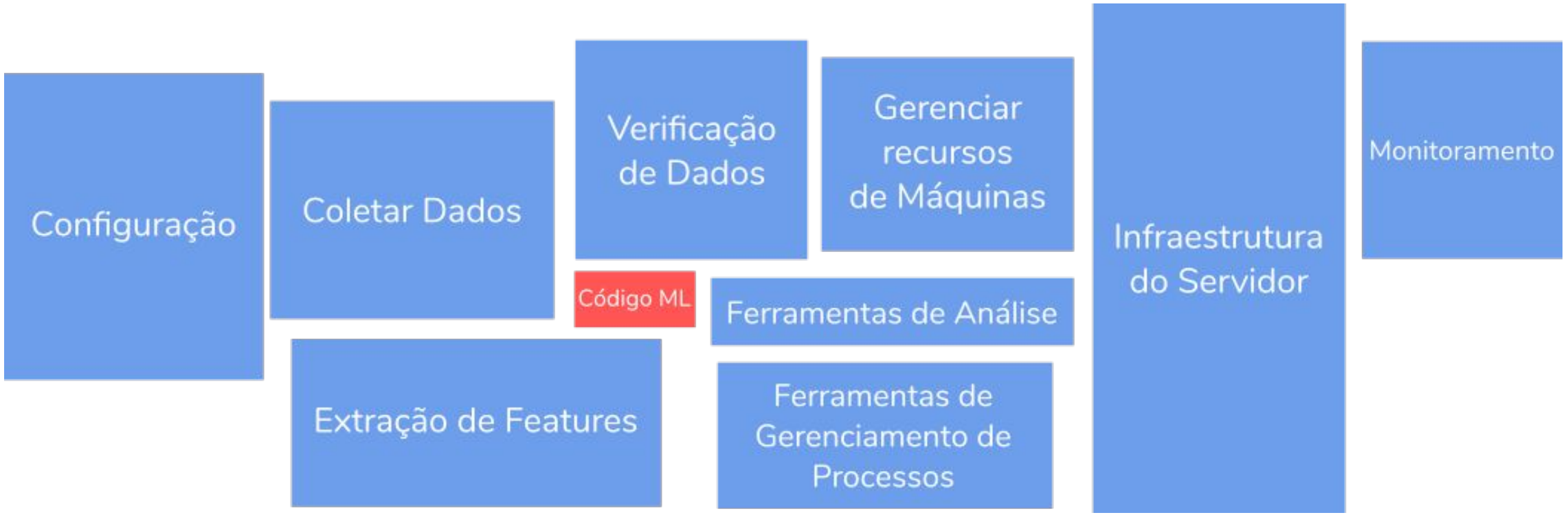
Pipeline de ML



- Pipelines
- Infraestrutura
- Escalabilidade

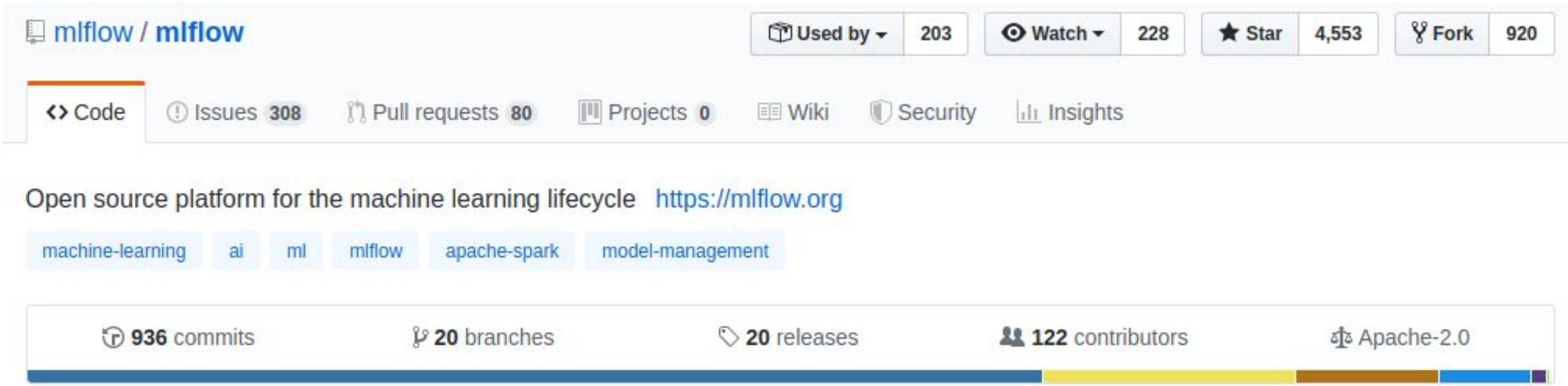
- Monitoramento
- Avaliação Online
- Predição de novos dados

Pipeline de ML




mlflow
TM

mlflow


mlflow / mlflow
Used by 203
Watch 228
Star 4,553
Fork 920
Code
Issues 308
Pull requests 80
Projects 0
Wiki
Security
Insights
Open source platform for the machine learning lifecycle https://mlflow.org
machine-learning
ai
ml
mlflow
apache-spark
model-management
936 commits
20 branches
20 releases
122 contributors
Apache-2.0

TensorFlow

PyTorch

Keras

Spark

scikit learn

H_2O.ai

python

R

Java

mleap

CONDA

docker

Amazon SageMaker

Azure Machine Learning

Google Cloud

mlflow


databricks


Microsoft


RStudio


UNIVERSITY of WASHINGTON


SK telecom


Outreach


splice
MACHINE


Telia Company


TEKION


Bering


SCRY ANALYTICS


cnrs


WiX.com


ABRAXAS
Intelligence


TIGER SHARK


BRANDLESS


virtusa


Health 2 Sync


mined.ai


dL

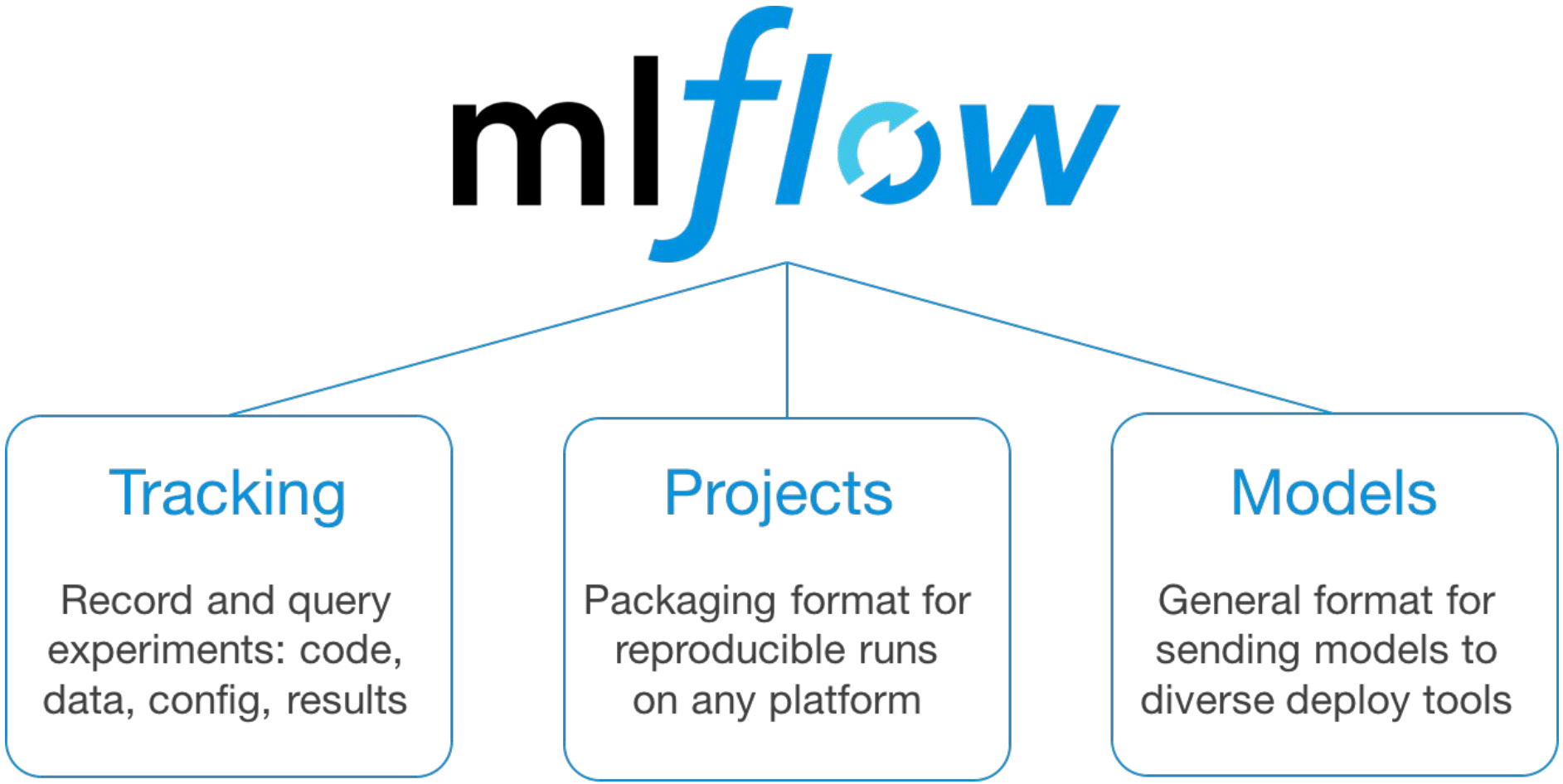

SYSTEMATIC


despegar.com


ModelWorks

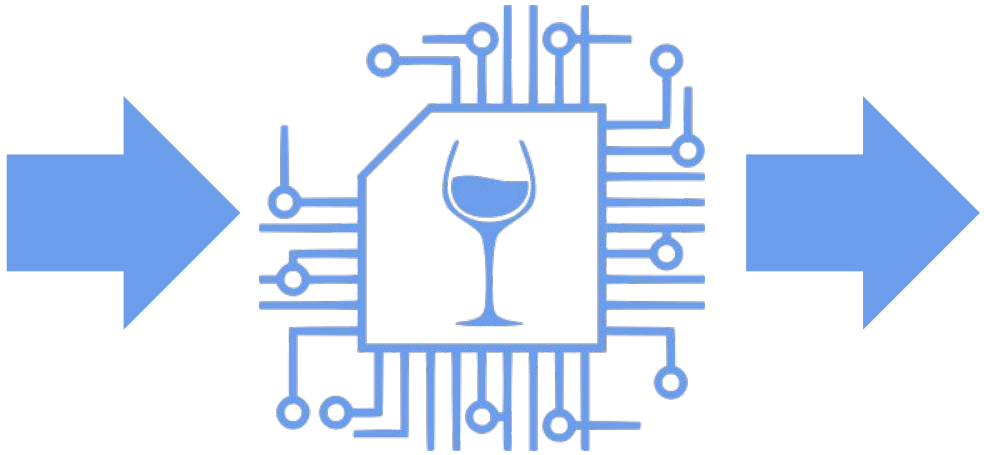

Brandfolder


ASML


mlflow
Tracking
Record and query experiments: code, data, config, results
Projects
Packaging format for reproducible runs on any platform
Models
General format for sending models to diverse deploy tools

Problema

Acidez fixa:	7
Acidez volátil:	0.27
Ácido cítrico:	0.36
Açúcar residual:	20.7
Cloretos:	0.045
Dióxido de Enxofre Livre:	45
Dióxido de Enxofre Total:	170
Densidade:	1.001
pH:	3
Sulfatos:	0.45
Álcool:	8.8



Qualidade
0 a 10

Tracking

mlflow

Tracking

Run Sources

mlflow Github Docs

Listing Price Prediction

Experiment ID: 0 Artifact Location: /Users/matei/mlflow/demo/mlruns/0

Search Runs: metrics.R2 > 0.24 Search

Filter Params: alpha, lr Filter Metrics: rmse, r2 Clear

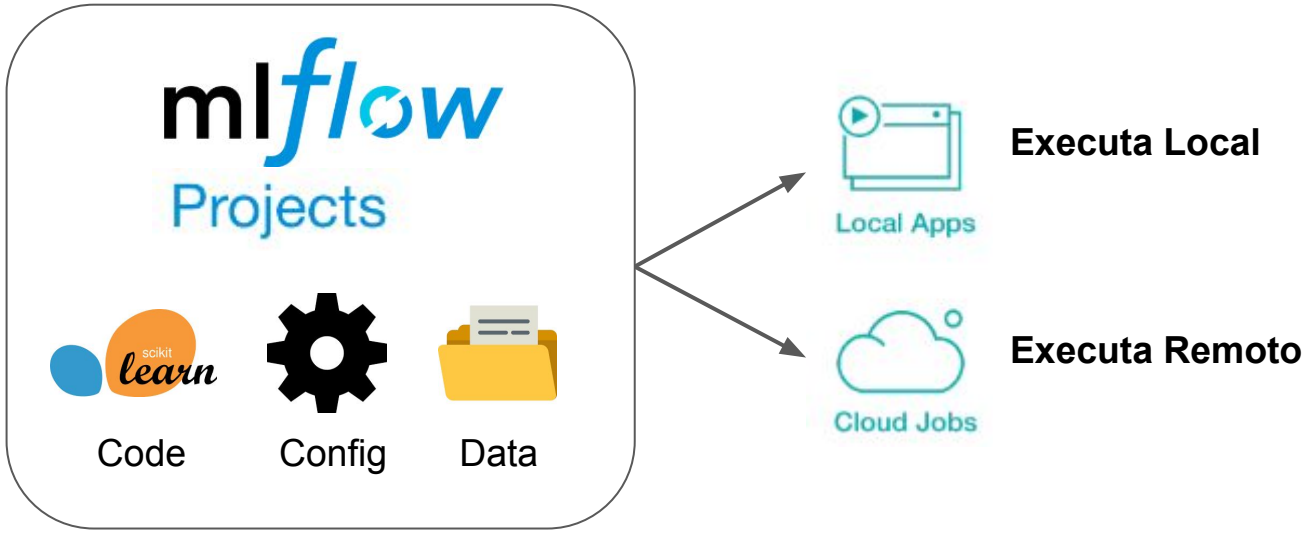
4 matching runs Compare Selected Download CSV

					Parameters		Metrics		
	Time	User	Source	Version	alpha	l1_ratio	MAE	R2	RMSE
☐	17:37	matei	linear.py	3a1995	0.5	0.2	84.27	0.277	158.1
☐	17:37	matei	linear.py	3a1995	0.2	0.5	84.08	0.264	159.6
☐	17:37	matei	linear.py	3a1995	0.5	0.5	84.12	0.272	158.6
☐	17:37	matei	linear.py	3a1995	0	0	84.49	0.249	161.2

```
# tracking
with mlflow.start_run():

    mlflow.log_artifacts("artifacts/")

    mlflow.log_param("alpha", alpha)
    mlflow.log_param("l1_ratio", l1_ratio)

    mlflow.log_metric("RMSE", rmse)
    mlflow.log_metric("MAE", mae)
    mlflow.log_metric("R2", r2)

    mlflow.sklearn.log_model(model, "model")
```

Projects

mlflow Projects

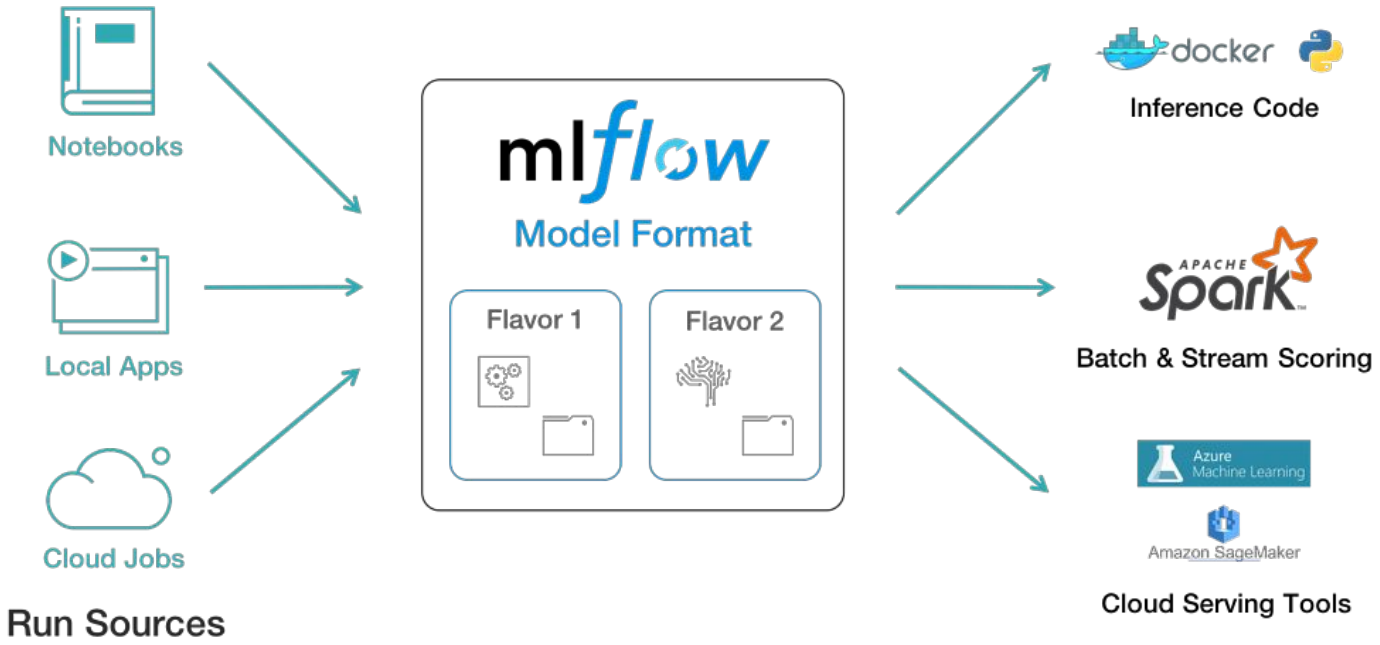


```
name: cp-wine-quality
conda_env: conda.yaml

entry_points:
  main:
    parameters:
      alpha: {type: float, default: 1.0}
      l1_ratio: {type: float, default: 0.1}
    command: "python elastic-train.py {alpha} {l1_ratio}"
  huber:
    parameters:
      alpha: {type: float, default: 0.0001}
      epsilon: {type: float, default: 1.35}
      max_iter: {type: int, default: 100}
    command: "python huber-train.py {alpha} {epsilon}
{max_iter}"
  classifier:
    command: "python clf-train.py"
```

Models

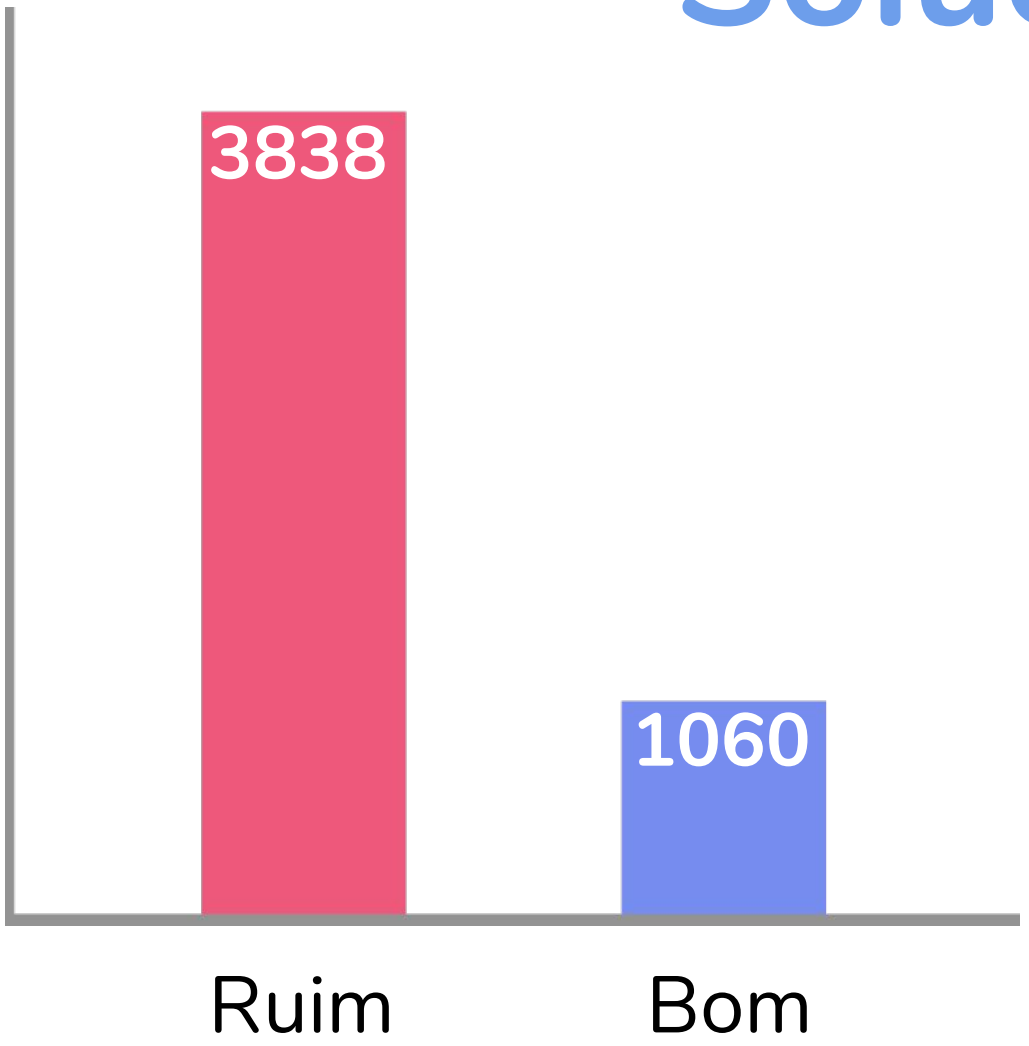
mlflow

Projects



Hands-On

Solução



Pré filtragem que elimina aproximadamente **78%** dos vinhos que chegam para análise, com uma acurácia de **88%** que o identifica como **Ruim**.

DataHackers

datahackers.com.br

Obrigado!

github: /arthurulacerda
LinkedIn: /arthurlacerda
arthurlacerda.ds@gmail.com