

SICO7A

SISTEMAS INTELIGENTES 1

Aula 04 A - Introdução ao
Aprendizado de Máquina

Prof. Rafael G. Mantovani



Apucarana - PR, Brasil

Universidade Tecnológica Federal do Paraná (UTFPR)
Engenharia de Computação

Roteiro

- 1 Introdução**
- 2 Dados**
- 3 Tipos de ML**
- 4 Pipeline**
- 5 Ferramentas de ML**
- 6 Referências**

Roteiro

1 Introdução

2 Dados

3 Tipos de ML

4 Pipeline

5 Ferramentas de ML

6 Referências

Introdução

Introdução

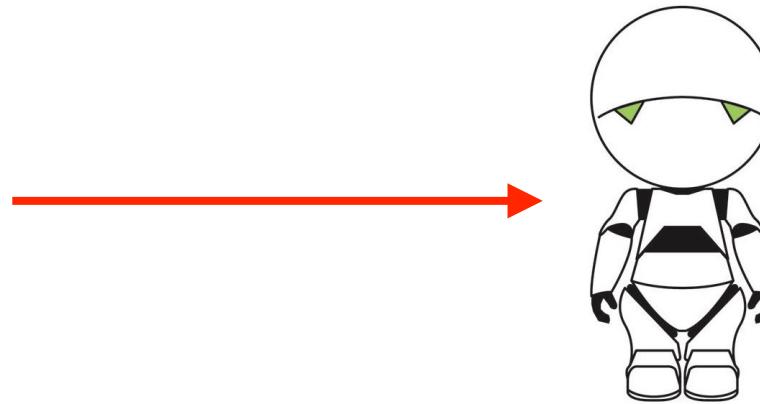


Introdução



Site que vende produtos

Introdução

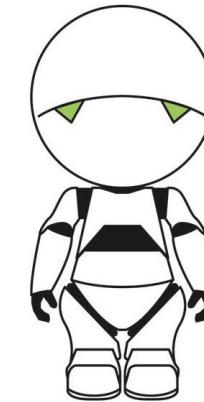


Cliente

Introdução



- **Dados:**
 - tipo de computador
 - browser
 - país
 - tempo
 - software/produto
 - forma de pagamento
 - ...



Cliente

Introdução

Cliente 1



**Mac OS, Safari, Inglaterra, manhã, Super Game 1,
cartão de crédito**

Cliente 2



Windows, Chrome, EUA, tarde, Super Game 1, PayPal

Cliente 3



Linux, Firefox, Brasil, noite, Super Game 2, Boleto

Introdução

Cliente 1



Mac OS, Safari, Inglaterra, manhã, Super Game 1,
cartão de crédito

Cliente 2

**Muitos clientes → Muitos dados
Dados: Previsões
Padrões / Similaridades**

1, PayPal

Cliente 3



Linux, Firefox, Brasil, noite, Super Game 2, Boleto

Introdução



Carros autônomos
(Tesla, ...)



Processamento de textos
(corretores de texto,
tradutores)



Reconhecimento de imagens



Processamento de Sinais
(Alexa, Siri, etc)

Introdução



**Sistemas para
Diagnóstico Médico**
(Clínica geral, Covid, etc)



Bioinformática
(identificar proteínas)



Sistemas de recomendação
(Serviços de Streaming,
e-commerce, Ad-ons, etc)



Detecção de fraudes
(clonagem de cartão,
contas indevidas ...)

Roteiro

1 Introdução

2 Dados

3 Tipos de ML

4 Pipeline

5 Ferramentas de ML

6 Referências

Datasets

Datasets



Datasets



Pessoas



Dados

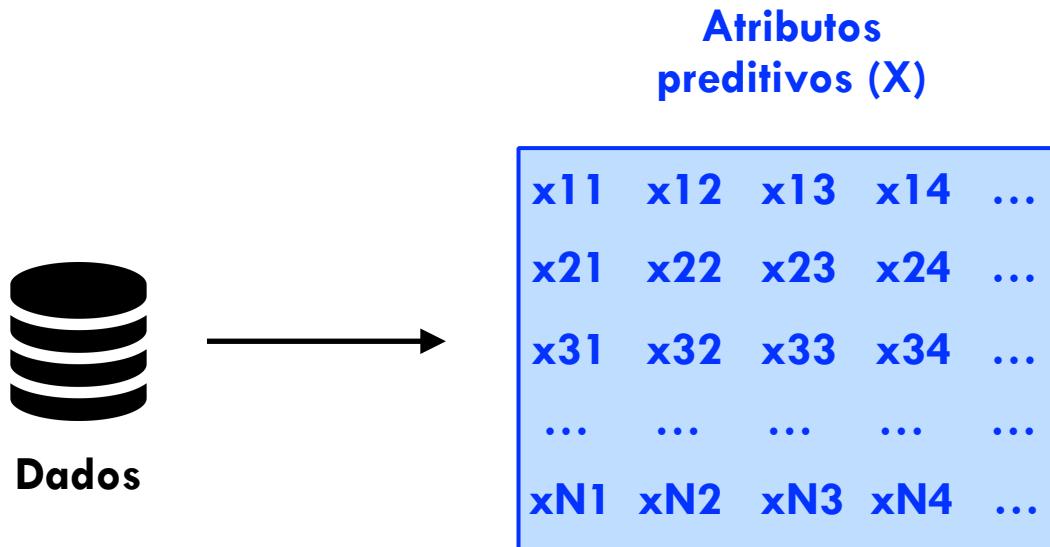
- Dados são **continuamente**: gerados, coletados, processados e transmitidos

Datasets

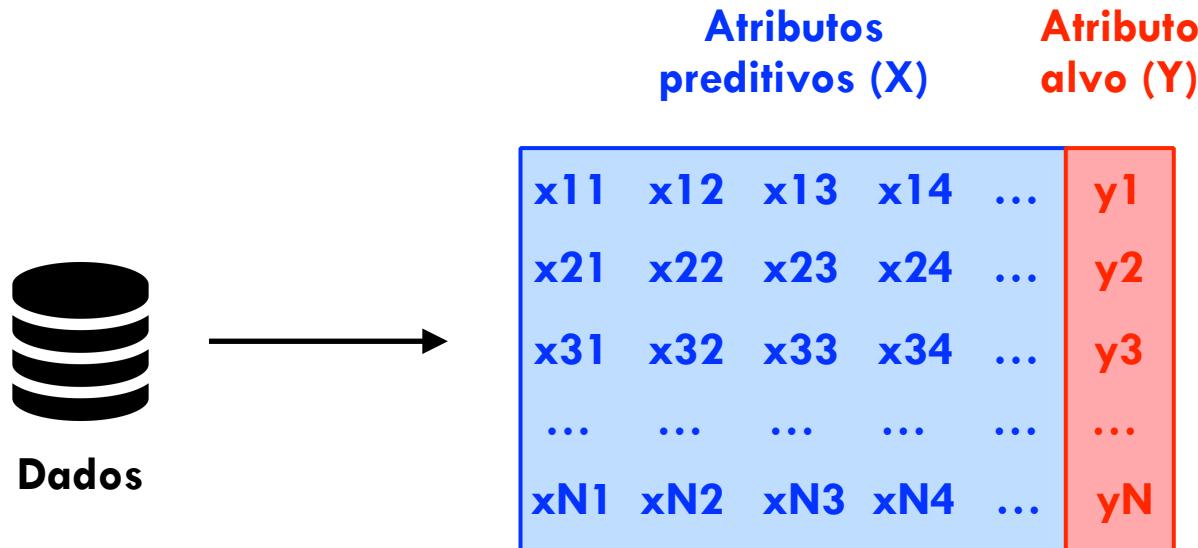


Dados

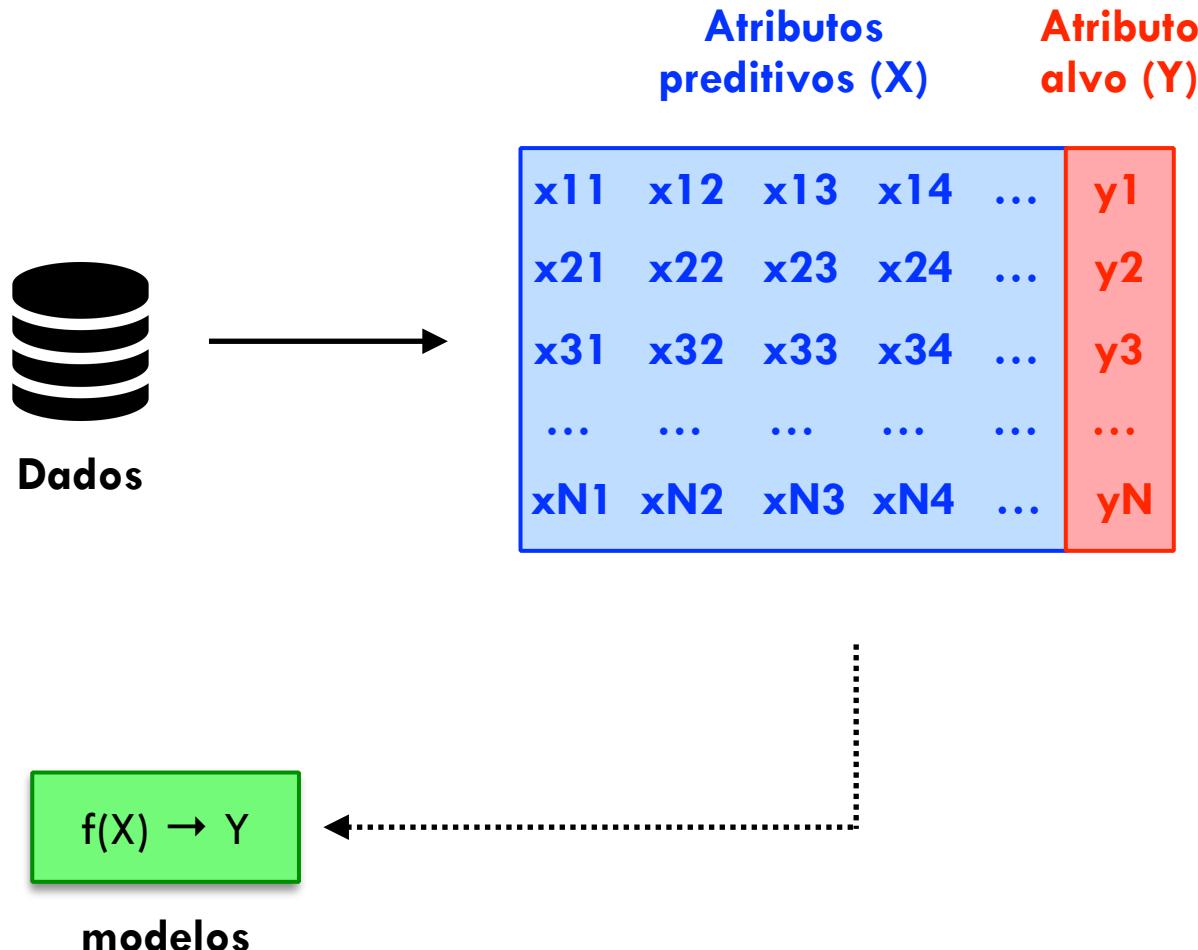
Datasets



Datasets



Datasets



Datasets



Iris

Atributos
preditivos (X)

Atributo
alvo (Y)

	sepallength	sepalwidth	petallength	petalwidth	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa
10	5.4	3.7	1.5	0.2	Iris-setosa
11	4.8	3.4	1.6	0.2	Iris-setosa
12	4.8	3.0	1.4	0.1	Iris-setosa
13	4.3	3.0	1.1	0.1	Iris-setosa
14	5.8	4.0	1.2	0.2	Iris-setosa
15	5.7	4.4	1.5	0.4	Iris-setosa

Showing 1 to 16 of 150 entries

Datasets

x1	x2	Classe
0.1	1	1
0.15	0.2	2
0.48	0.6	3
0.1	0.6	1
0.2	0.15	2
0.5	0.55	3
0.2	1	1
0.3	0.25	2
0.52	0.6	3
0.3	0.6	1
0.4	0.2	2
0.52	0.5	3

Datasets

x1	x2	Classe
0.1	1	1
0.15	0.2	2
0.48	0.6	3
0.1	0.6	1
0.2	0.15	2
0.5	0.55	3
0.2	1	1
0.3	0.25	2
0.52	0.6	3
0.3	0.6	1
0.4	0.2	2
0.52	0.5	3



?

Roteiro

1 Introdução

2 Dados

3 Tipos de ML

4 Pipeline

5 Ferramentas de ML

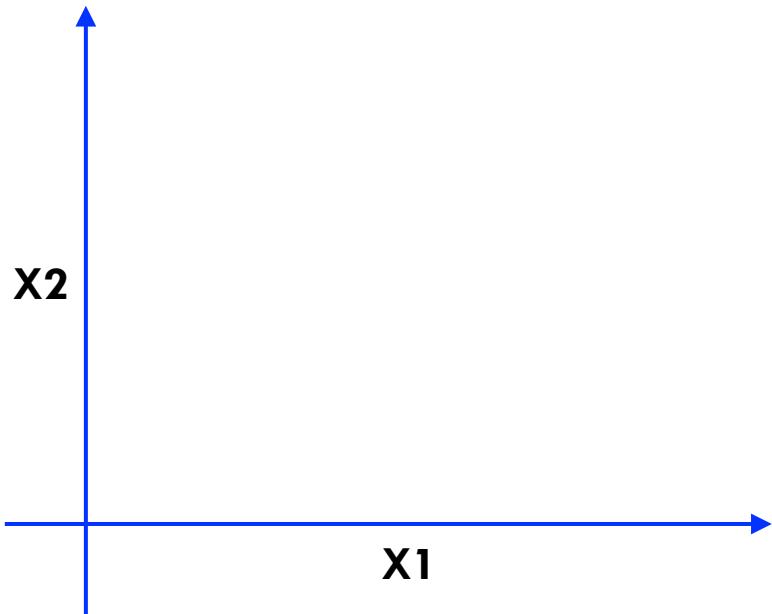
6 Referências

Tipos de ML

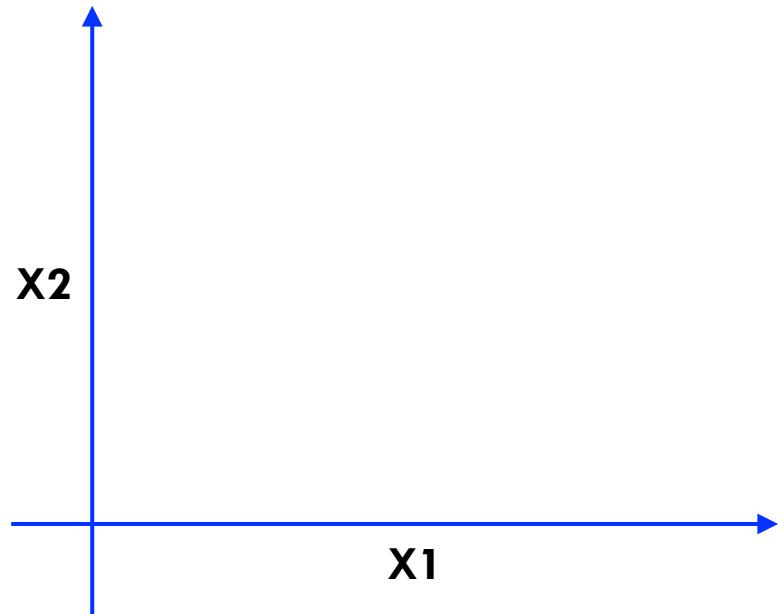
Tarefa descritiva
(Clustering)

Tarefa preditiva
**(Classificação/
Regressão)**

Tipos de ML

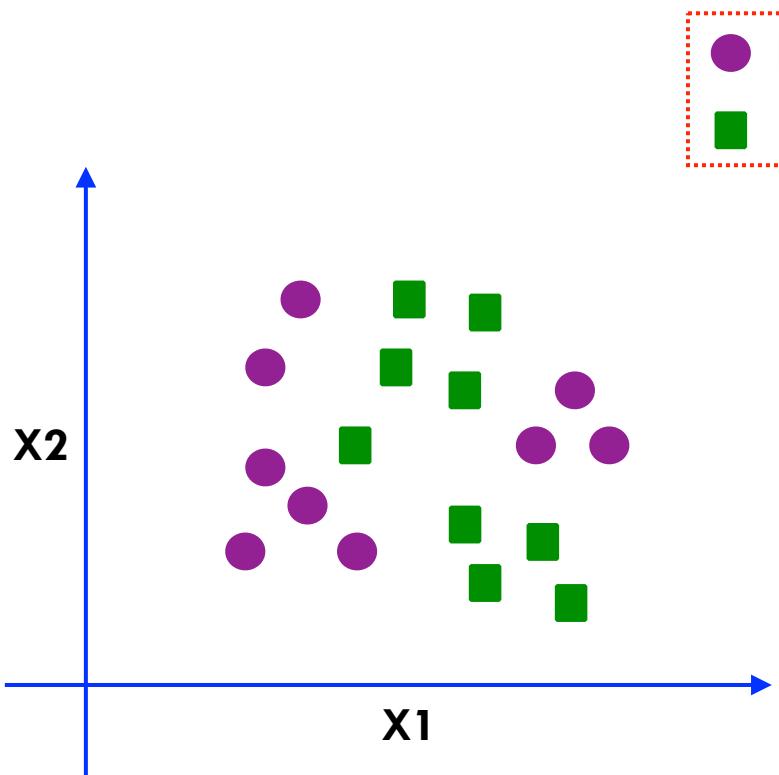


Tarefa descritiva
(**Clustering**)



Tarefa preditiva
(**Classificação**/
Regressão)

Tipos de ML

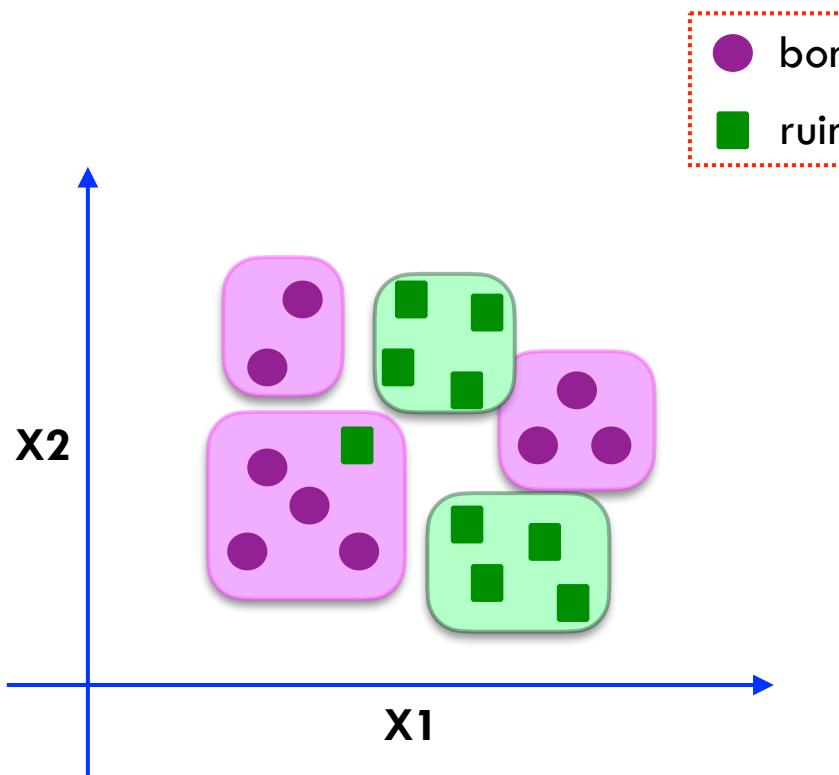


Tarefa descritiva
(Clustering)



Tarefa preditiva
**(Classificação/
Regressão)**

Tipos de ML

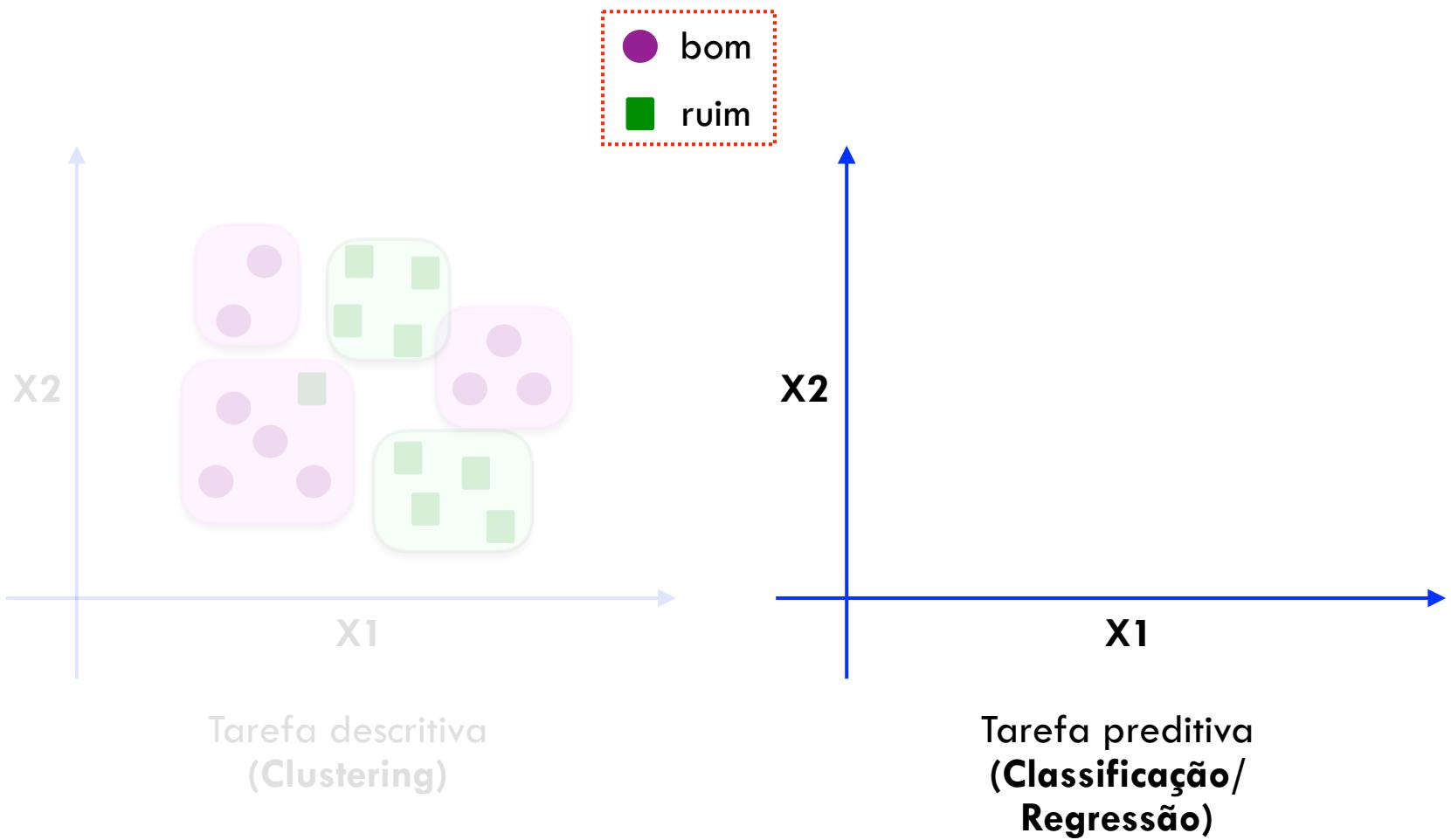


Tarefa descritiva
(Clustering)

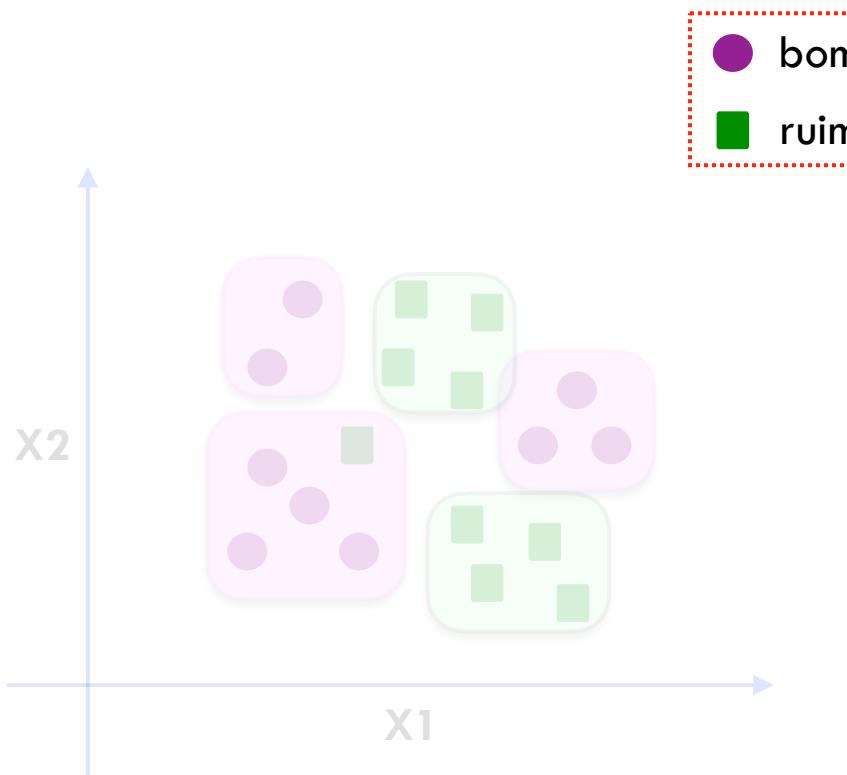


Tarefa preditiva
**(Classificação /
Regressão)**

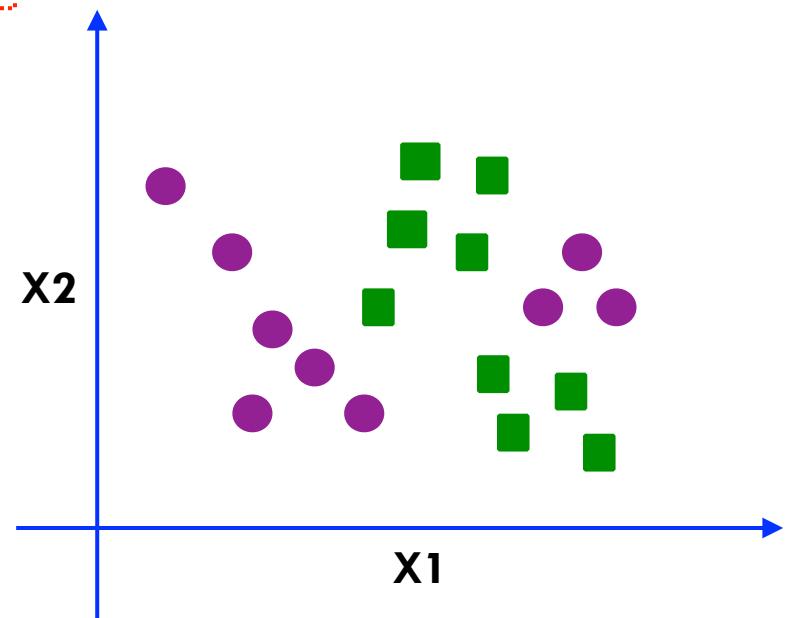
Tipos de ML



Tipos de ML

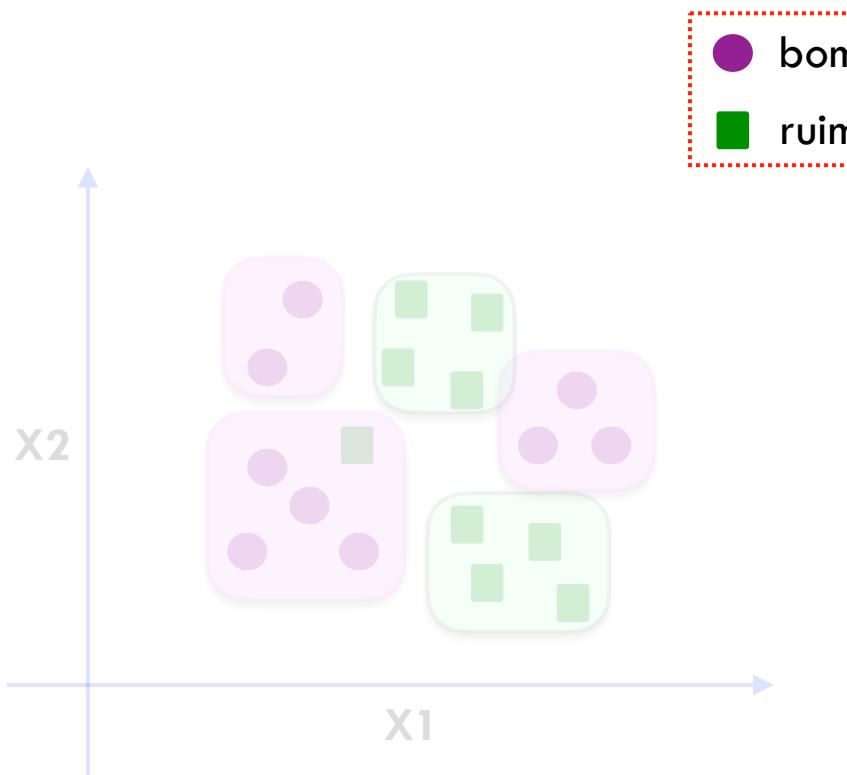


Tarefa descritiva
(Clustering)

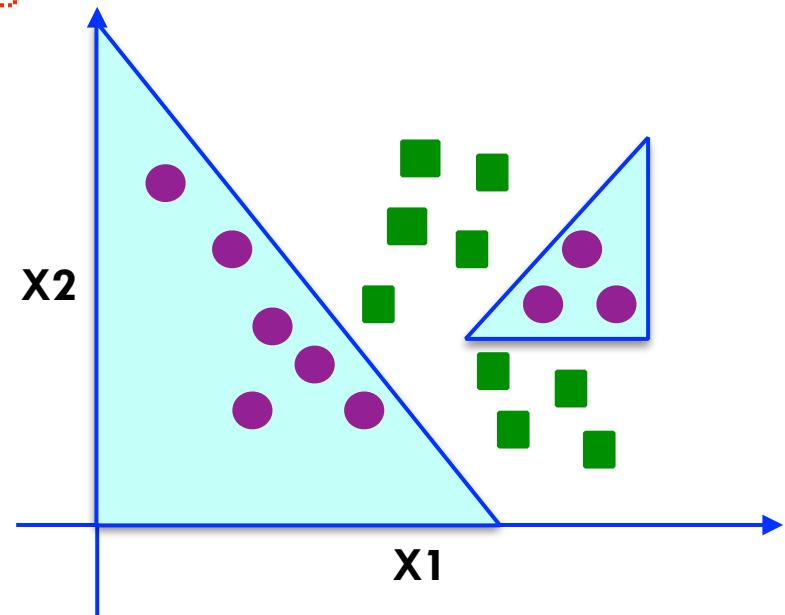


Tarefa preditiva
(Classificação/
Regressão)

Tipos de ML

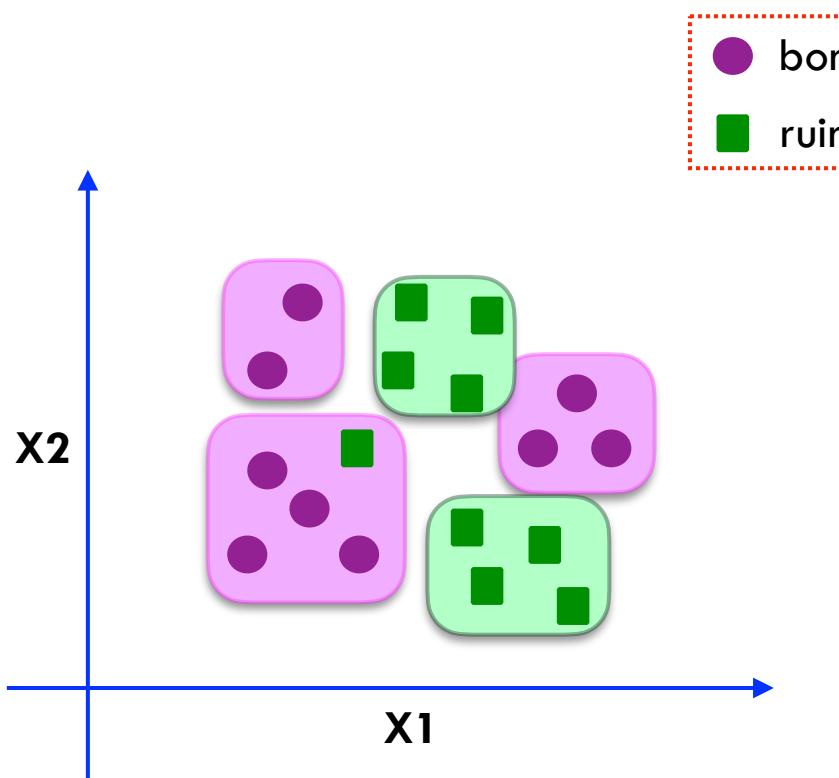


Tarefa descritiva
(Clustering)

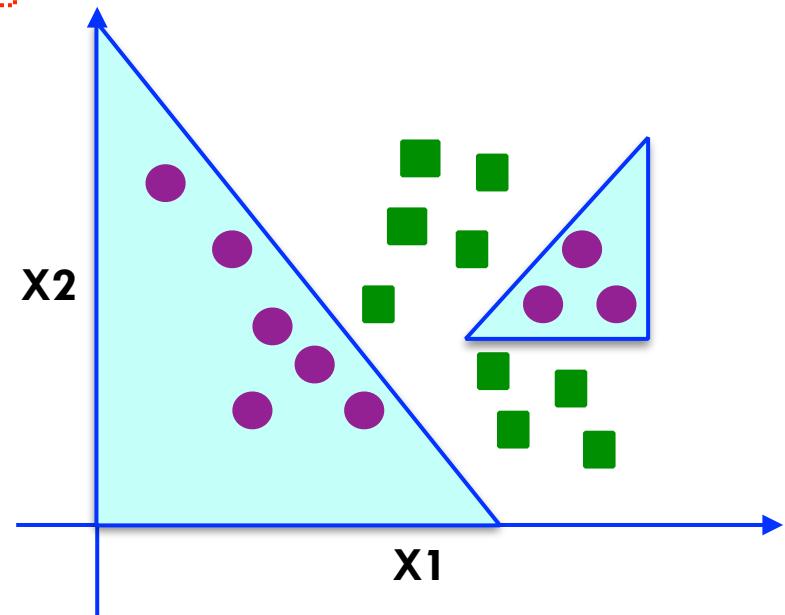


Tarefa preditiva
(Classificação/
Regressão)

Tipos de ML



Tarefa descritiva
(Clustering)



Tarefa preditiva
(Classificação/
Regressão)

Tipos de ML

x	t
0	0
0.5236	1.5
1.0472	-2.5981
1.5708	3.0
2.0944	-2.5981
2.6180	1.5
3.1416	0

Tipos de ML

x	t
0	0
0.5236	1.5
1.0472	-2.5981
1.5708	3.0
2.0944	-2.5981
2.6180	1.5
3.1416	0



?

Roteiro

1 Introdução

2 Dados

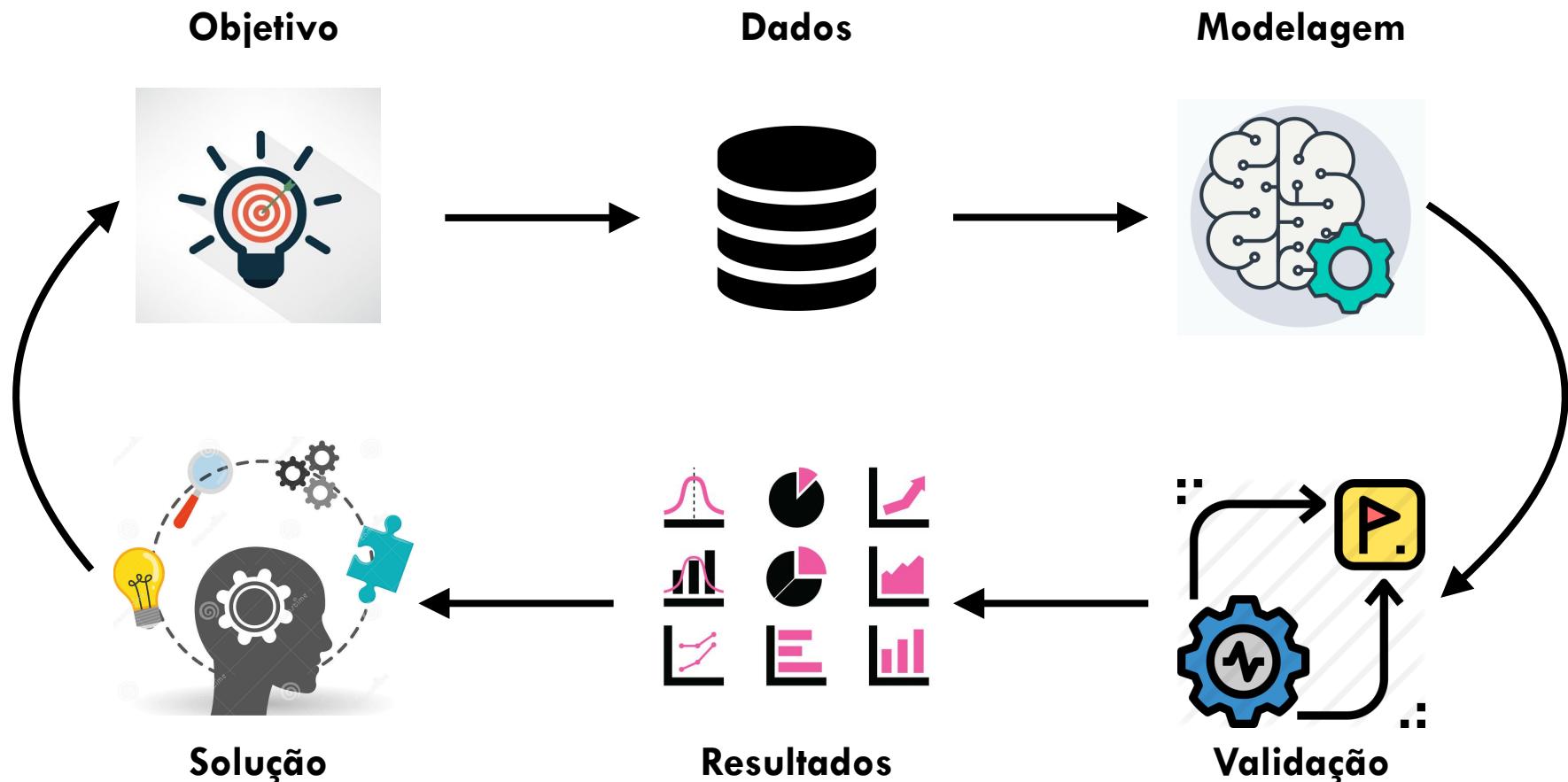
3 Tipos de ML

4 Pipeline

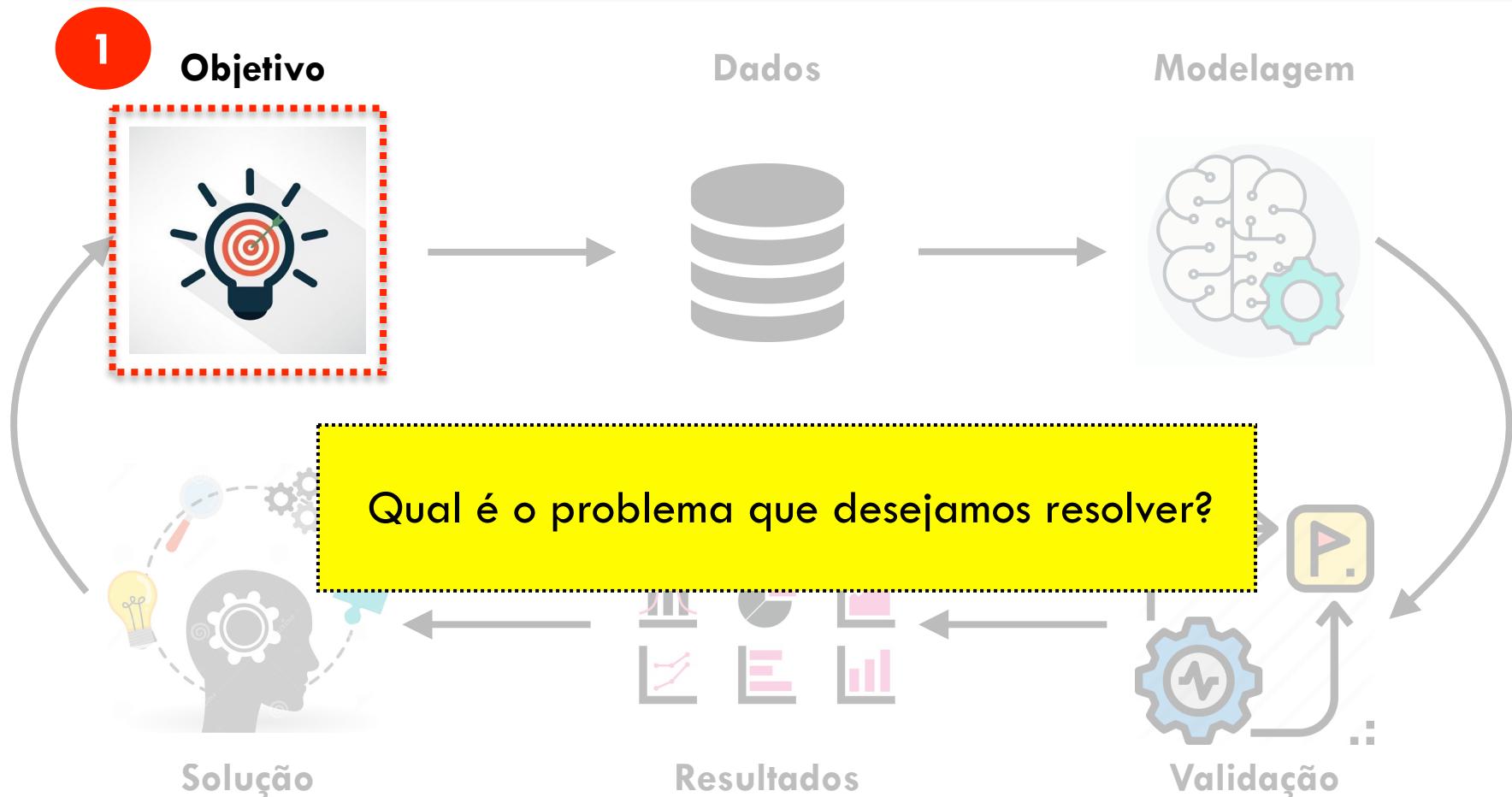
5 Ferramentas de ML

6 Referências

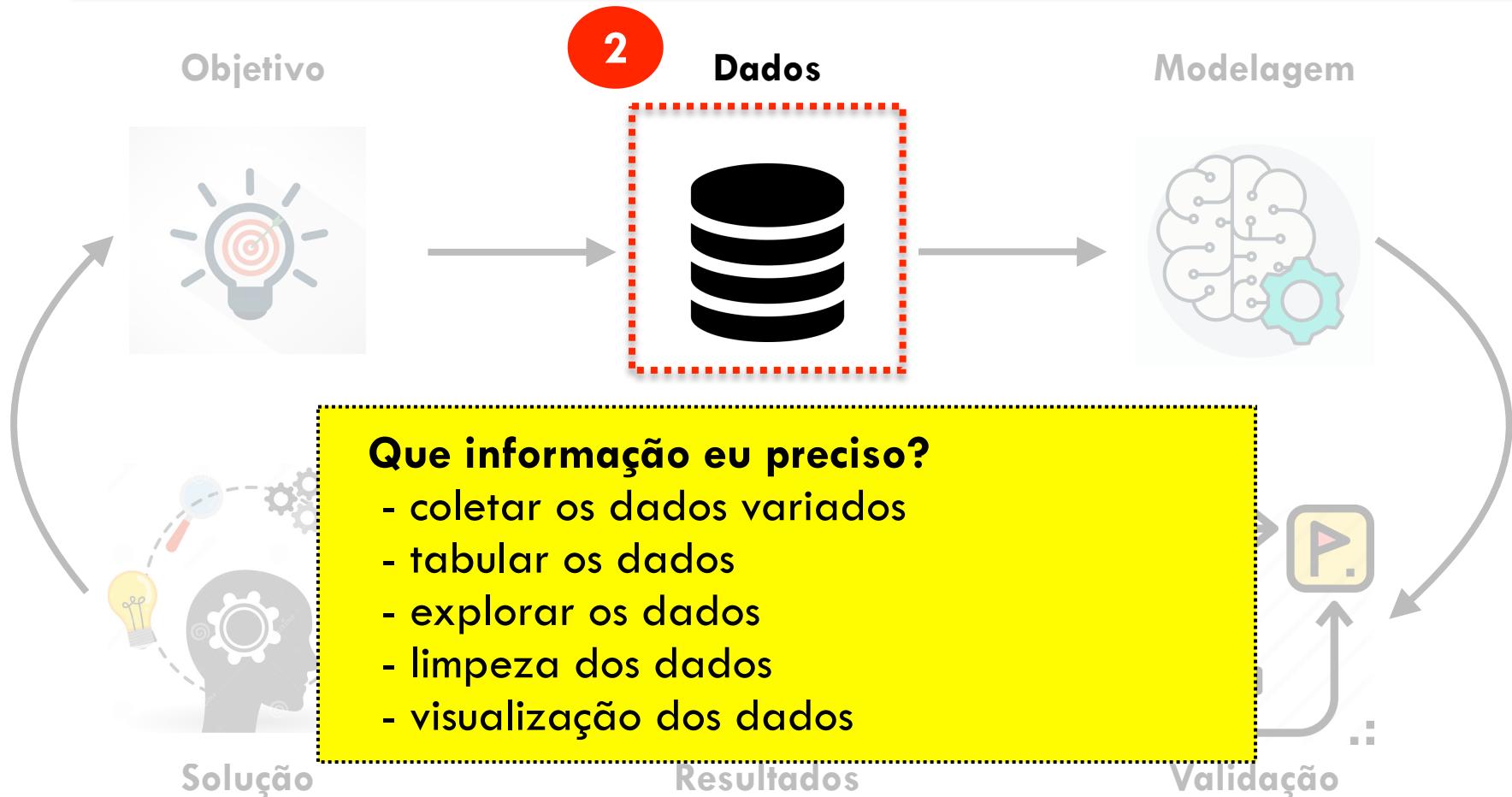
Pipeline



Pipeline



Pipeline

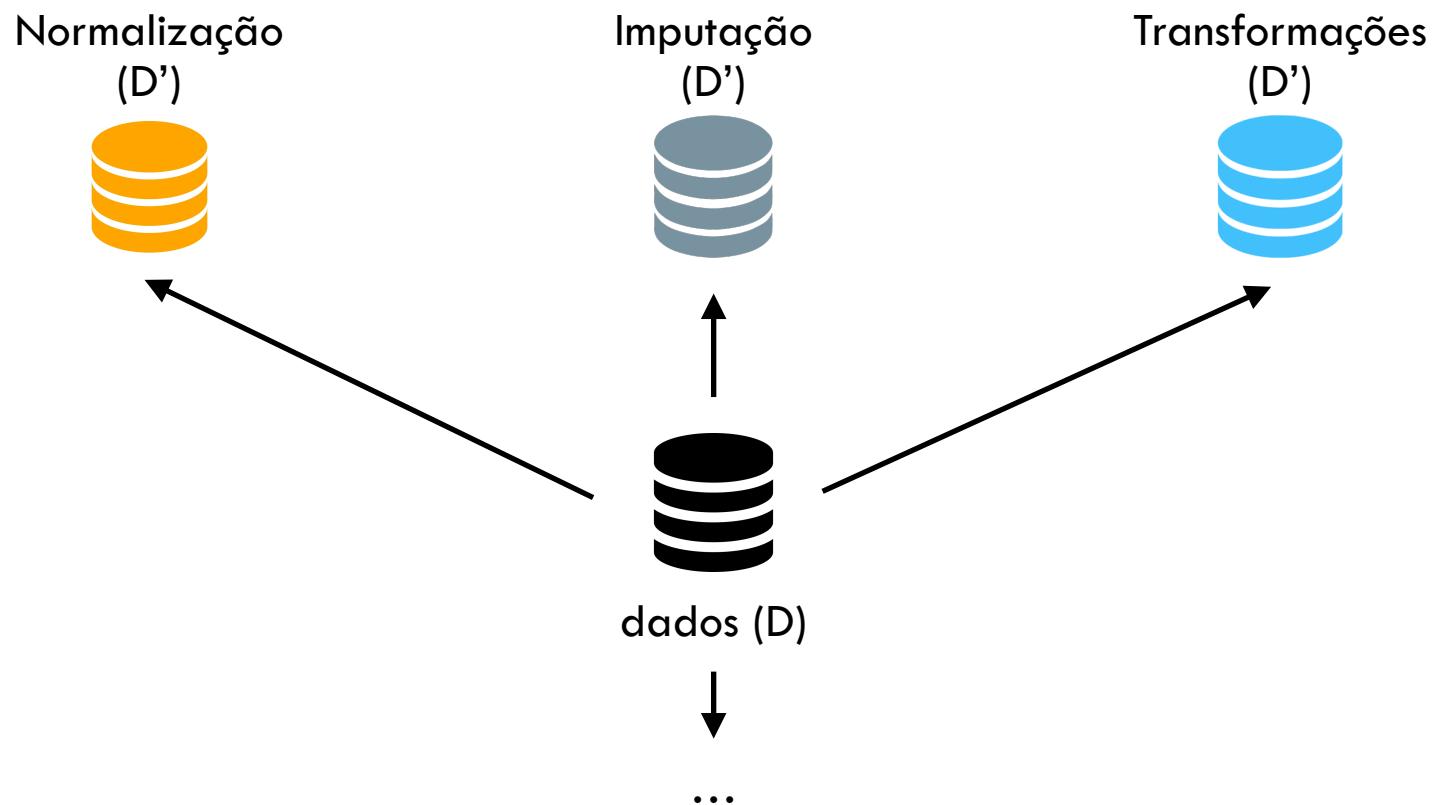


Pipeline

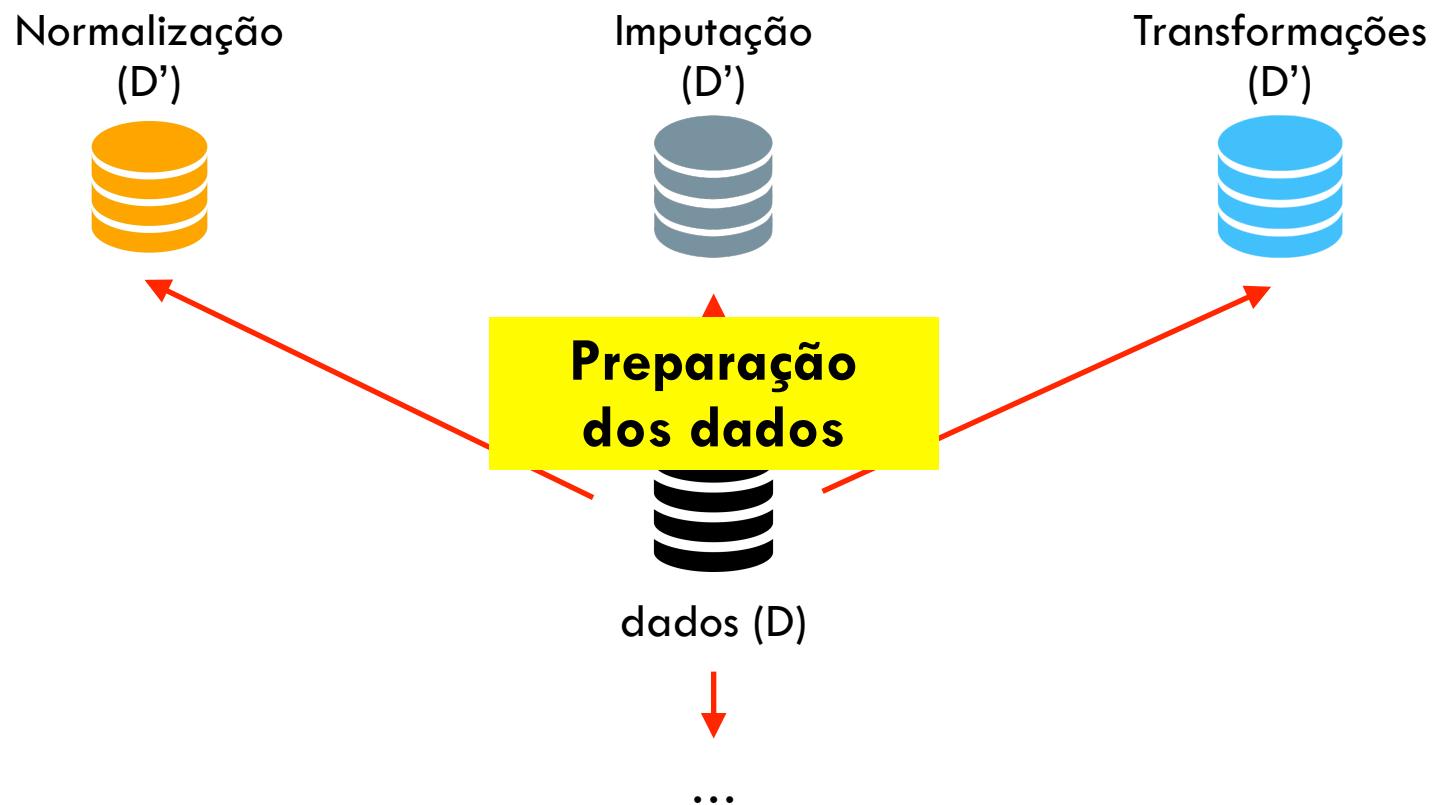


dados (D)

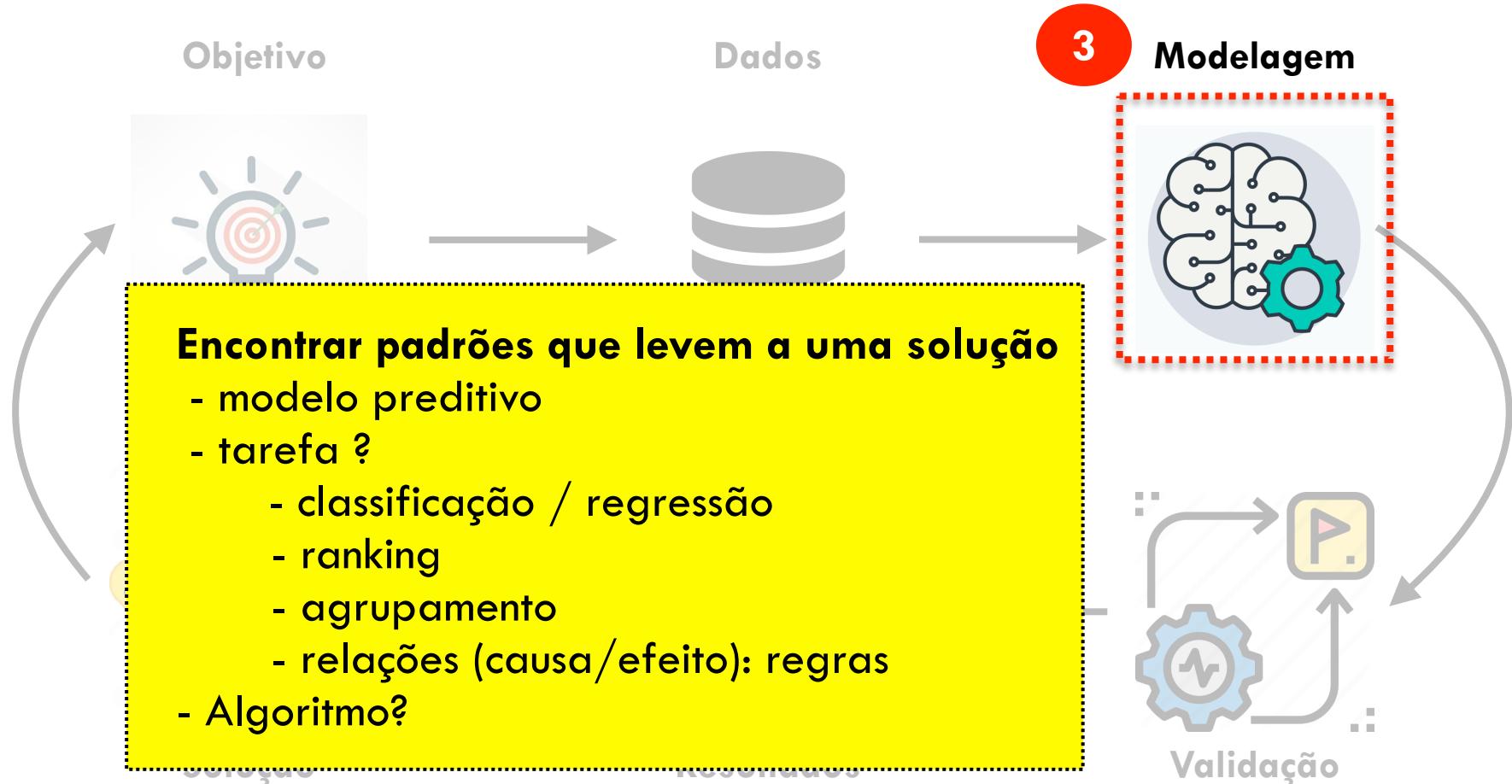
Pipeline



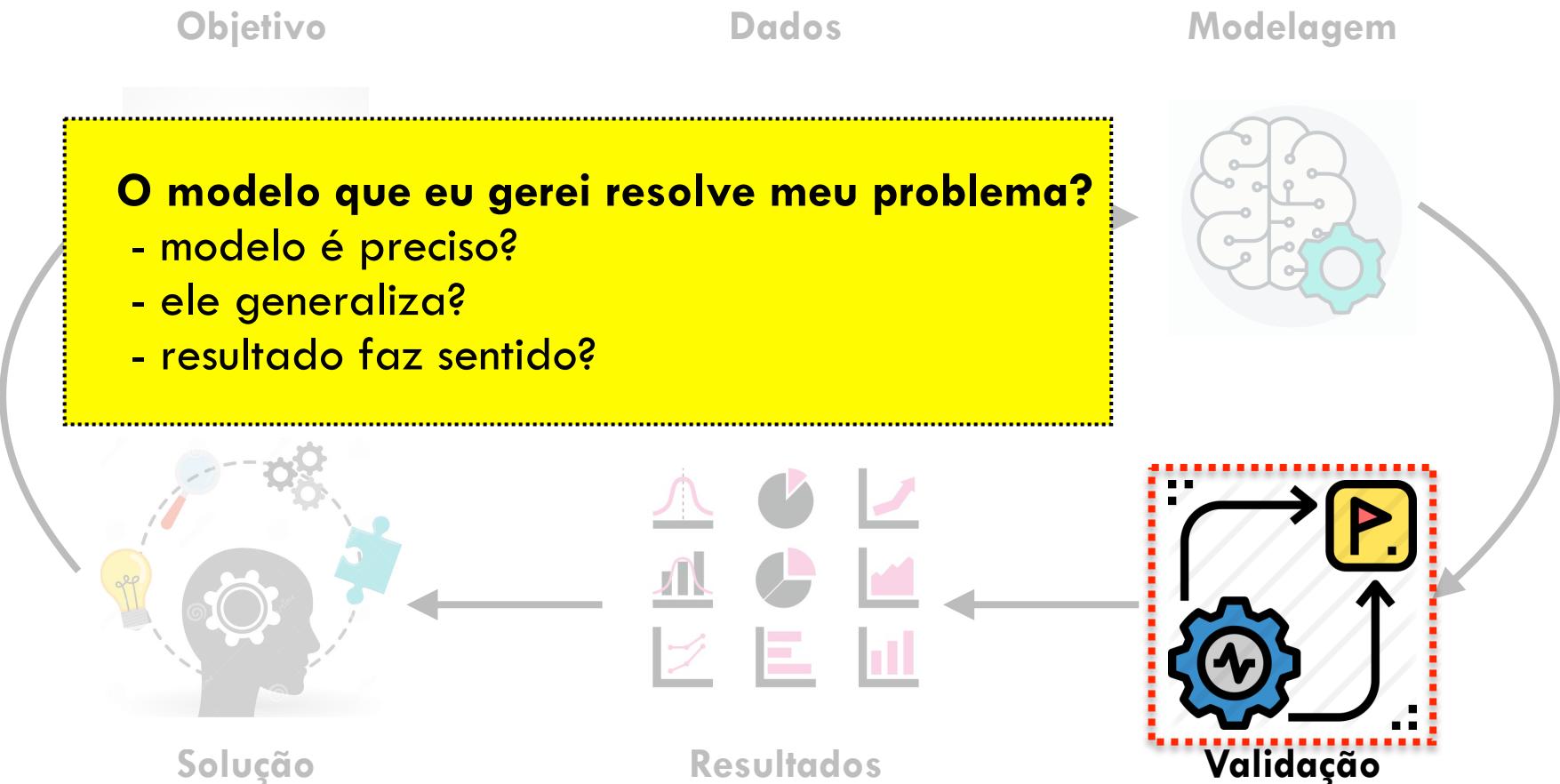
Pipeline



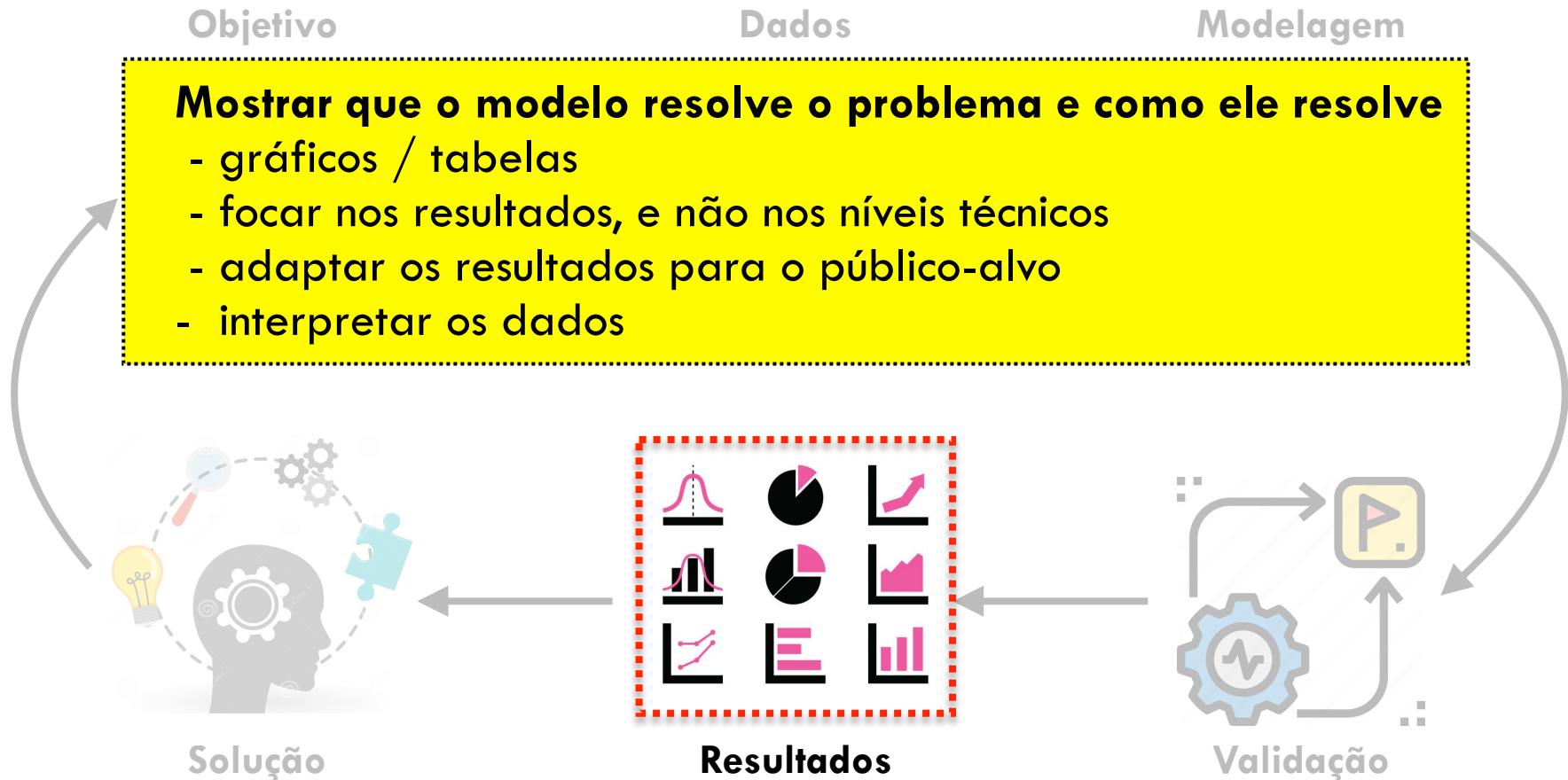
Pipeline



Pipeline



Pipeline



Pipeline

Objetivo



Dados



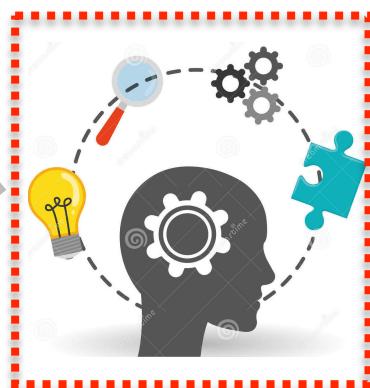
Modelagem



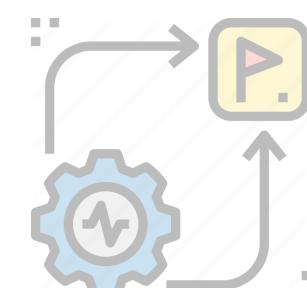
Colocar a solução para funcionar :)

- modelo deve ser atualizável
- começar com um piloto (escopo limitado)
- aumentar aos poucos
- escalar

Solução

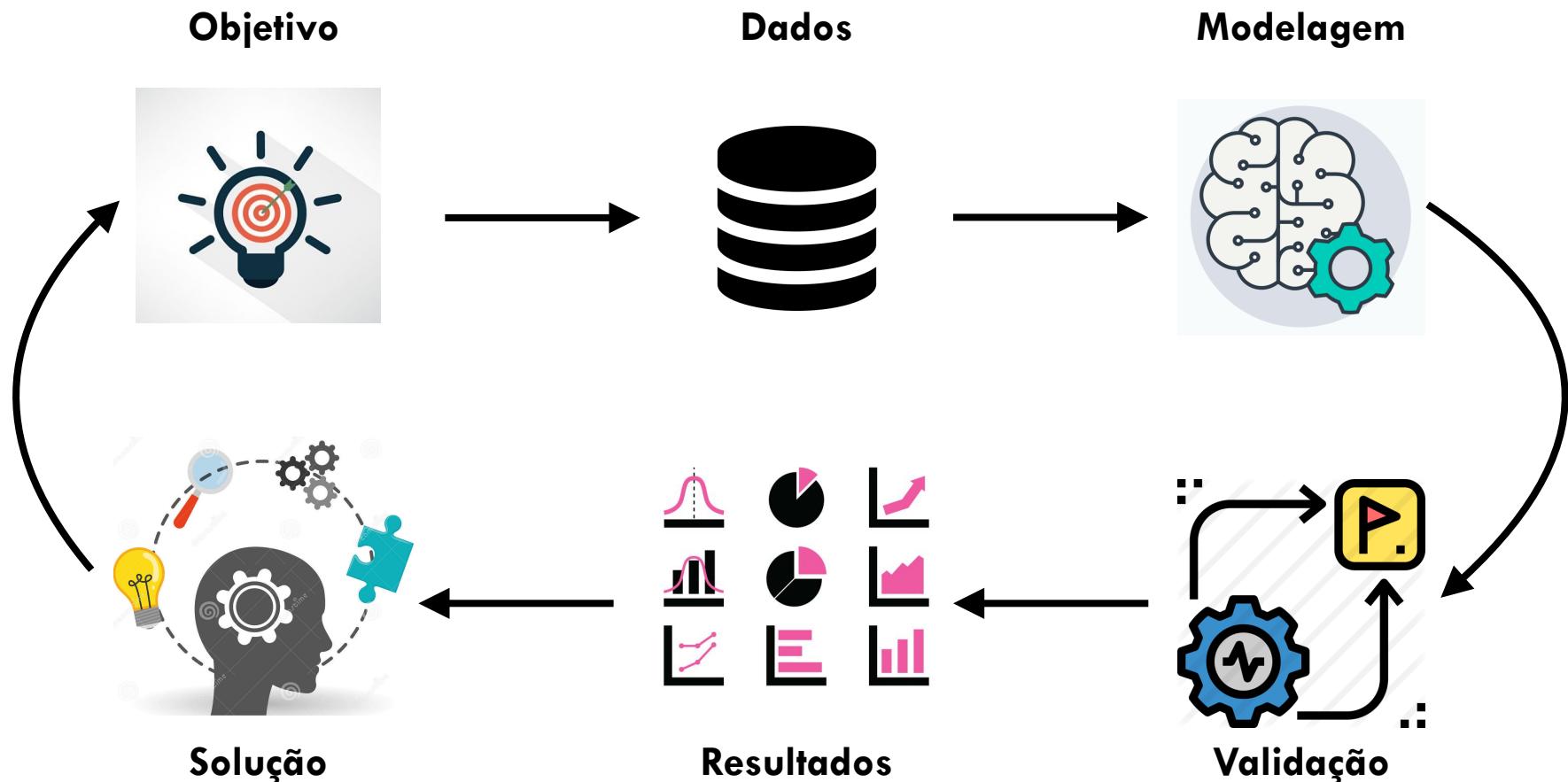


Resultados

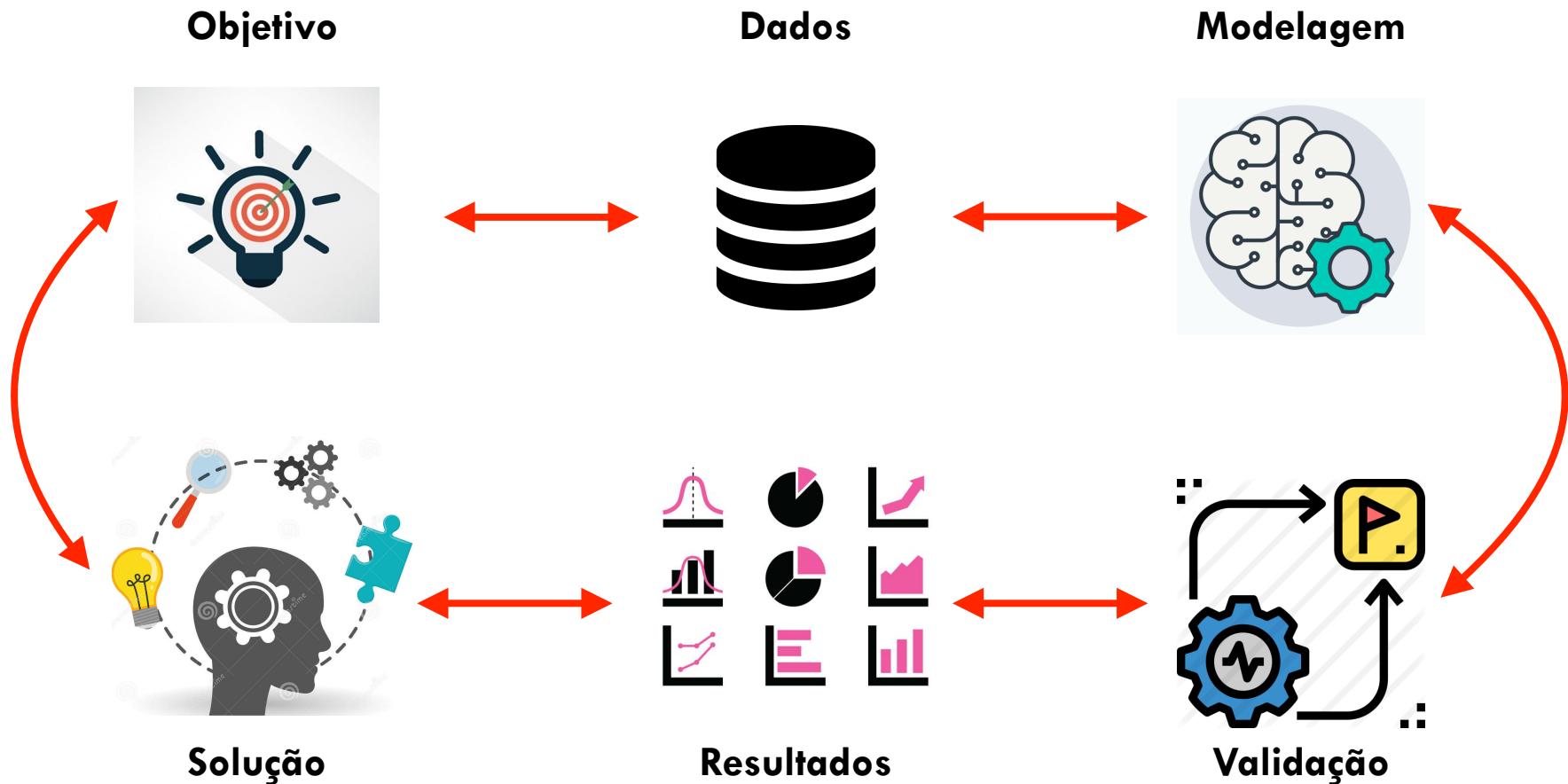


Validação

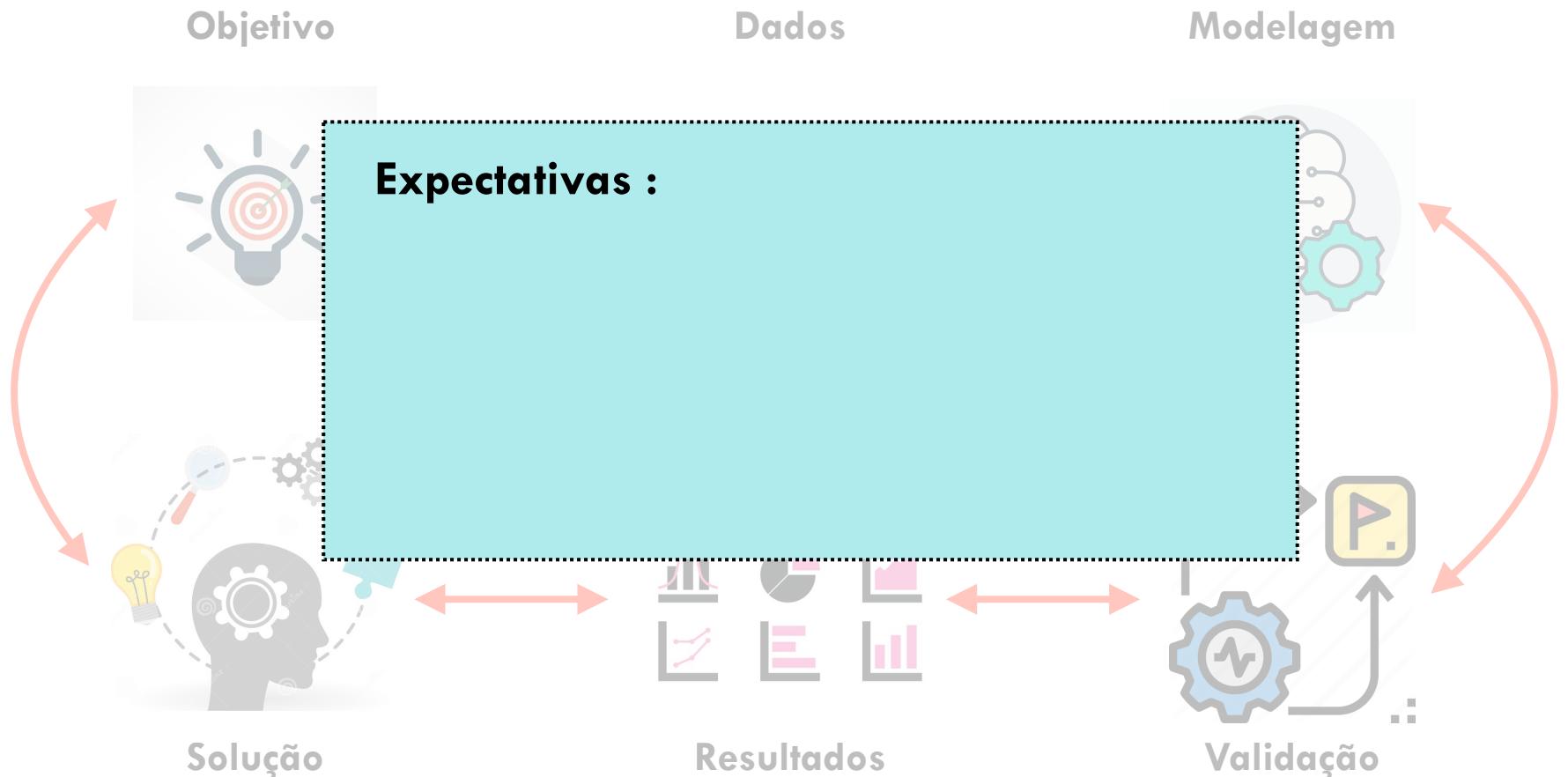
Pipeline



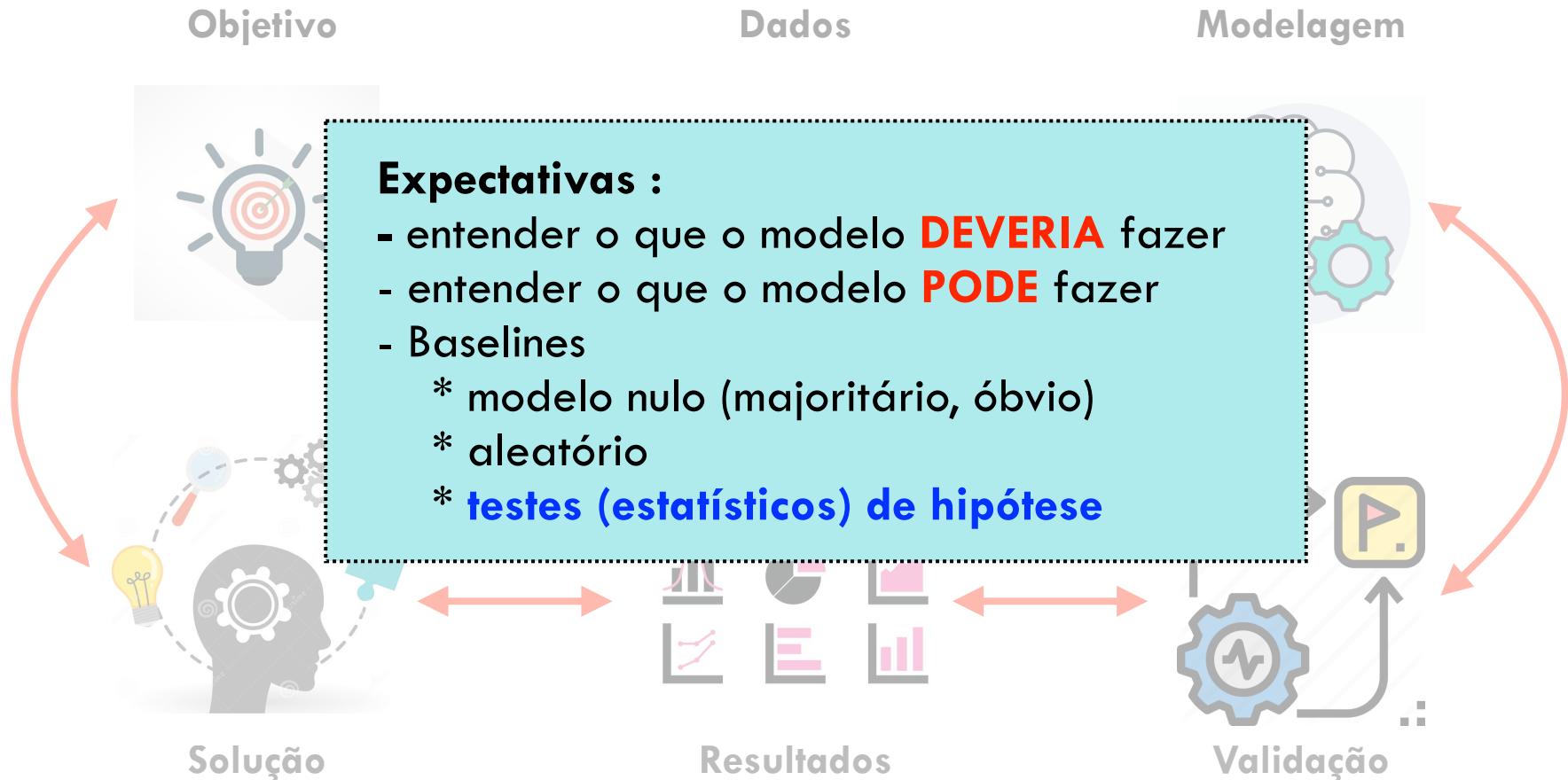
Pipeline



Pipeline



Pipeline



Roteiro

1 Introdução

2 Dados

3 Tipos de ML

4 Pipeline

5 Ferramentas de ML

6 Referências

Ferramentas



rapidminer



TensorFlow

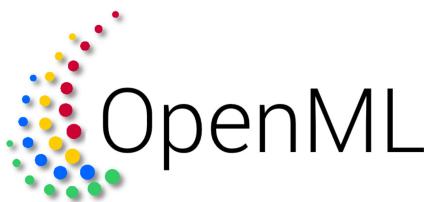


Keras



...

Ferramentas



OpenML

The screenshot shows the homepage of the OpenML website. At the top, there is a navigation bar with a menu icon, the text "OpenML", and a search bar labeled "Search". The main content area has a colorful gradient background (blue, purple, red, green, orange). It features the text "OpenML" and "A worldwide machine learning lab". Below this, a paragraph explains the platform's mission: "Machine learning research should be easily accessible and reusable. OpenML is an open platform for sharing datasets, algorithms, and experiments - to learn how to learn better, together." At the bottom, there is a search bar with the placeholder "Search OpenML", a "Datasets" dropdown menu, and a "Sign Up" button. A message next to the button says "to start tracking and sharing your own work. OpenML is open and free to use." To the right of the main content, there is a diagram illustrating the OpenML ecosystem. It shows a central brain icon labeled "OpenML" with a gear and a flask icon. Two arrows point from a desktop computer icon and a laptop icon towards the central node. Blue speech bubbles between the devices and the central node contain the text "I shared a new data set" and "I found a better model!".

OpenML

Search

OpenML

A worldwide machine learning lab

Machine learning research should be easily accessible and reusable. OpenML is an open platform for sharing datasets, algorithms, and experiments - to learn how to learn better, together.

Search OpenML

Datasets ▾

Sign Up to start tracking and sharing your own work. OpenML is open and free to use.

I shared a new data set

I found a better model!

OpenML

Desktop Computer Icon

Laptop Icon

Gear and Flask Icons

OpenML / Dados

7983 results

FILTERS

FOR SEARCH OPTIONS, SE



iris (4)

This is perhaps the best known database to be found in the pattern

★ 0 runs ❤ 0 likes 🌐 0 downloads 📺 0 reach ↗ 3 impact

150 instances - 5 features - 3 classes - 0 missing values



iris (1)

This is perhaps the best known database to be found in the pattern

★ 8030 runs ❤ 5 likes 🌐 82 downloads 📺 87 reach ↗ 31 impact

150 instances - 5 features - 3 classes - 0 missing values



Subgroup Discovery on iris

★ 1298 runs ❤ 0 likes 🔩 0 downloads 📺 0 reach ↗ 2 impact

uploader_id : 1 - quality_measure : Information gain - target_feature : class - target_value : Iris-setosa - reuse : 0



Learning Curve on iris-test

OpenML / Dados

7983 results FILTERS FOR SEARCH OPTIONS, SEE

 iris (4)	This is perhaps the best known database to be found in the pattern ★ 0 runs ❤ 0 likes 🌐 0 downloads 📺 0 reach ⚡ 3 impact 150 instances - 5 features - 3 classes - 0 missing values
 iris (1)	This is perhaps the best known database to be found in the pattern ★ 8030 runs ❤ 5 likes 🌐 82 downloads 📺 87 reach ⚡ 31 impact 150 instances - 5 features - 3 classes - 0 missing values
 Subgroup Discovery on iris	★ 1298 runs ❤ 0 likes 🔩 0 downloads 📺 0 reach ⚡ 2 impact uploader_id : 1 - quality_measure : Information gain - target_feature : class - target_value : Iris-setosa - reuse
 Learning Curve on iris-test	

active [ARFF](#) [Publicly available](#) Visibility: public Uploaded 06-04-2014 by [Jan van Rijn](#)

5 likes downloaded by 82 people, 104 total downloads 0 issues 0 downvotes

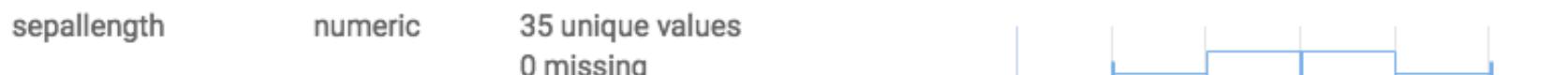
[study_1](#) [study_25](#) [study_4](#) [study_41](#) [study_50](#) [study_52](#) [study_7](#) [study_86](#) [study_88](#) [study_89](#) [uci](#) [+ Add tag](#)[Help us complete this description →](#) Edit**Author:** R.A. Fisher**Source:** [UCI](#) - 1936 - Donated by Michael Marshall**Please cite:**

Iris Plants Database

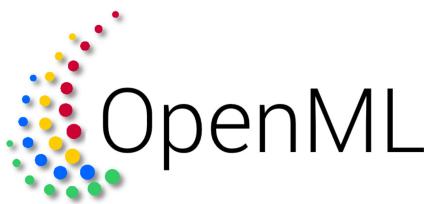
This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda & Hart, for example.) The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly

[▼ Show all](#)

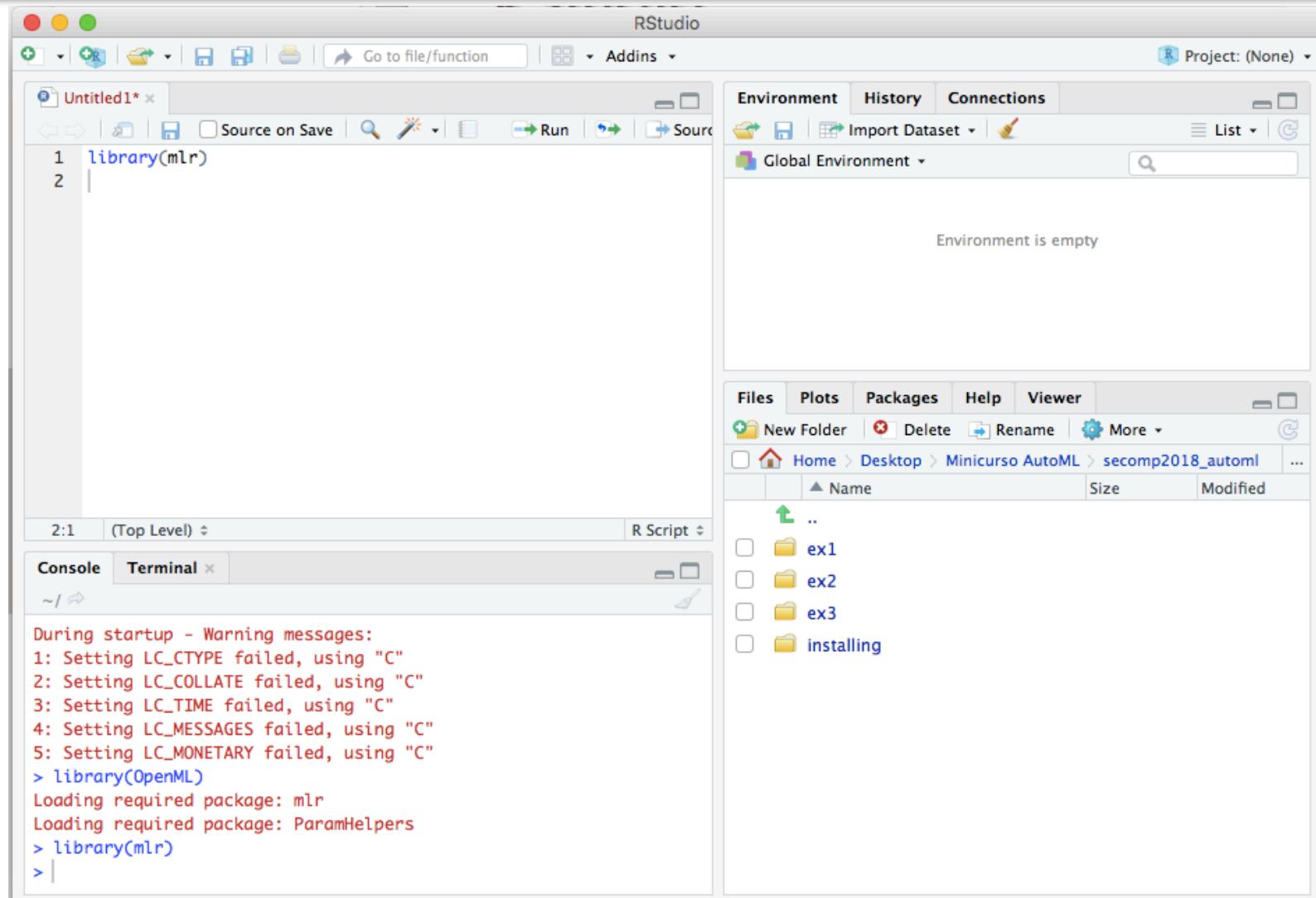
5 features



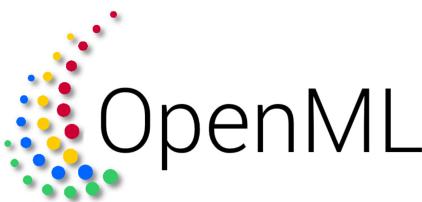
Ferramentas



Studio / IDE para R



Ferramentas



mlr / framework em R

Machine Learning in R



[build failing](#) [build failing](#) [CRAN 2.13](#) [downloads 7732/month](#) [stackoverflow mlr](#)

- [CRAN release site](#)
- Detailed Tutorial: [Online as HTML](#)
- [mlr cheatsheet](#)
- Install the development version

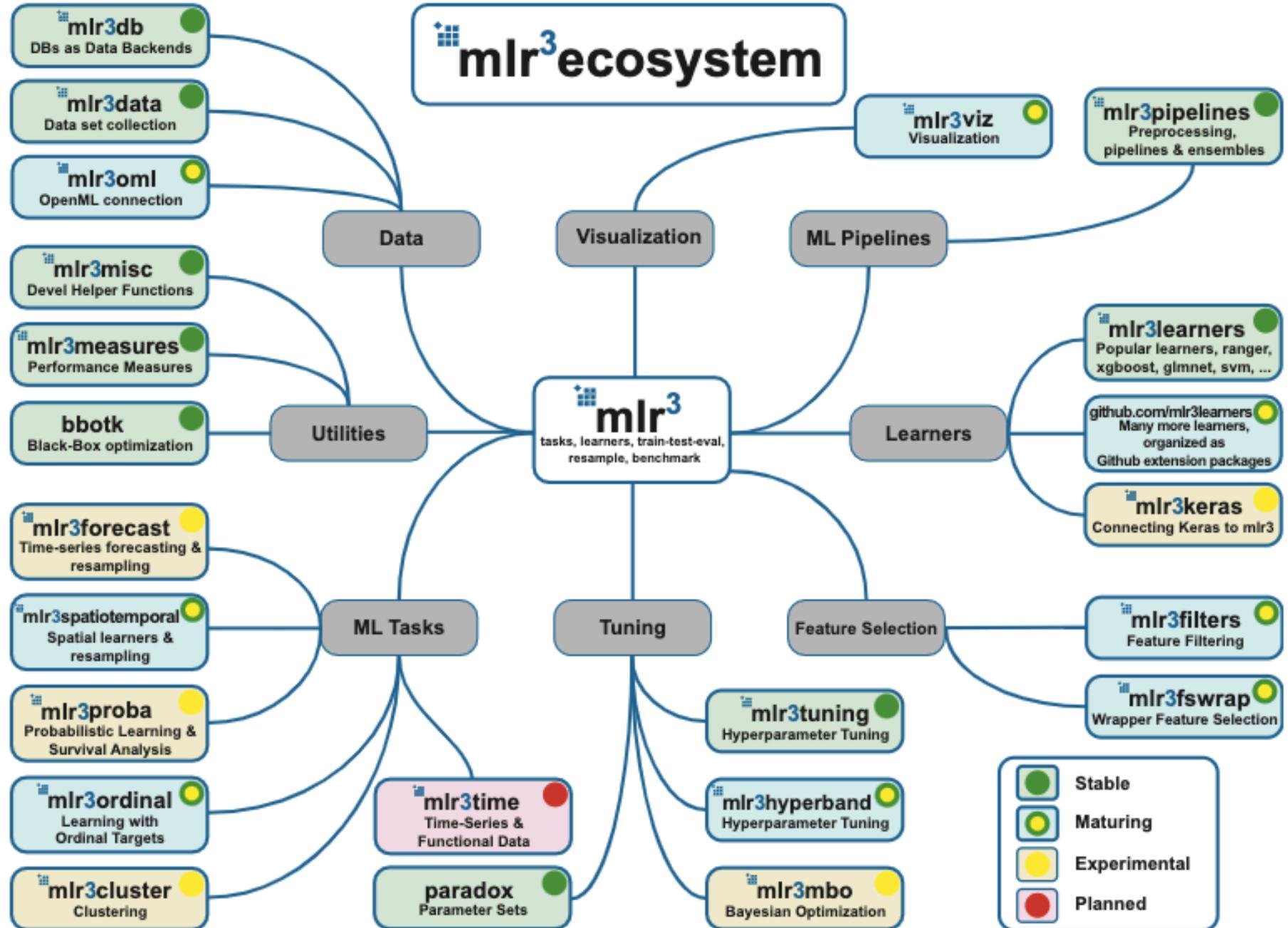
```
devtools::install_github("mlr-org/mlr")
```

- [Further installation instructions](#)
- [Ask a question about mlr on Stackoverflow](#)
- [We are on Slack](#) (Request invitation: code{at}jakob-r.de)
- [We have a blog on mlr](#)
- A list of possible enhancements to mlr is available on the [wiki](#) - contributors welcome!
- We are in the top 20 of the most starred R packages on Github, as reported by [metacran](#).

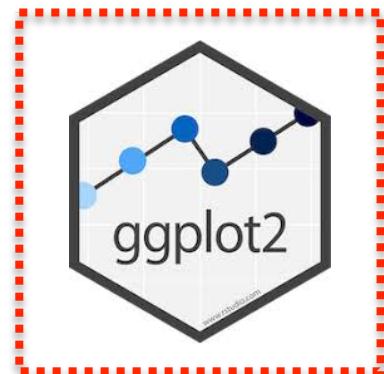
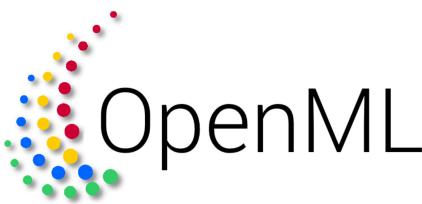
mlr / framework em R

- Página principal:
 - <https://github.com/mlr-org/mlr>
- Tutoriais:
 - <https://mlr-org.github.io/mlr/>
 - <https://mlr-org.github.io/mlr/articles/wrapper.html>
 - https://mlr-org.github.io/mlr/articles/integrated_learners.html
 - <https://mlr-org.github.io/mlr/articles/measures.html>
 - https://mlr-org.github.io/mlr/articles/advanced_tune.html

Extension Packages



Ferramentas



ggplot2



ggplot2 part of the [tidyverse](#)

3.2.1

Overview

ggplot2 is a system for declaratively creating graphics, based on [The Grammar of Graphics](#). It relieves you from having to think about how you want to map variables to aesthetics, what graphical primitives to use, and it takes care of

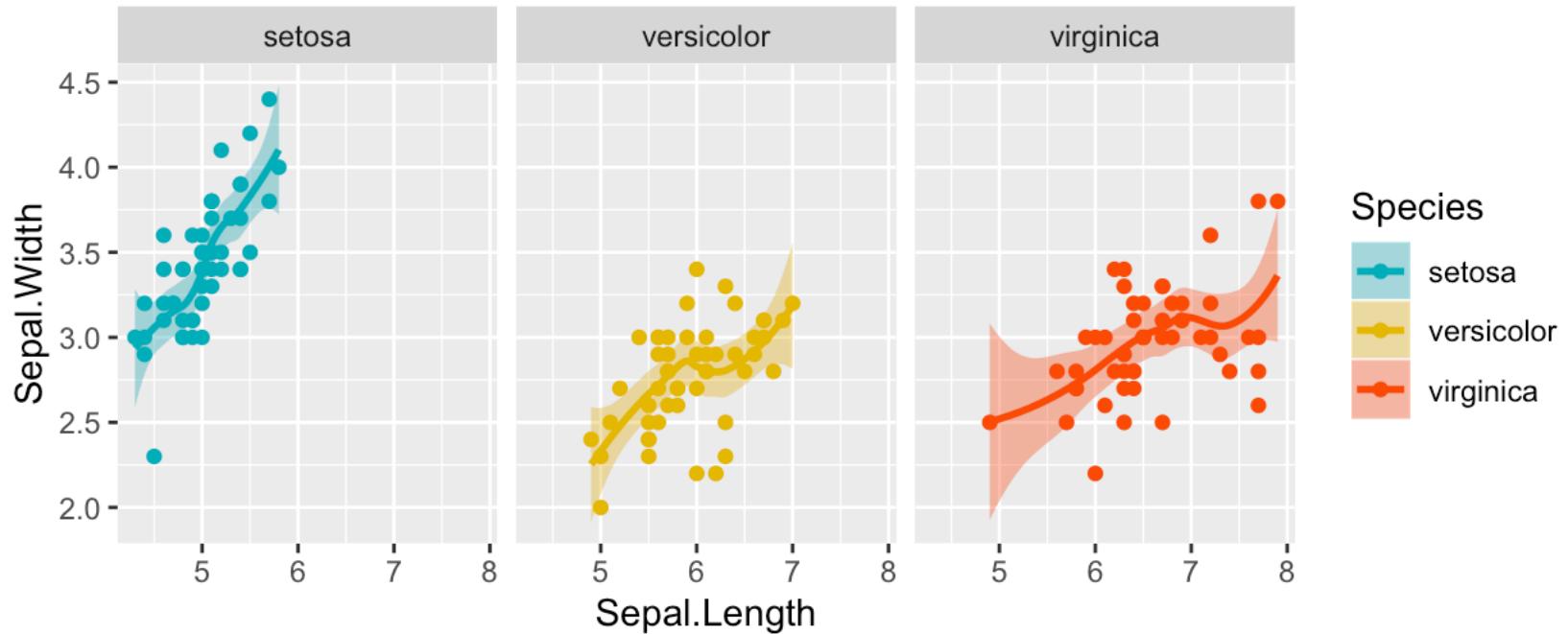
Installation

```
# The easiest way to get ggplot2 is to install the whole tidyverse:  
install.packages("tidyverse")
```

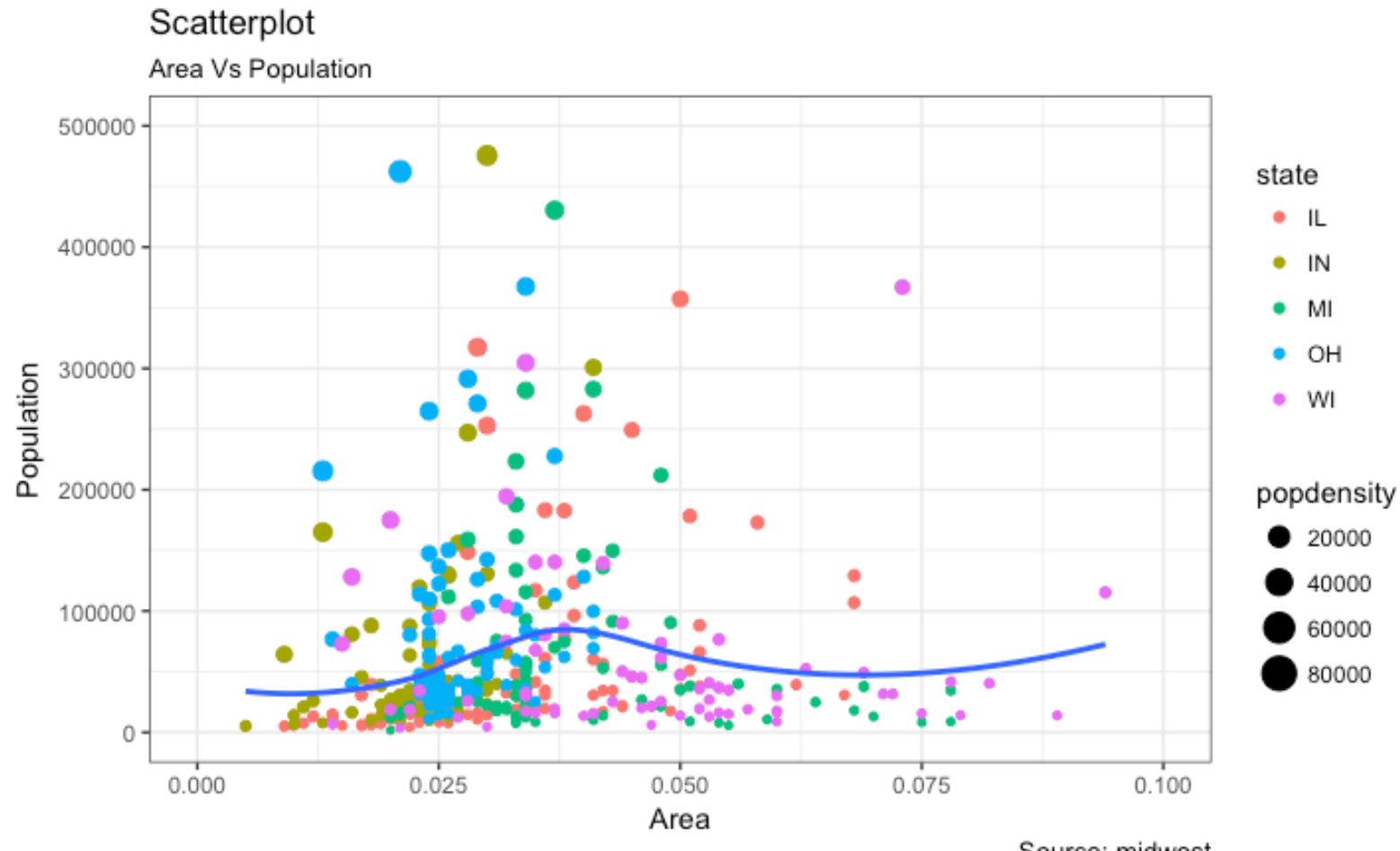
```
# Alternatively, install just ggplot2:  
install.packages("ggplot2")
```

ggplot2

- Visualização dos dados :)



ggplot2



Hello world

- O que faremos?
 - ler dados no R
 - ver características dos dados
 - plotar

Hello world

- O que faremos?
 - ler dados no R
 - ver características dos dados
 - plotar

helloWorld.R

helloWorld.R

main ▾

intelligentSystems1 / codes / 03_supervisedLearning / helloWorld.R



rgmantovani initial content with ml

1 contributor

48 lines (33 sloc) | 823 Bytes

```
1 # carregando o pacote ggplot2
2 library(ggplot2)
3
4 # acessando o dataset
5 mpg
6 View(mpg)
7
8 # contando numero de linhas do dataset
9 nrow(mpg)
10
11 #contando numero de colunas do dataset
12 ncol(mpg)
13
14 # principais caracteristicas do dataset
15 summary(mpg)
```

Exercício 01

- Fazer o mesmo com o dataset: iris
 - visualizar par a par as coordenadas
 - o que pode ser visto?

- Usar:
 - `library(ggplot2)`
 - `plot()`
 - `geom_point()`

Exercício 02

- Encontrar outro dataset no OpenML
 - baixar e ler no R
 - visualizar informação
 - Qual informação o dataset nos mostra?

- Usar:

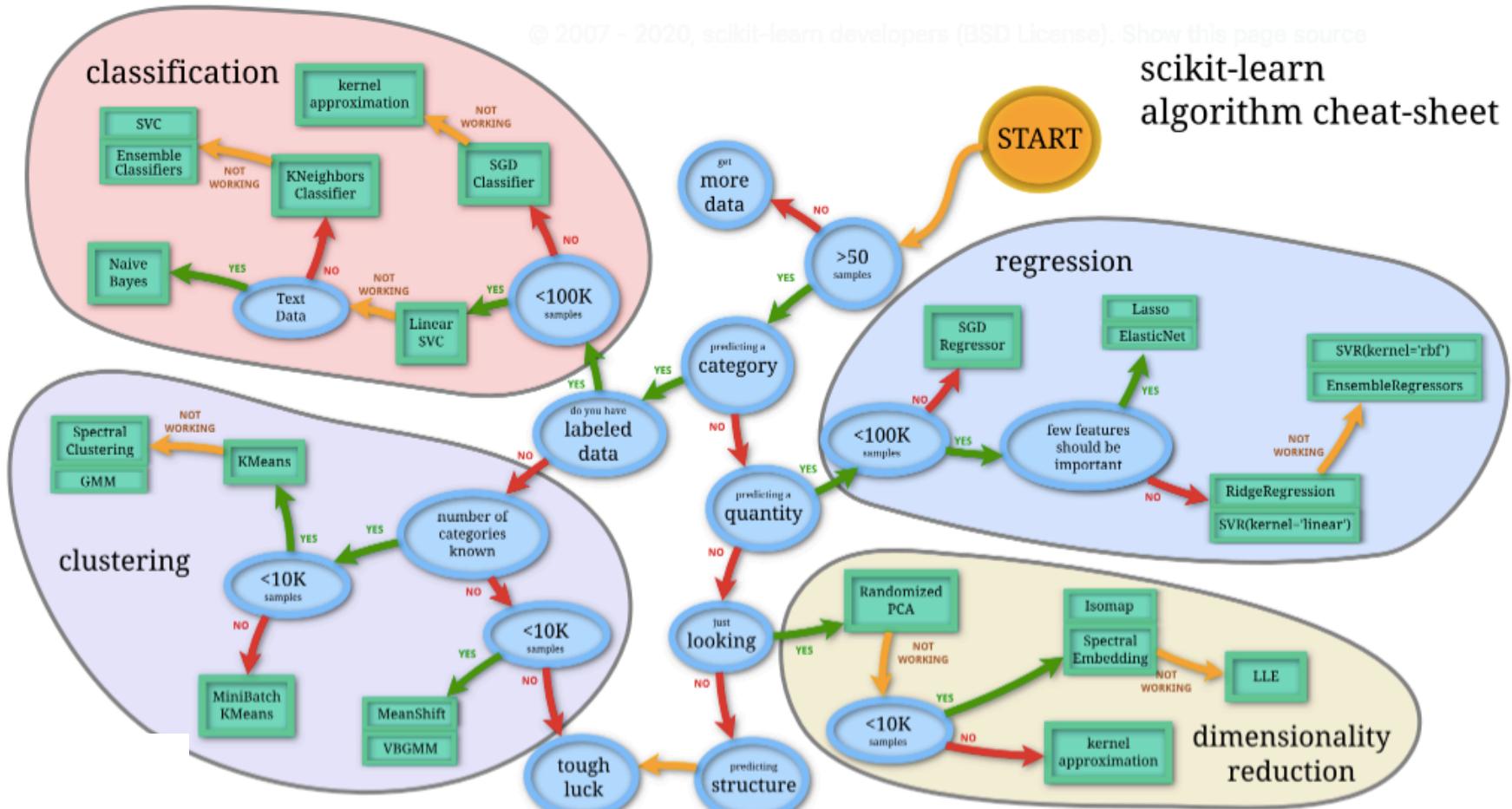
- `library(ggplot2)`
 - `library(OpenML)`
 - `read.csv()`
 - `read.table()`
 - `read.arff()`
 - `getOMLDataSet(data.id = <id>)`

Links Interessantes :)

- R for Data Science: <https://r4ds.had.co.nz>
- Tidyverse: <https://www.tidyverse.org>
- mlr: <https://mlr.mlr-org.com>
- mlr3: <https://mlr3.mlr-org.com>
- Skicit learn: <https://www.tidyverse.org>
- matplotlib: <https://matplotlib.org>
- OpenML: <https://www.openml.org>
- UCI: <https://archive.ics.uci.edu/ml/index.php>
- RStudio: <https://rstudio.com>
- Spyder: <https://www.spyder-ide.org>

scikit-learn / framework em Python

scikit-learn algorithm cheat-sheet



Roteiro

1 Introdução

2 Dados

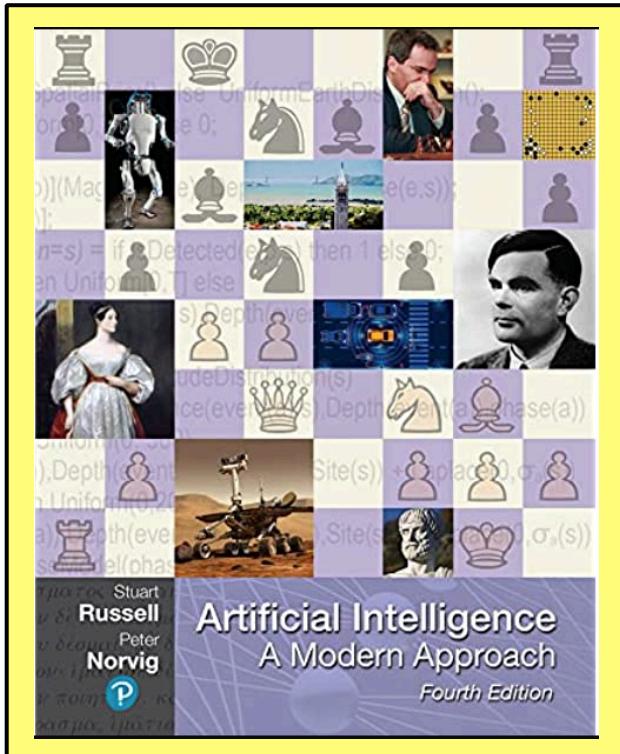
3 Tipos de ML

4 Pipeline

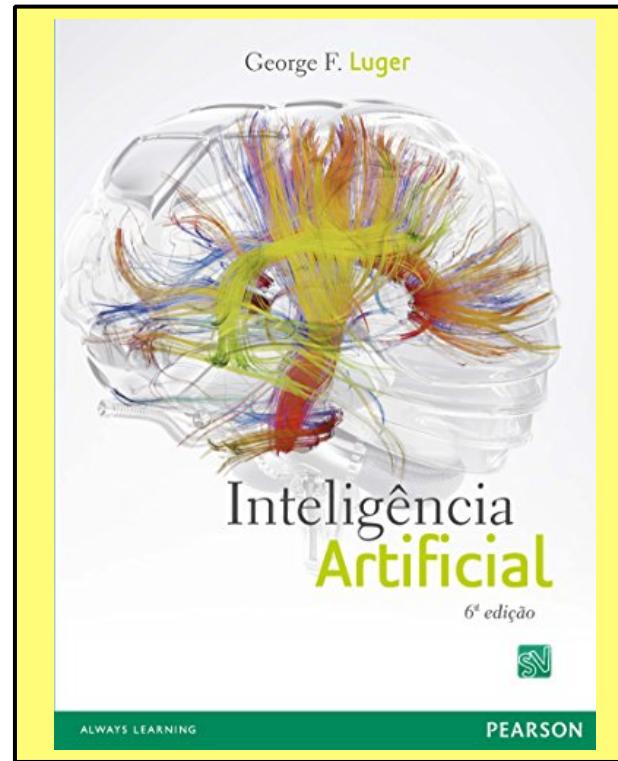
5 Ferramentas de ML

6 Referências

Referências

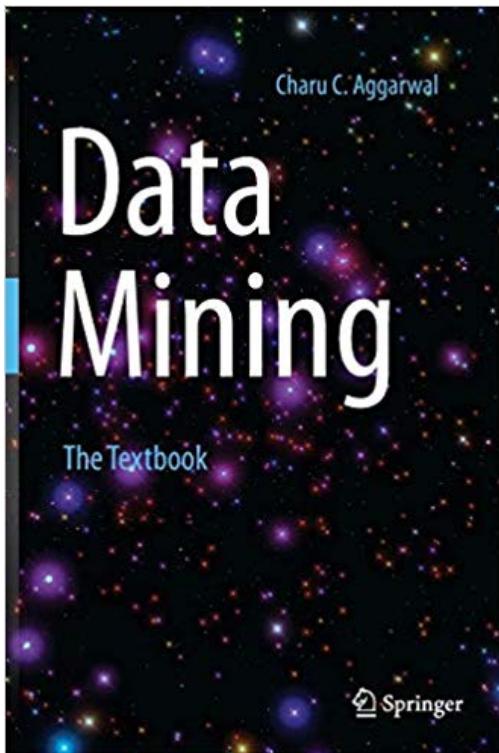


[Russel & Norvig, 2021]

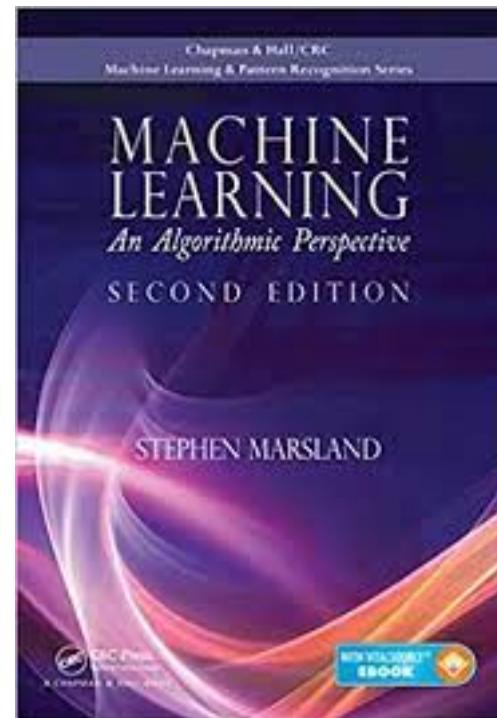


[Luger, 2013]

Referências



[Aggarwal, 2015]



[Marsland, 2014]

Perguntas?

Prof. Rafael G. Mantovani

rafaelmantovani@utfpr.edu.br