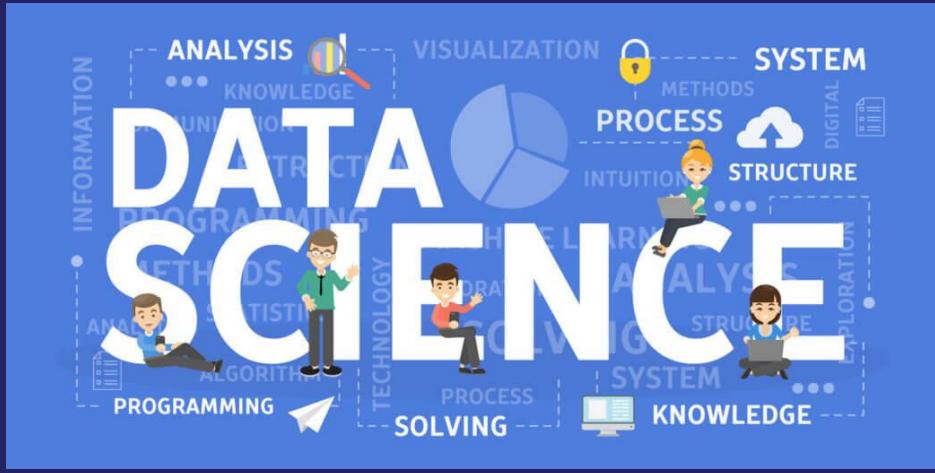




institut d'intelligence
artificielle appliquée

i2a2.academy

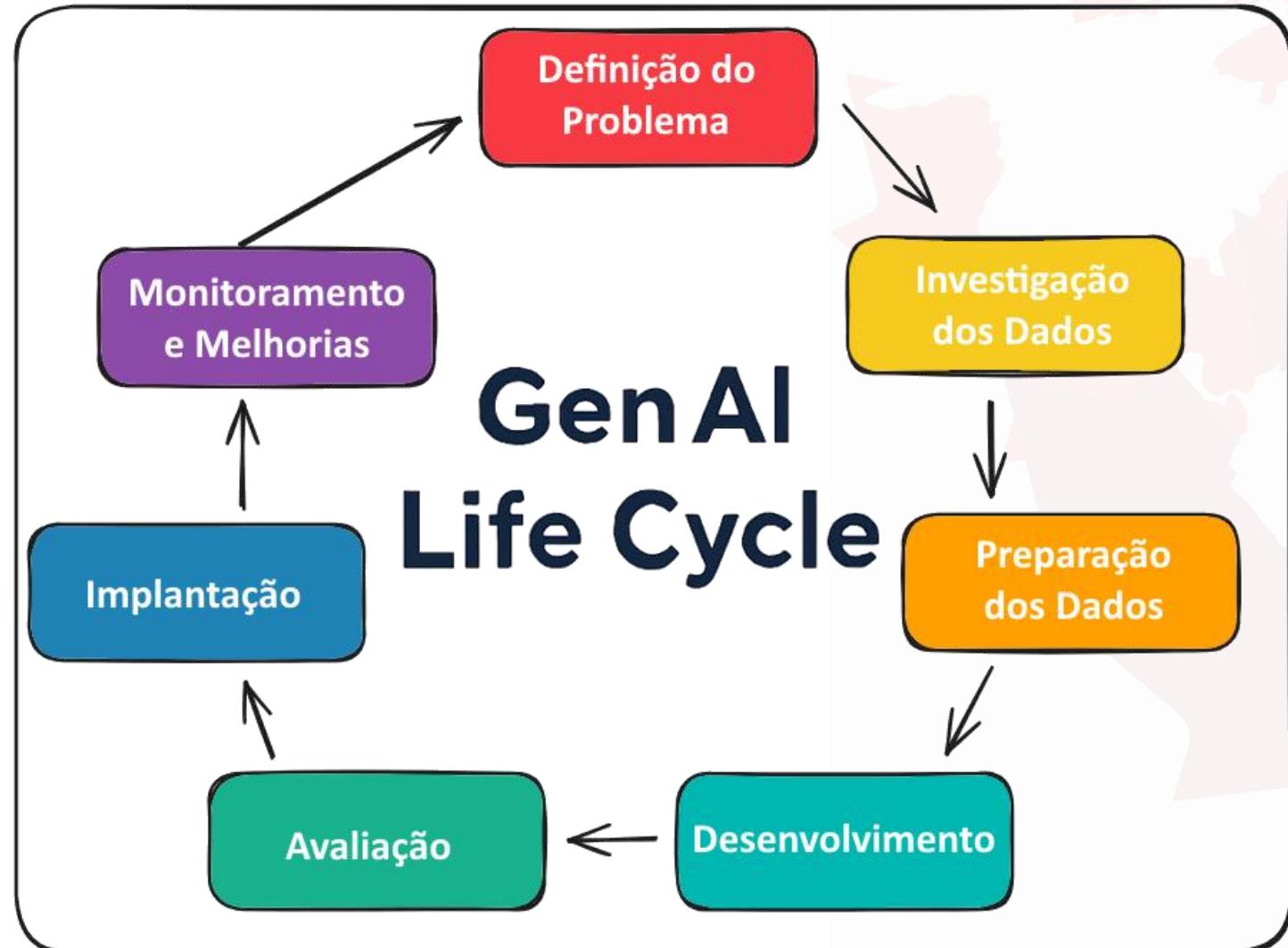




Fonte: <https://www.cienciaedados.com/por-que-e-como-data-science-e-mais-do-que-apenas-machine-learning/>

Data Science

Ciclo de vida da IA Generativa



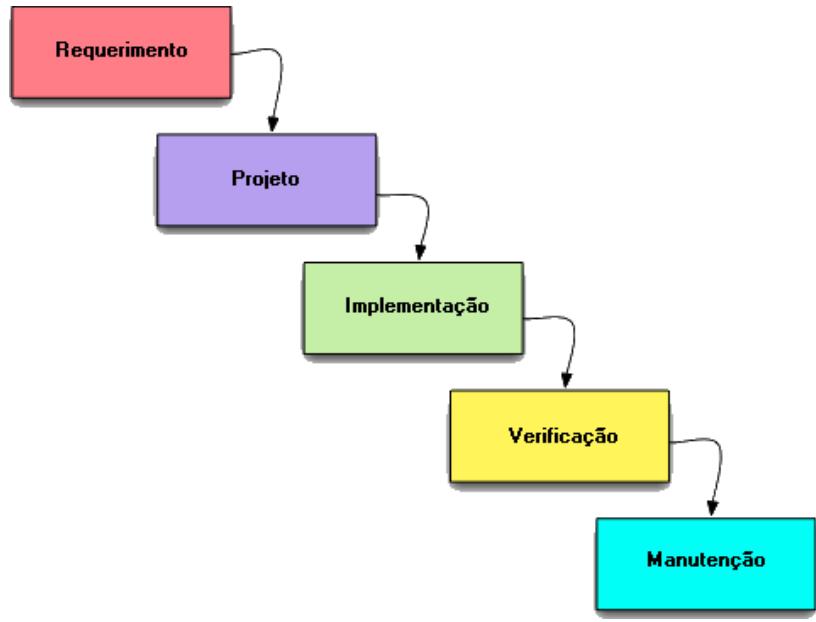
Metodologias Cíclicas



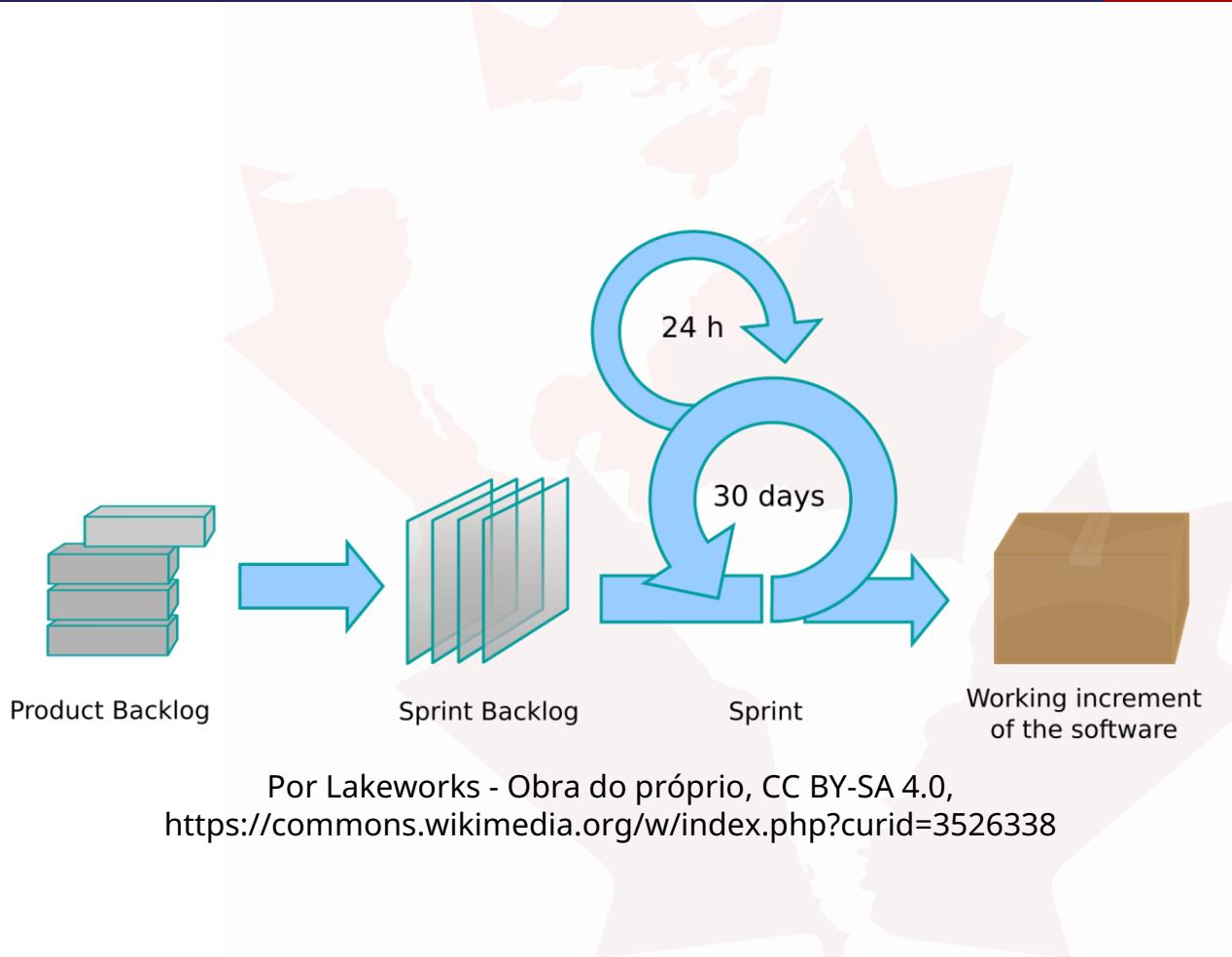
Metodologias Cíclicas



No mundo do software...



Por own work - Obra do próprio, Domínio público,
<https://commons.wikimedia.org/w/index.php?curid=1552158>



No mundo do DM e IA...

CRoss Industry Standard Process for Data Mining

Criado no anos 90

Método geral para conduzir projetos de Mineração de Dados ou de Ciência de Dados

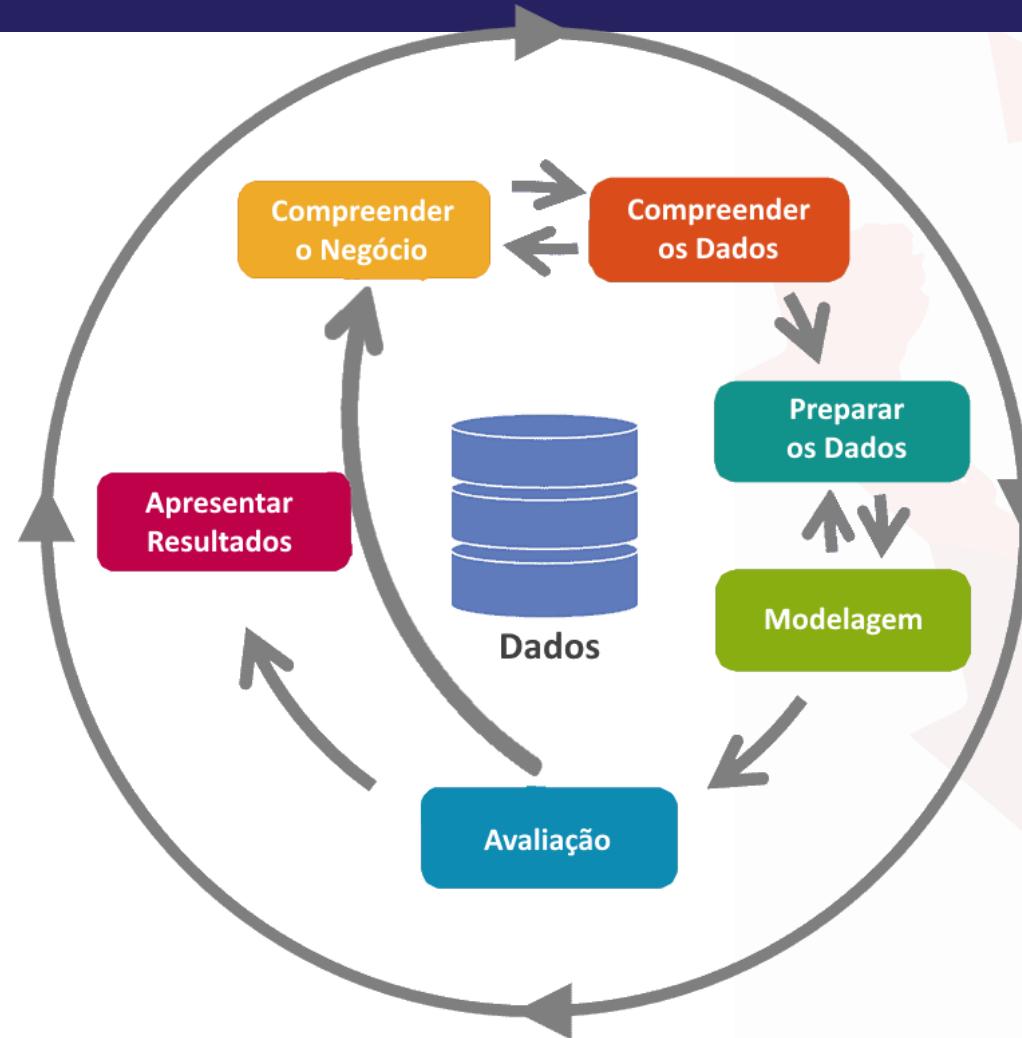
Dividido em 6 fases

Focado em iteração, melhoria incremental e entregas

Benefícios:

- Hipóteses organizadas
- Cíclico
- Flexível
- Agrega valor frequentemente

CRISP-DM



1. Compreender o Negócio



Quais são as necessidades do negócio?

- Primeira etapa do CRISP-DM
- A compreensão errada pode comprometer o sucesso.
 - Muito esforço empregado para produzir as respostas certas para as perguntas erradas
- O que você quer atingir sob uma perspectiva de negócios.
 - Compreender os OBJETIVOS da pessoa que está perguntando
 - Perguntas claramente definidas
 - Critérios de sucesso

1. Compreender o Negócio



O objetivo da Iniciativa

- Clareza sobre o que deve ser resolvido

As premissas

- Quais itens serão assumidos como verdadeiros

Os riscos envolvidos

- Quais riscos foram identificados e como mitiga-los?

Custo x Benefícios

- Quais os custos e benefícios esperados

Critérios de Sucesso

- **Quais métricas de negócio devem ser atingidas?**

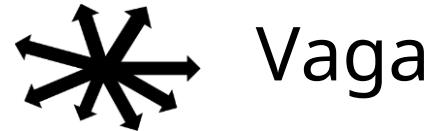
Planejamento da iniciativa

- Quais passos devem ser executados

Documentação inicial

- Documentar as definições acordadas

1. Perguntas vagas x objetivas



Vaga

Não é respondida por um nome ou um número:

- Como a IA pode me ajudar?
- O que meus dados podem dizer sobre o meu negócios?
- O que devo fazer?
- Como posso aumentar meu lucro?



Objetiva

Deve ser respondida com um nome ou um número:

- Quantas vezes esta nova função será utilizada por novos usuários?
- Qual caminho para através do centro me fará chegar mais rápido ao trabalho?

2. Compreender os dados

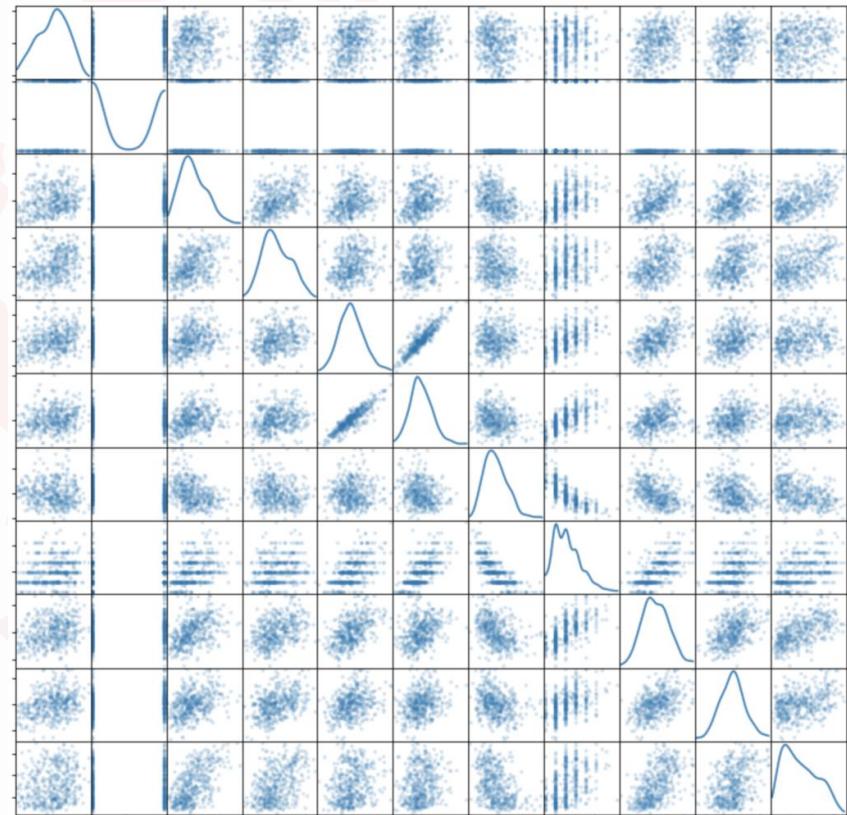


Quais dados precisamos?

Quais dados possuímos?

Eles estão “limpos”?

- Obter os dados preliminares
- Descrever os dados obtidos
- Explorar os dados (estatística descritiva)
- Incorporar novos dados
- Analisar a qualidade dos dados



2. Compreender os dados



Procure responder as seguintes perguntas:

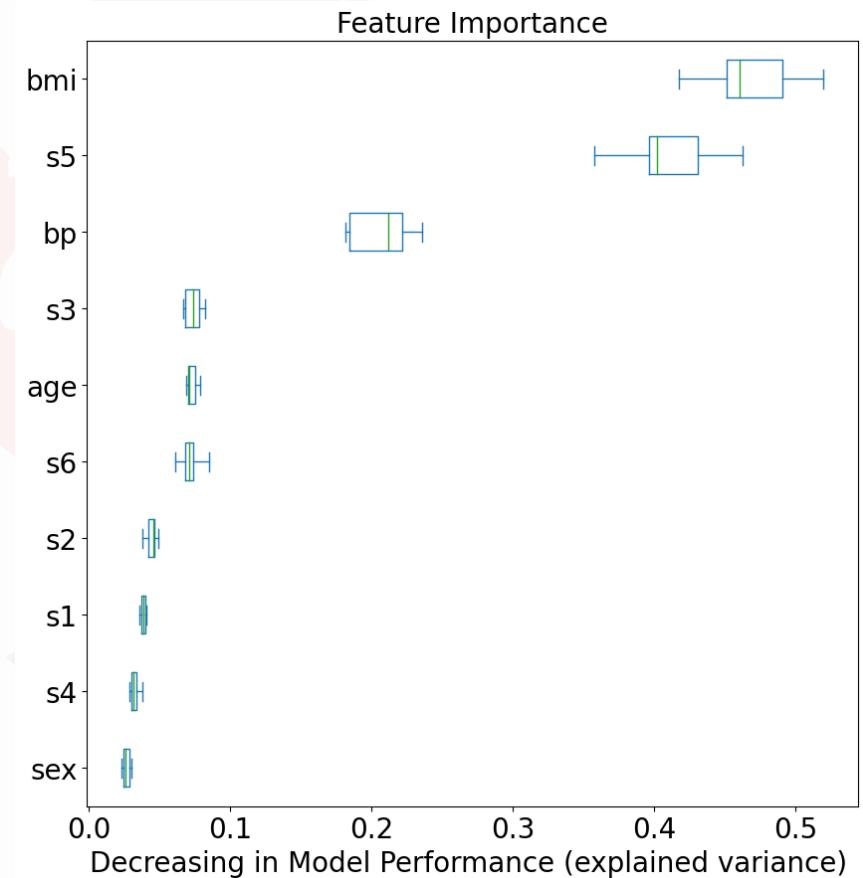
- Os dados estão estruturados?
- A empresa tem banco de dados? Ou em um data lake?
- Os dados serão acessados de que forma?
- Existem regras de compliance que devem ser respeitadas?
- Quantas fontes de dados serão utilizadas?
- Quais serão os formatos dos dados?

3. Preparar os dados



Como vamos organizar os dados para a fase de modelagem?

- Selecionar os dados
- Limpar os dados
- Construir dados derivados
- Incorporar novos dados
- Modificar o formato dos dados



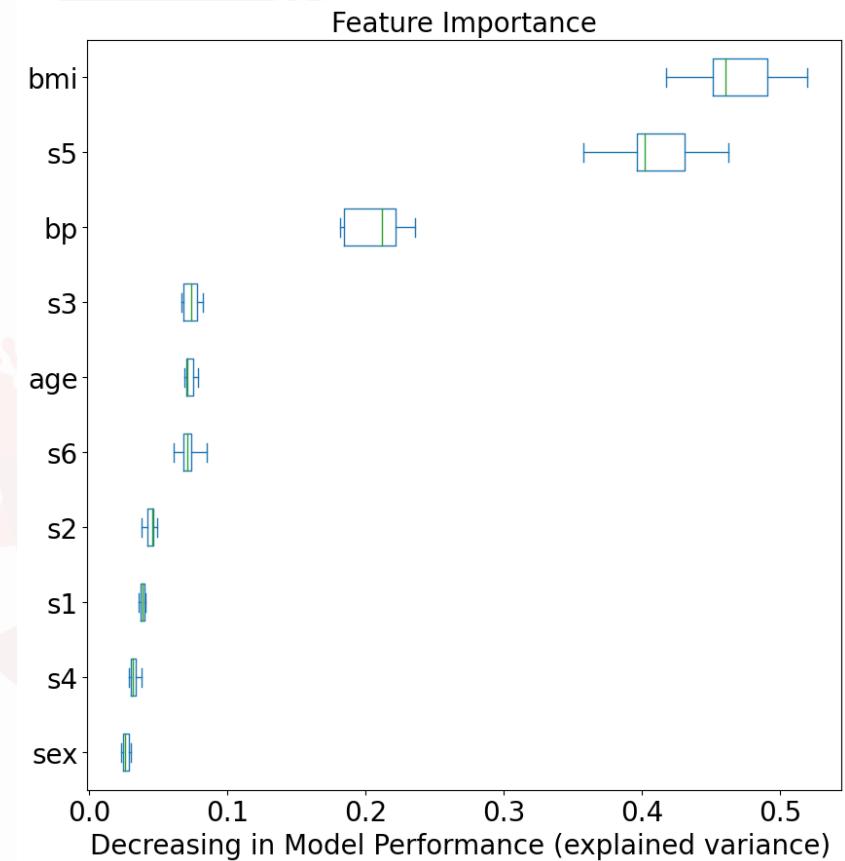
3. Preparar os dados



Data
Preparation

Algumas perguntas que precisamos responder:

- Como os valores nulos devem ser tratados?
- Os atributos estão nos formatos corretos?
- Será necessário enriquecer os dados?
- Quais variáveis serão utilizadas na modelagem?



4. Modelagem



Quais técnicas de modelagem vamos utilizar?

- Selecionar o(s) algoritmo(s)
- Projetar o teste
- Criar o modelo
- Interpretar os resultados

name	class	transform	prior	trainable	shape	dtype	value
GPR.kernel.kernels[0].variance	Parameter	Softplus		True	()	float64	26876.1
GPR.kernel.kernels[0].lengthscales	Parameter	Softplus		True	()	float64	8.45629
GPR.kernel.kernels[1].variance	Parameter	Softplus		True	()	float64	2444.36
GPR.likelihood.variance	Parameter	Softplus + Shift		True	()	float64	381.861

5. Avaliação



Quais resultados atingem melhor os objetivos de negócios?

- Coletar métricas
- Avaliar resultados em relação aos objetivos de negócios
- Rever as etapas
- Determinar os próximos passos

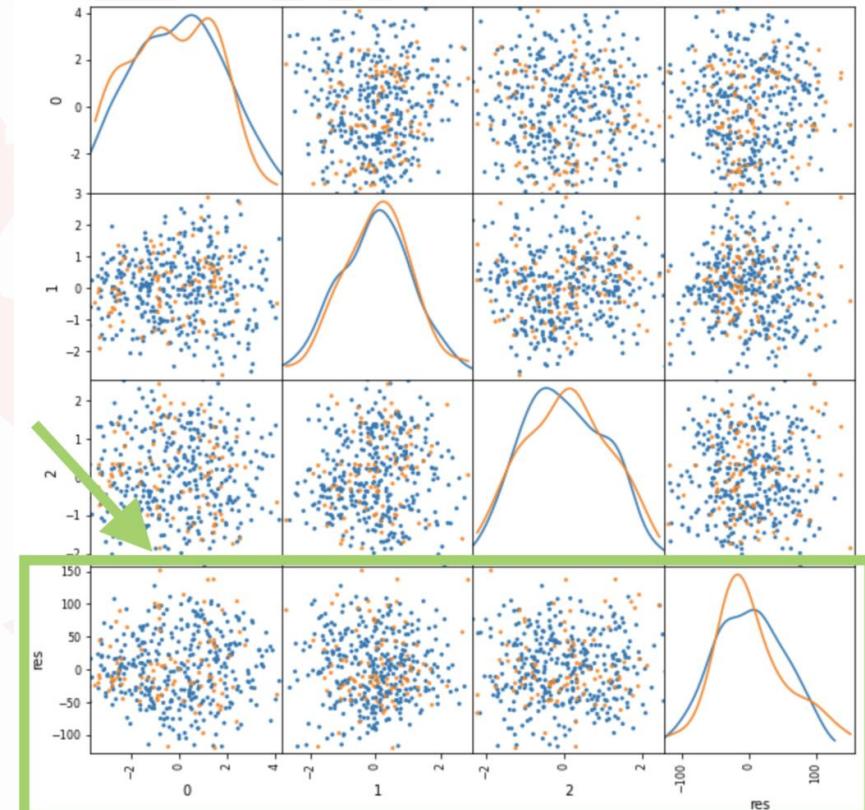
5. Avaliação



Evaluation

Algumas perguntas que precisamos responder:

- O modelo é robusto?
- Atende às expectativas?
- É eficaz?
- É confiável?
- É escalável?
- Produz resultados dentro das expectativas de performance?



6. Apresentar Resultados



Tornar públicas as conquistas realizadas

- Formatar as visões e insights obtidos
- Apresentar aos stakeholders
- Planejar a evolução
- Implantar?

- É a hora de gerar valor para o negócio.

6. Apresentar Resultados

The screenshot shows a REST API documentation interface with the URL `0.0.0.0/docs#/default/predict_predict_post` in the address bar. The page title is `default`. There are two main sections: a `GET` method for `/ Home` and a `POST` method for `/predict Predict`.

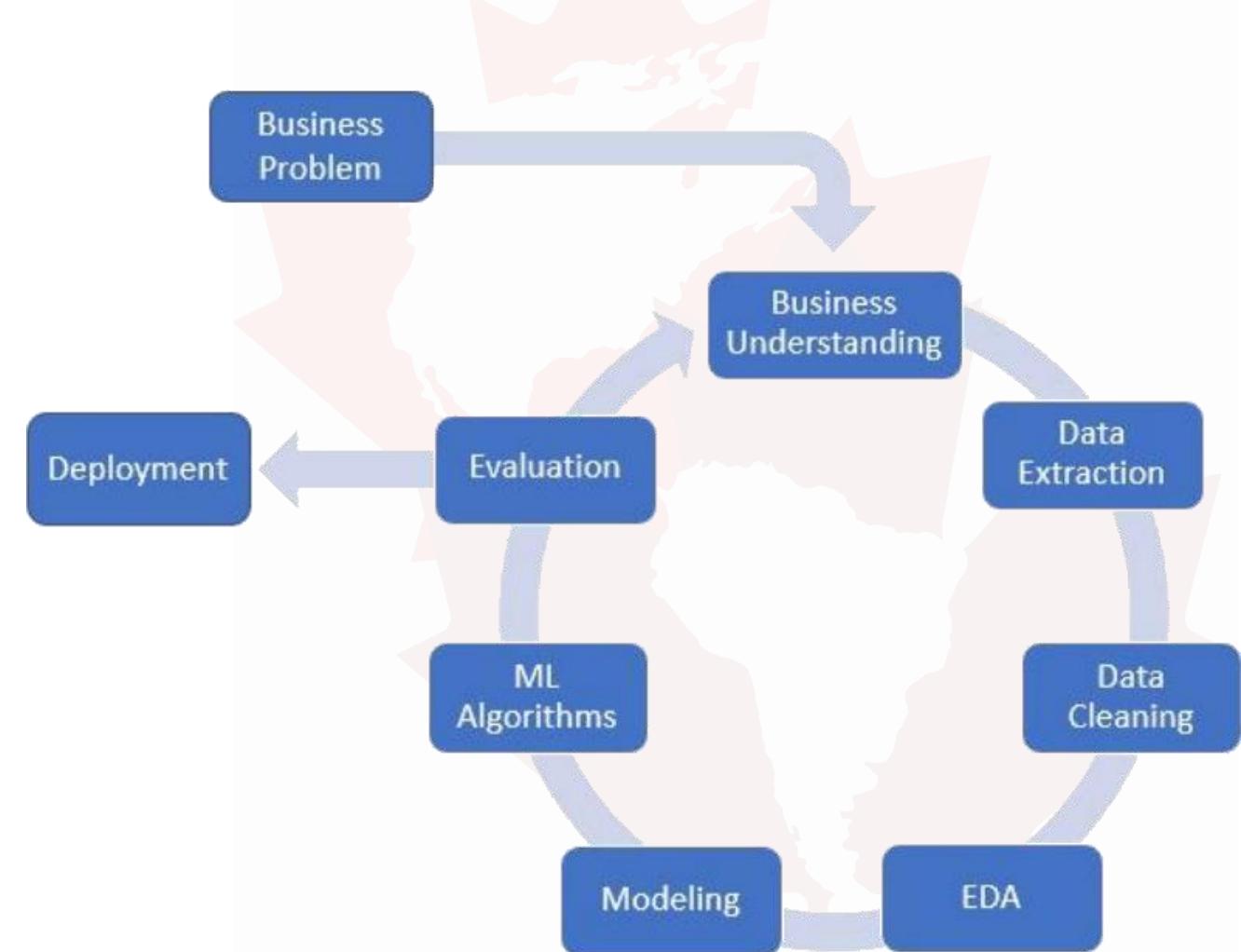
The `POST` method section includes a `Parameters` section which states "No parameters". It also includes a `Request body` section with the `required` status, a `application/json` content type dropdown, and a JSON placeholder containing the following data:

```
{  
    "age": 0.038076,  
    "sex": 0.050680,  
    "bmi": 0.061696,  
    "bp": 0.021872,  
    "s1": -0.044223,  
    "s2": -0.034821,  
    "s3": -0.043401,  
    "s4": -0.002592,  
    "s5": 0.019908,  
    "s6": -0.017646  
}
```

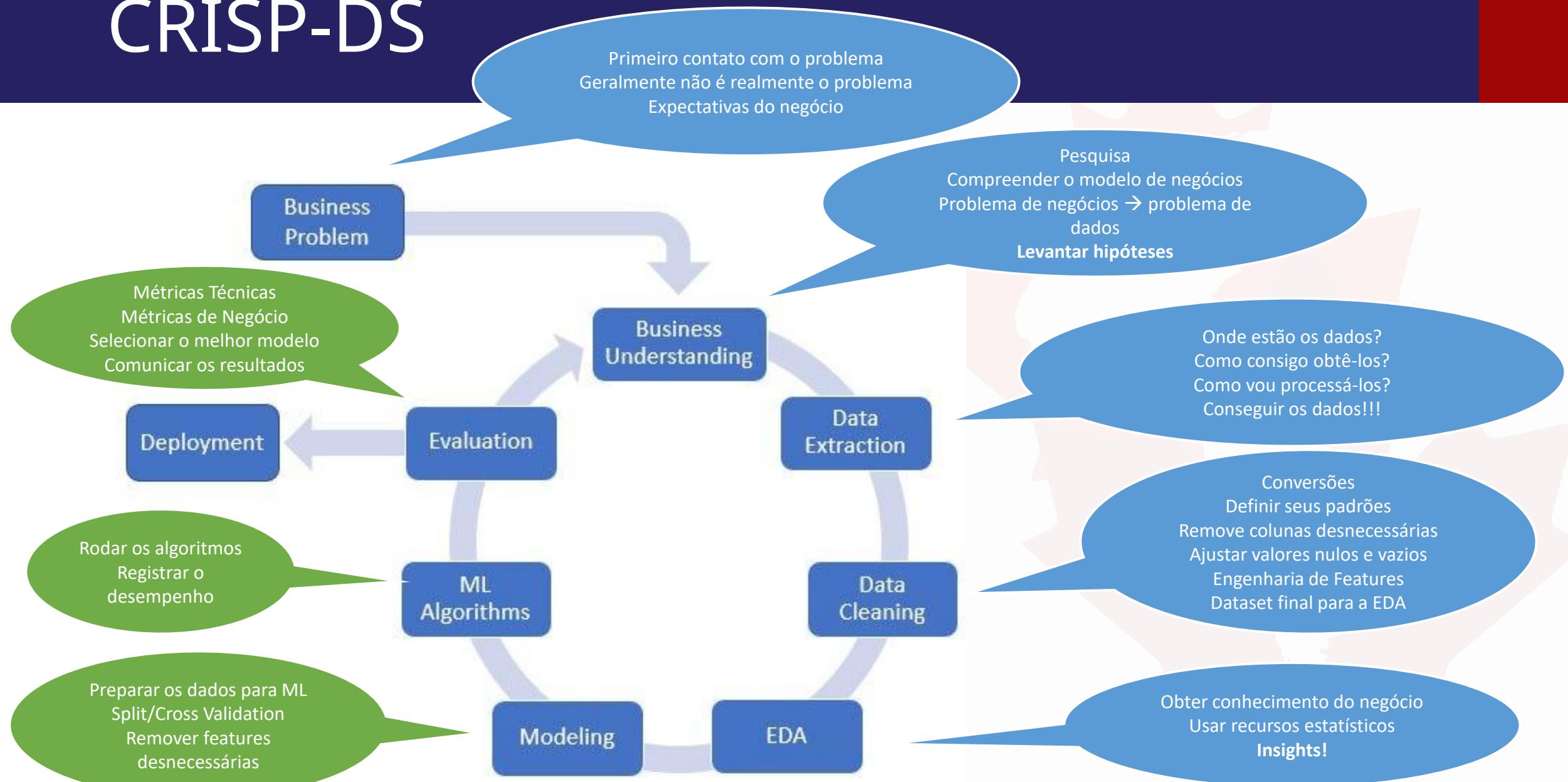
At the bottom of the `POST` section are `Execute` and `Clear` buttons.

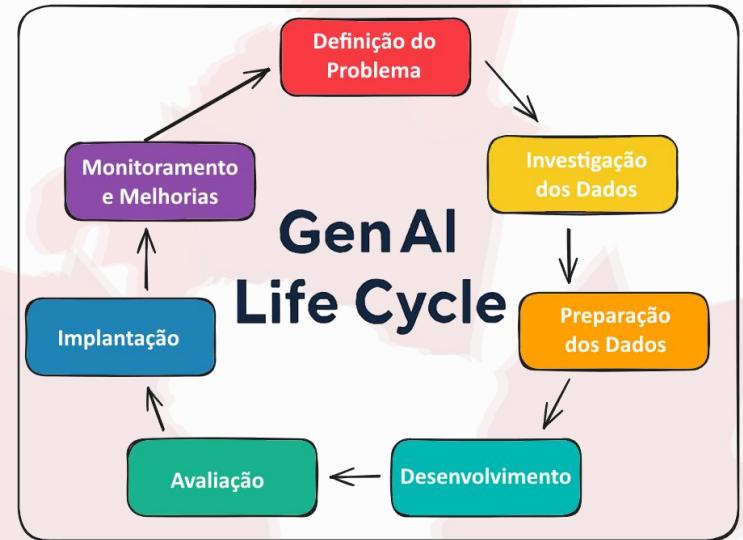
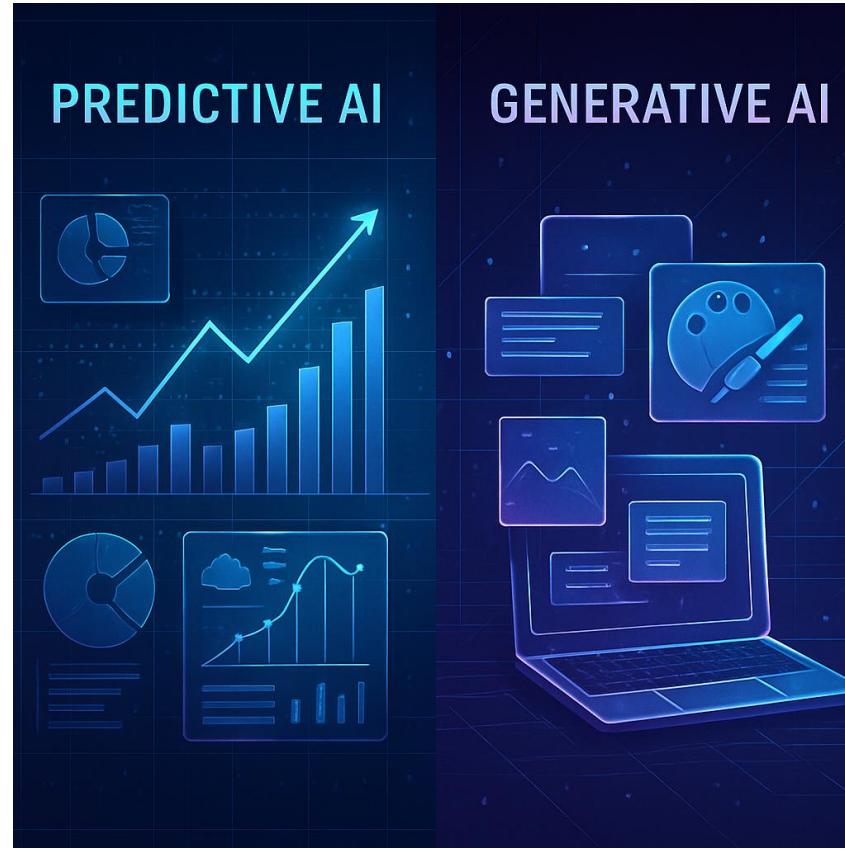
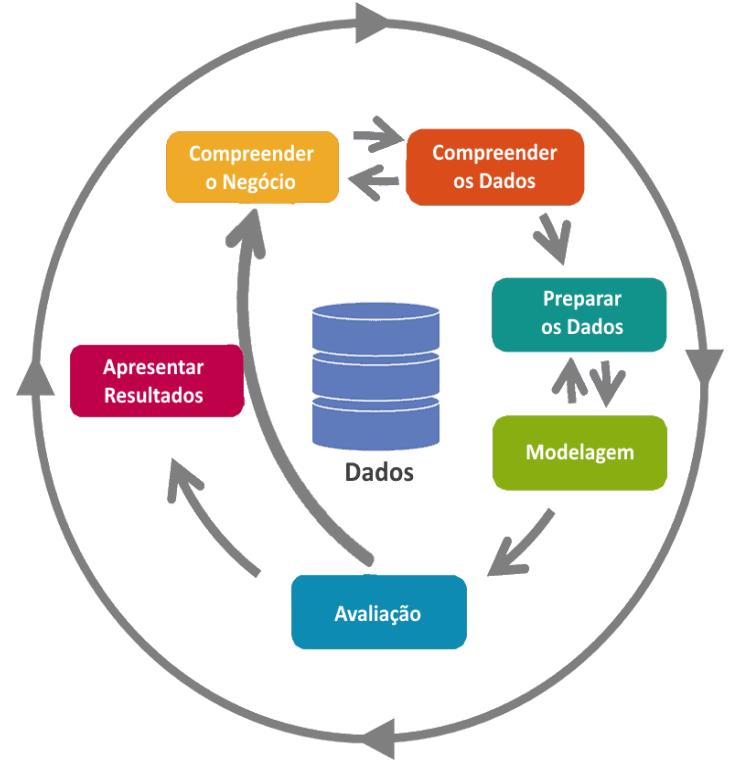
CRISP-DS

Adaptions



CRISP-DS





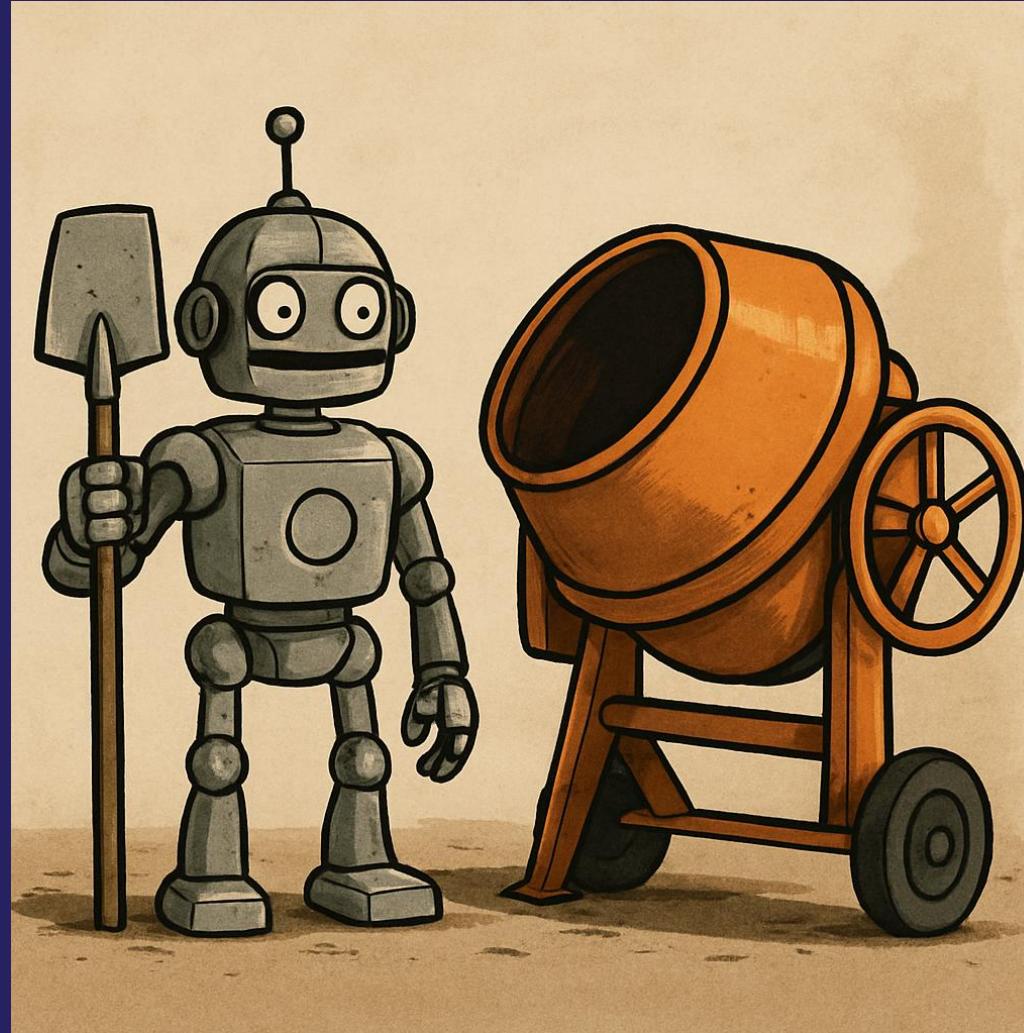
Diferença entre as abordagens

Predictive AI

- Foco em **predições**
- Treinada a partir de dados históricos
- Utilizada em aplicações de negócios
- Necessita de robustez
- Ajustada ao MLOps clássico

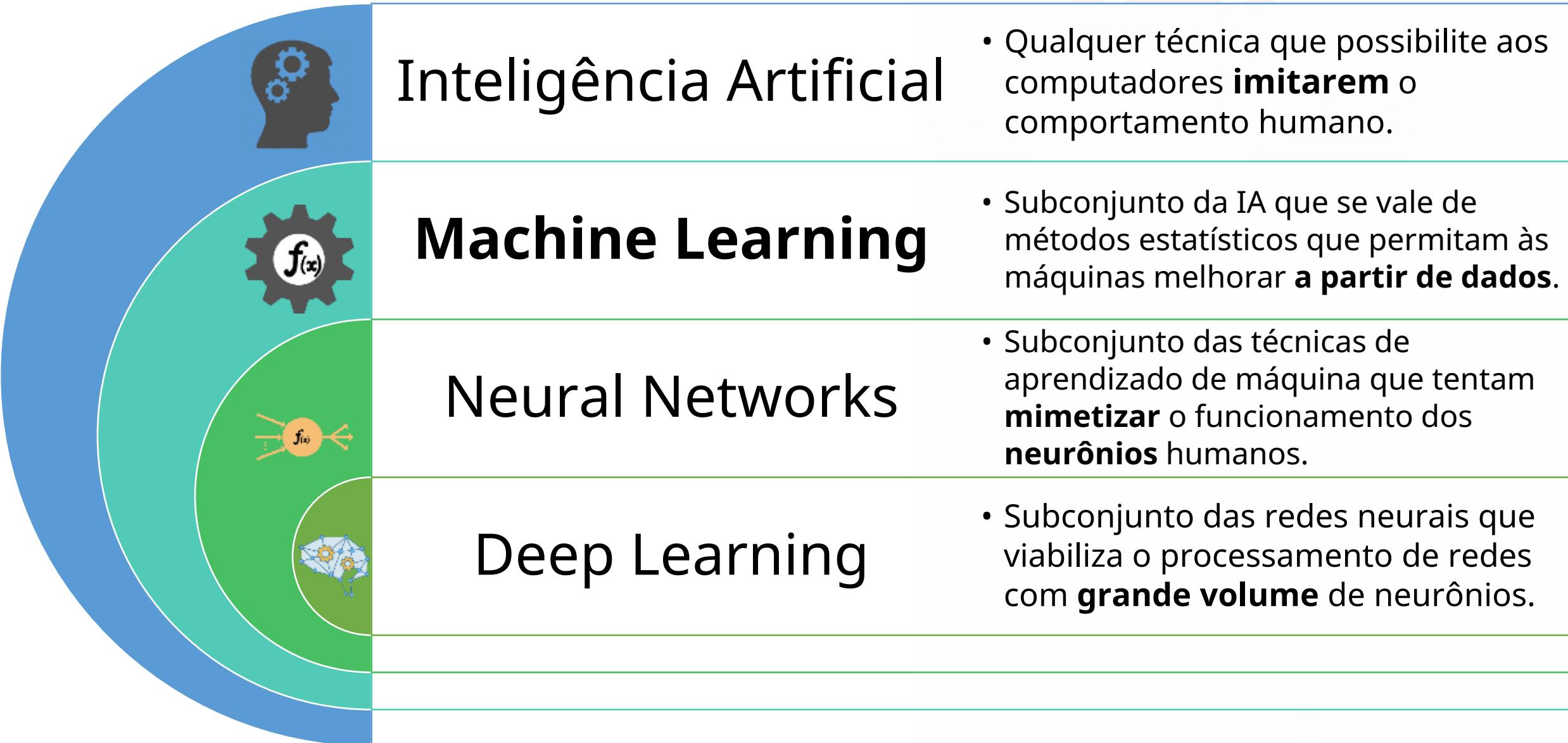
Generative AI

- Foco em **criar novos conteúdos**
- Treinada a partir de um enorme dataset
- Usada em aplicações criativas
- Necessita de guard rails
- Nuances distintos de MLOPs



Como criar modelos de ML ?

O que é Inteligência Artificial?



A Visão Geral da Ciência de Dados

Créditos:



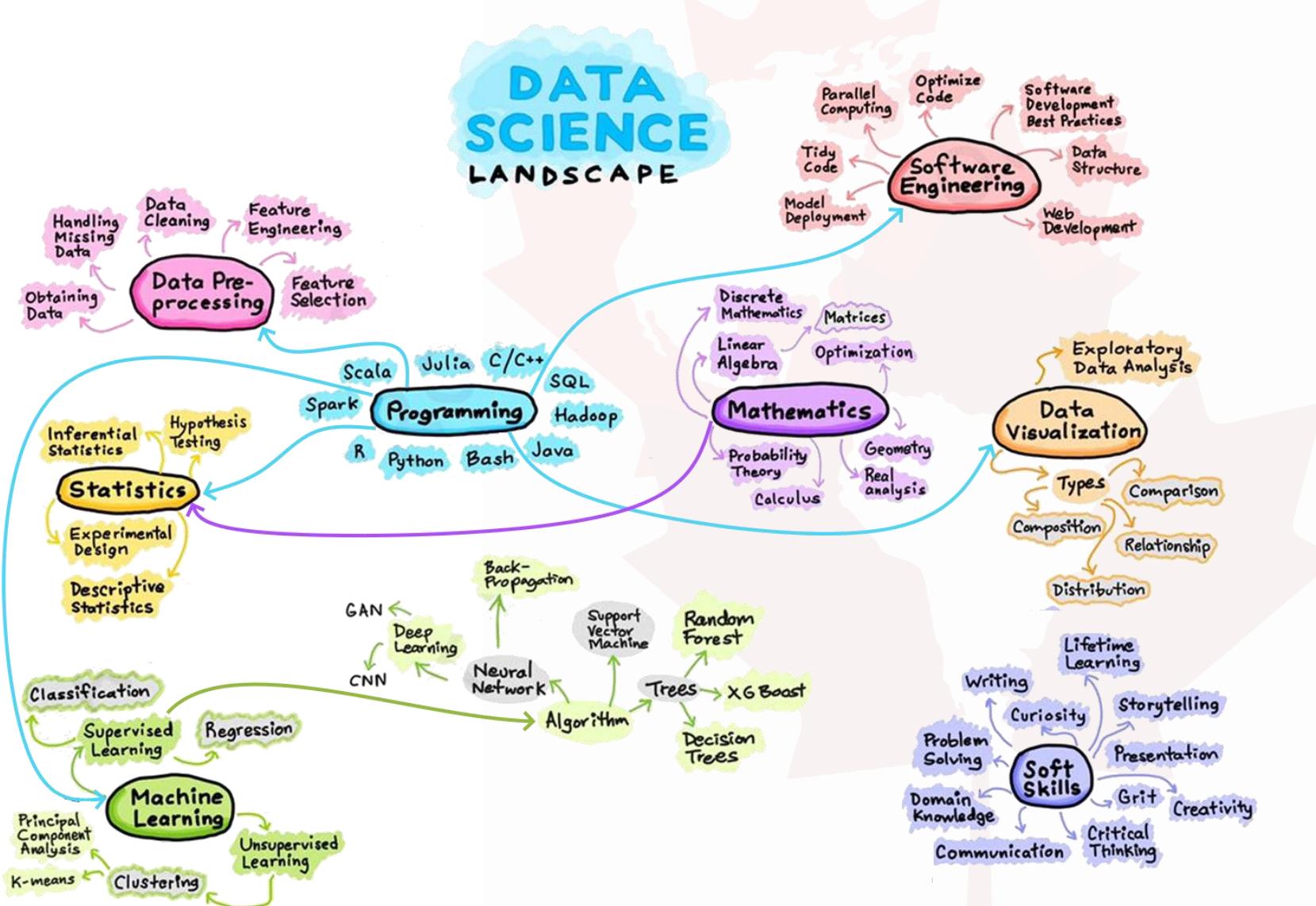
Chanin Nantasenamat

<https://www.instagram.com/data.professor/>

By: Chanin Nantasenamat

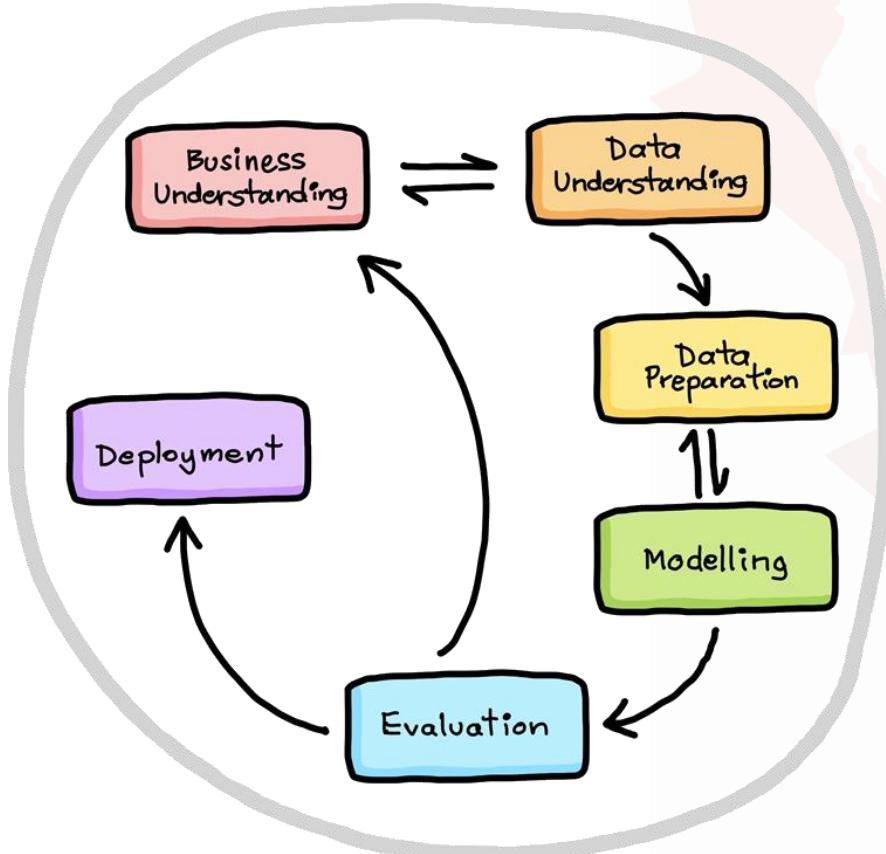
DATA PROFESSOR

<http://youtube.com/dataprofessor>



O Framework

CRISP-DM



Fase 2 – Data Understanding

Análise Exploratória de Dados

Obtenção
dos Dados

Pré-processamento
dos Dados

Análise e
Insights

Visualizações

Apresentação
de Resultados

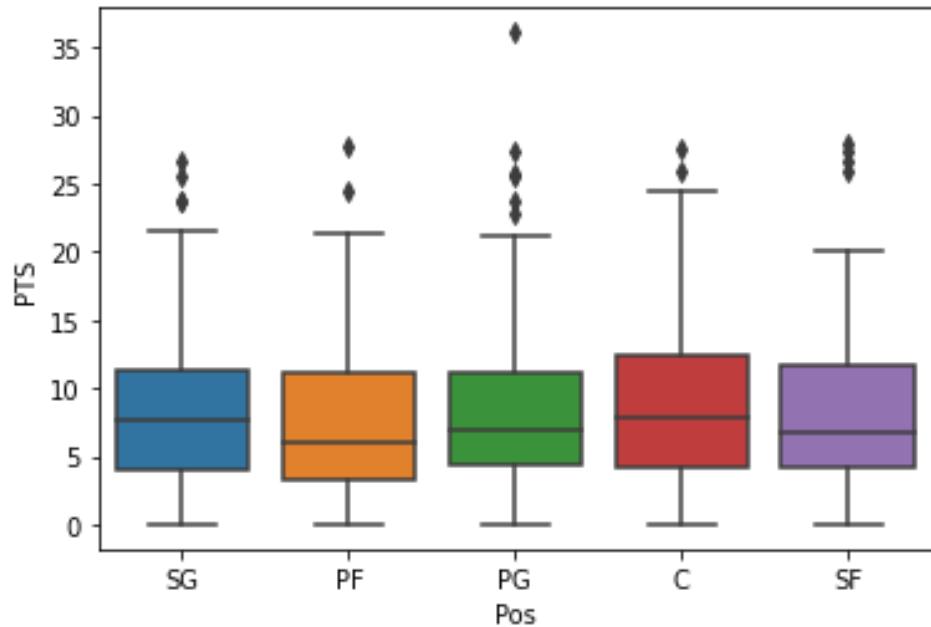


Pré-processamento

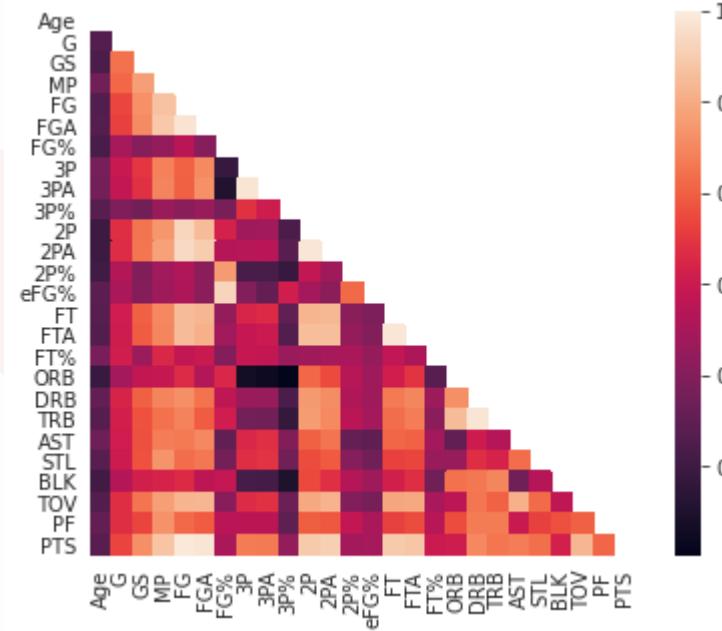
- Raramente os dados brutos podem ser utilizados diretamente
- Necessidade de pré-processamento para adequá-los
- Dados com baixa qualidade podem comprometer os resultados
- É importante entender a natureza do problema para guiar esta etapa



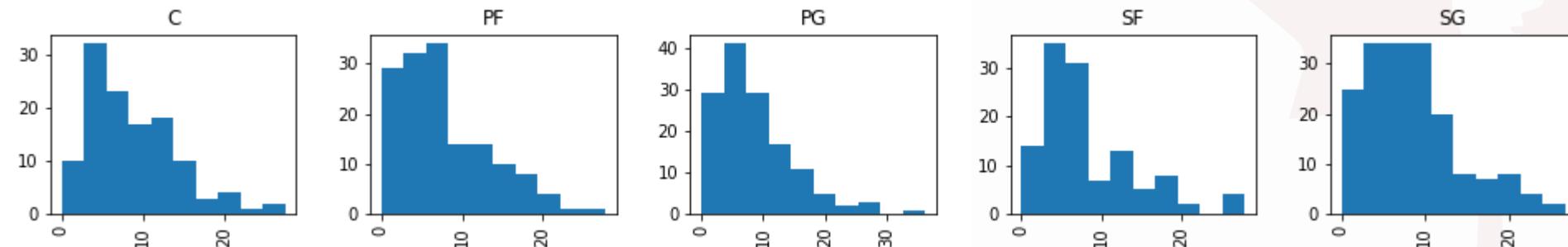
Análise Exploratória dos Dados



Example box plot of NBA player stats data.



Example correlation heatmap of NBA player stats data.

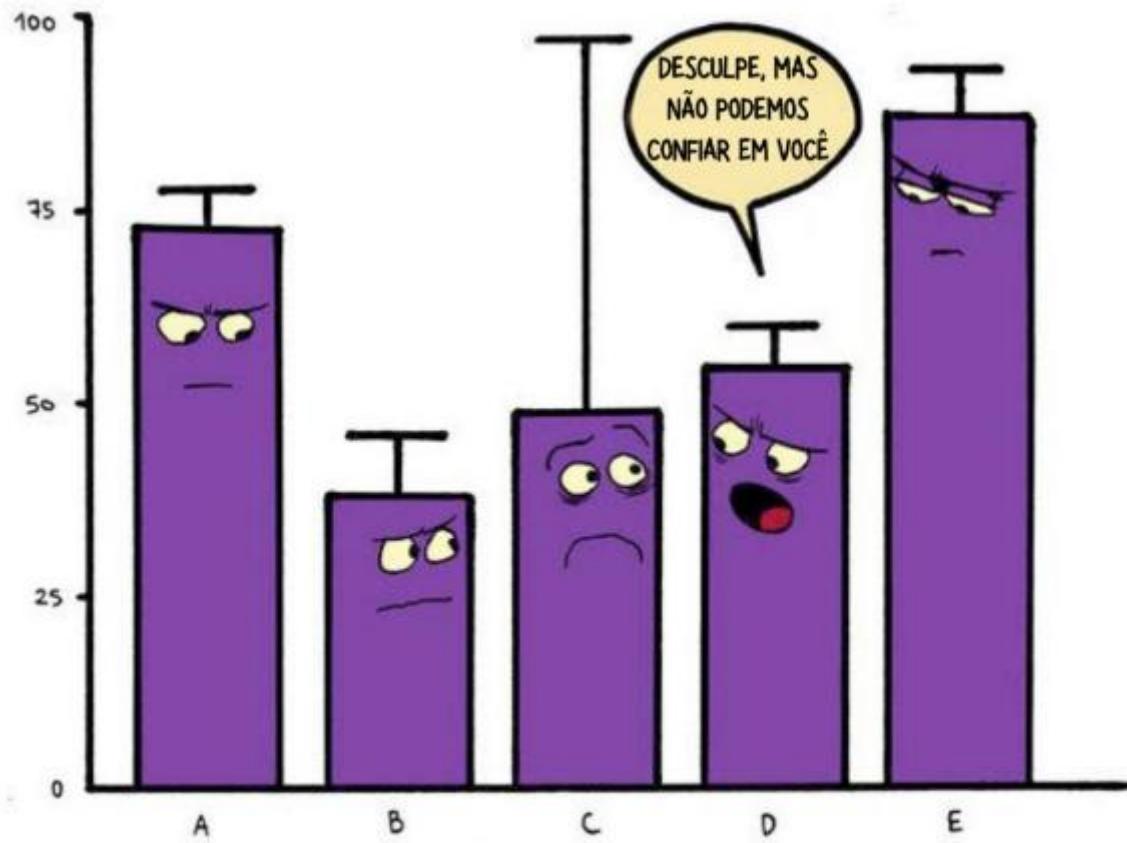


Example histogram plot of NBA player stats data.

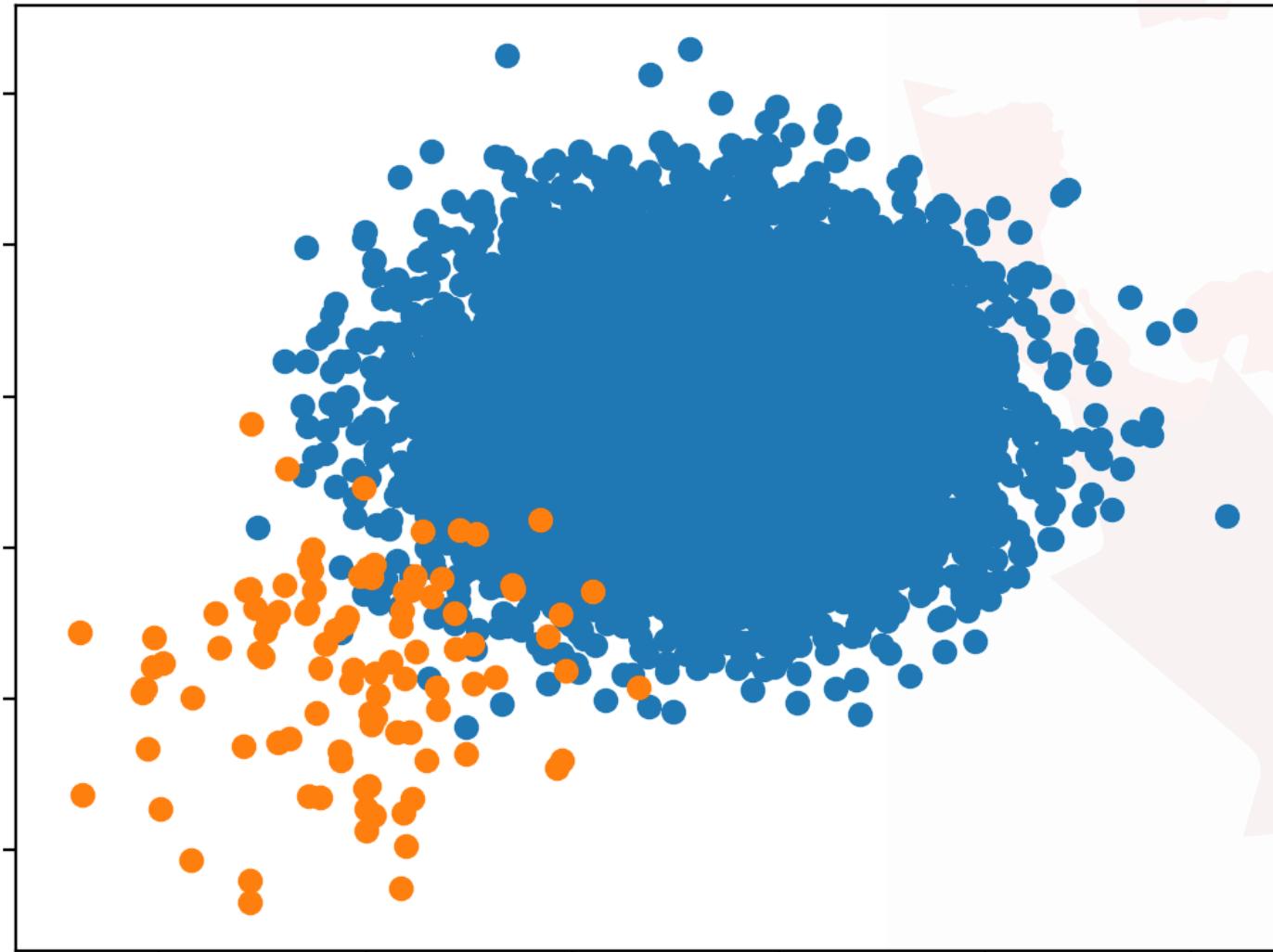
Dados Ausentes

- Como representar adequadamente os dados ausentes?
 - Ex. Qual é o ano da morte de uma pessoa viva?
- O que fazer com um campo deixado em branco ou preenchido com um valor obviamente estranho?
- Ou eventos raros demais para serem notados?
- Definir dados ausentes como **zero** é geralmente **errado**

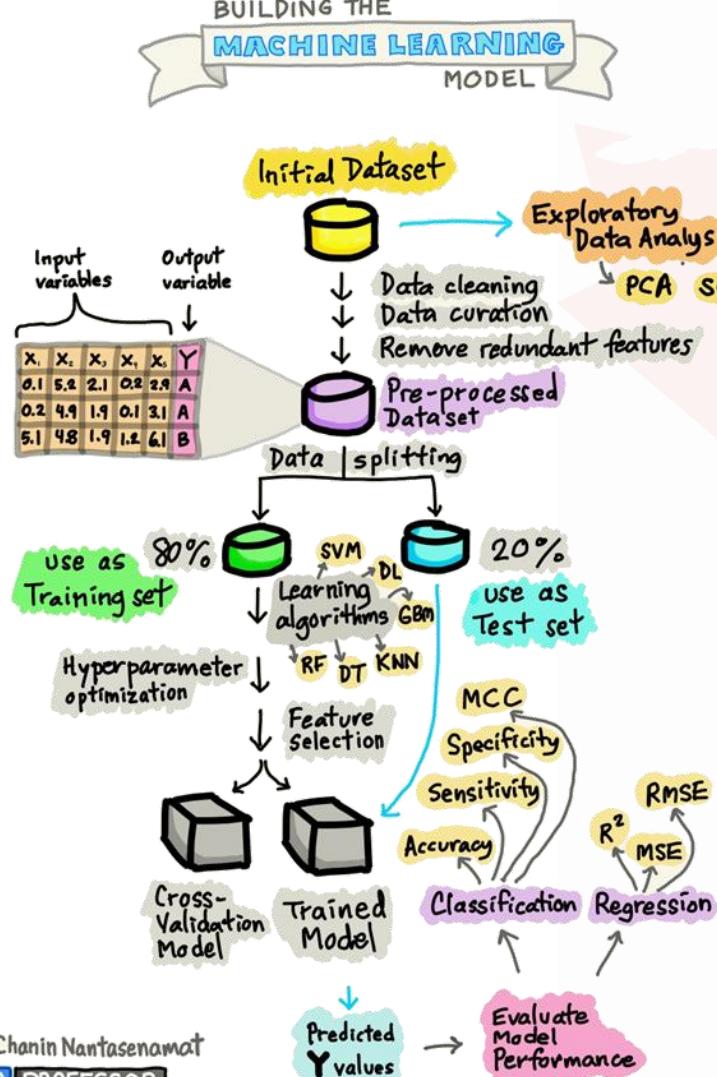
Outliers



Dados desbalanceados



Fases 3, 4 e 5 – Criando o Modelo



By Chanin Nantasenamat

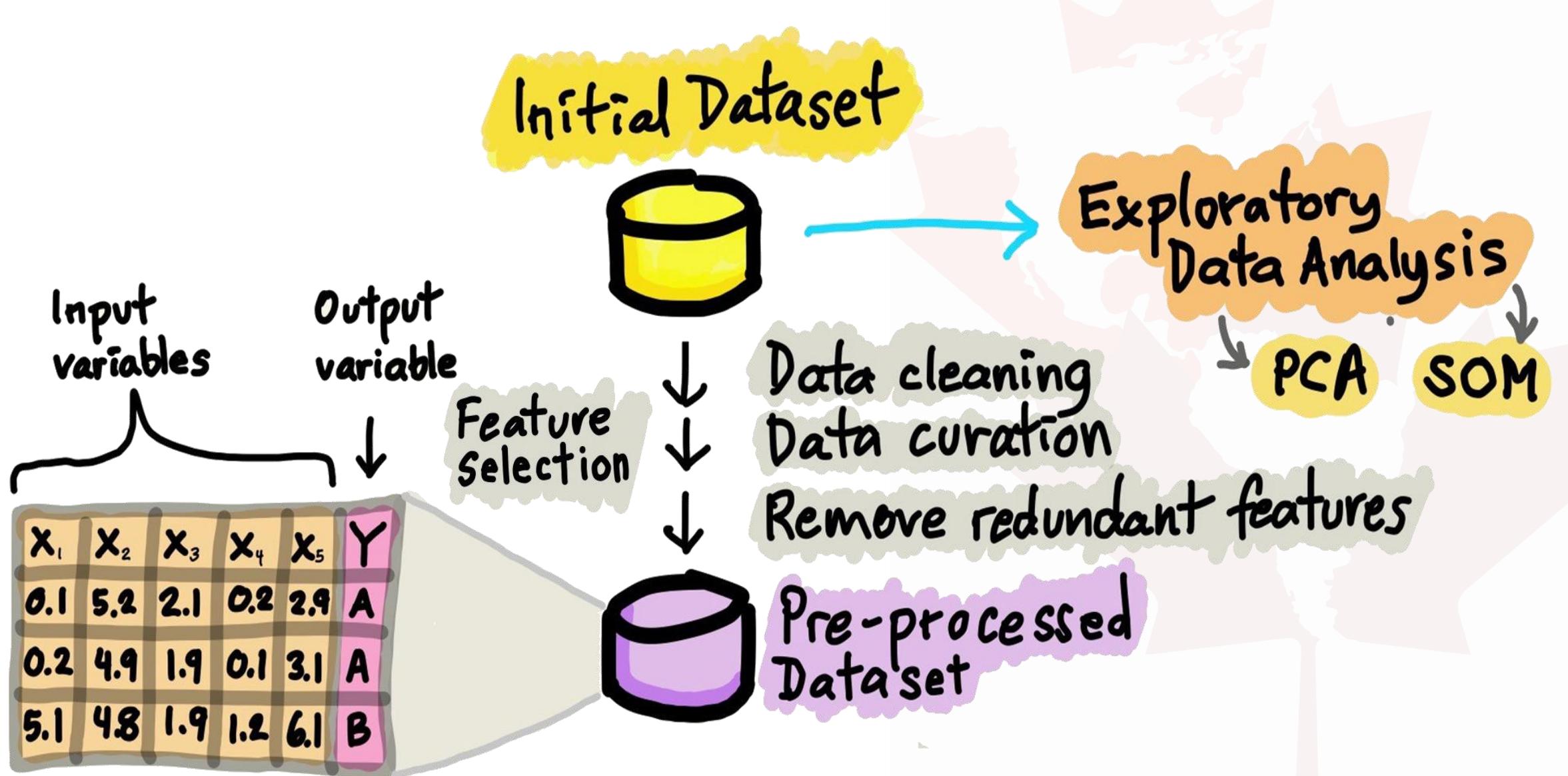
DATA PROFESSOR



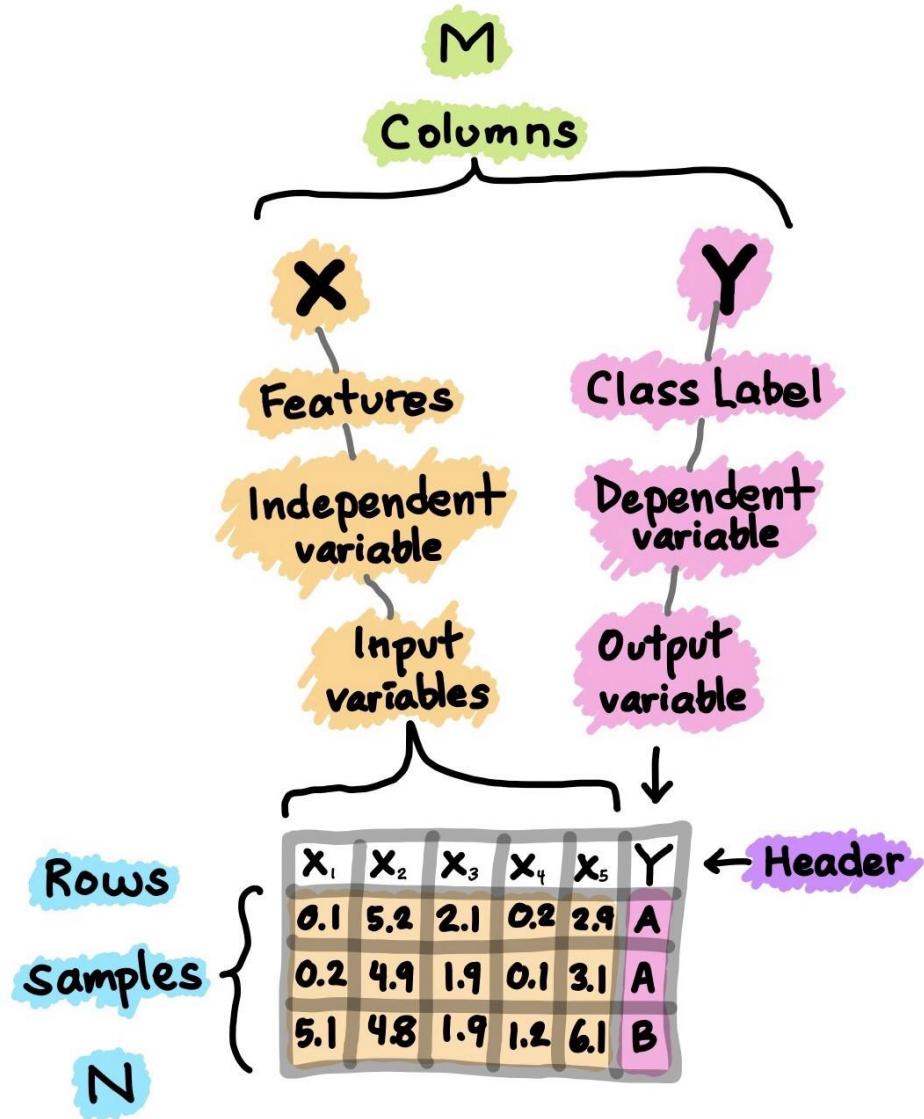
<http://youtube.com/dataprofessor>

January 1, 2020

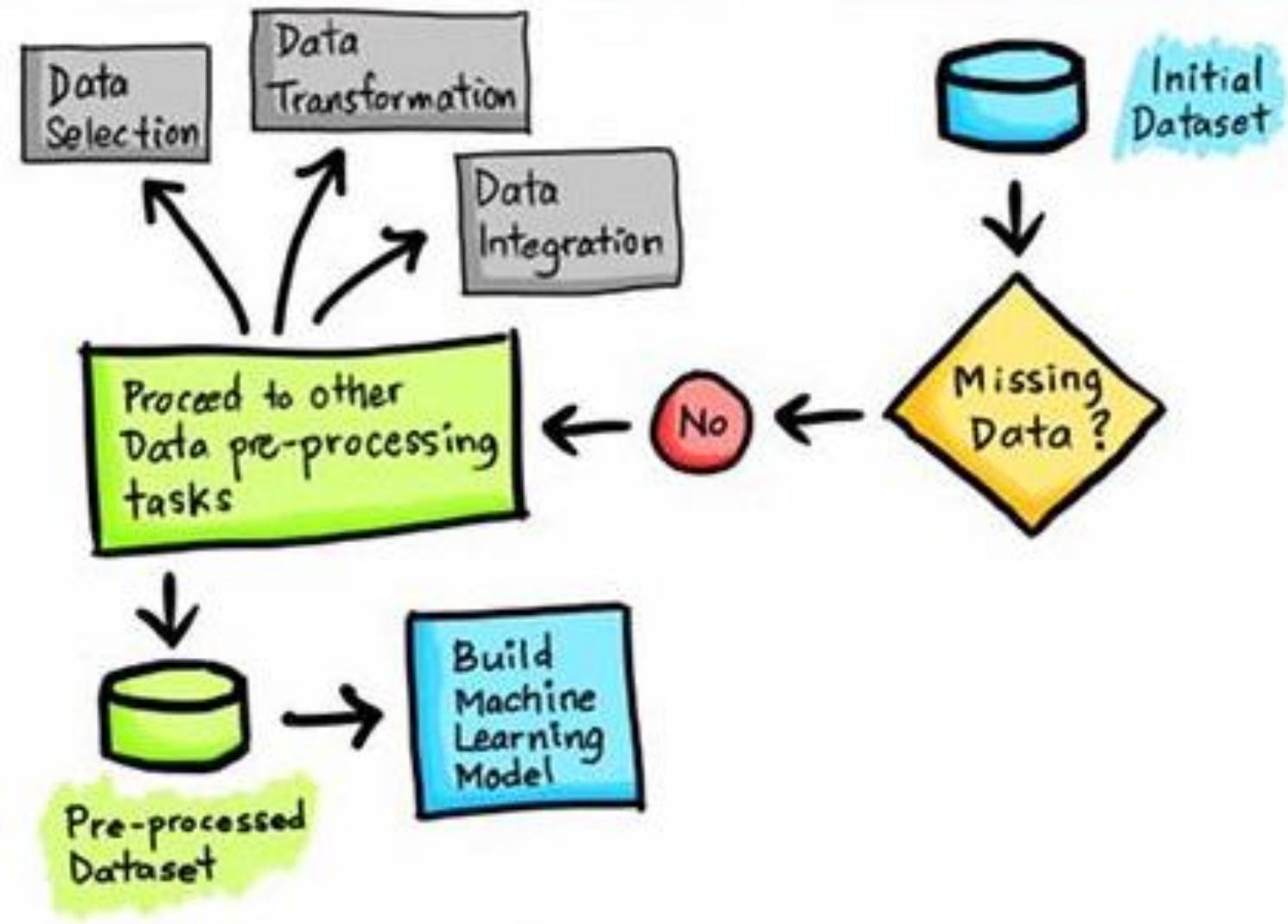
Fase 3 - Preparando o Dataset



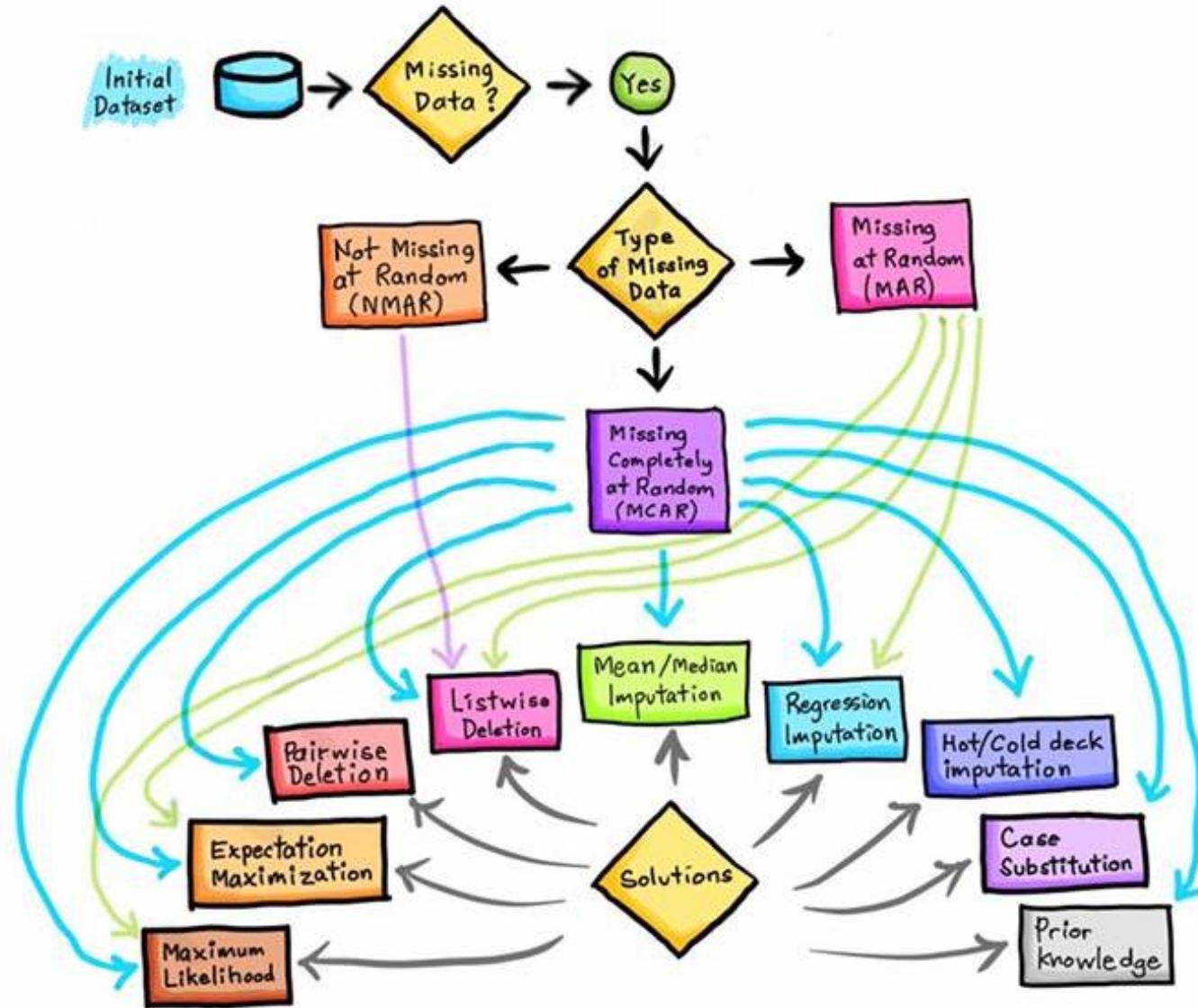
Preparando o Dataset



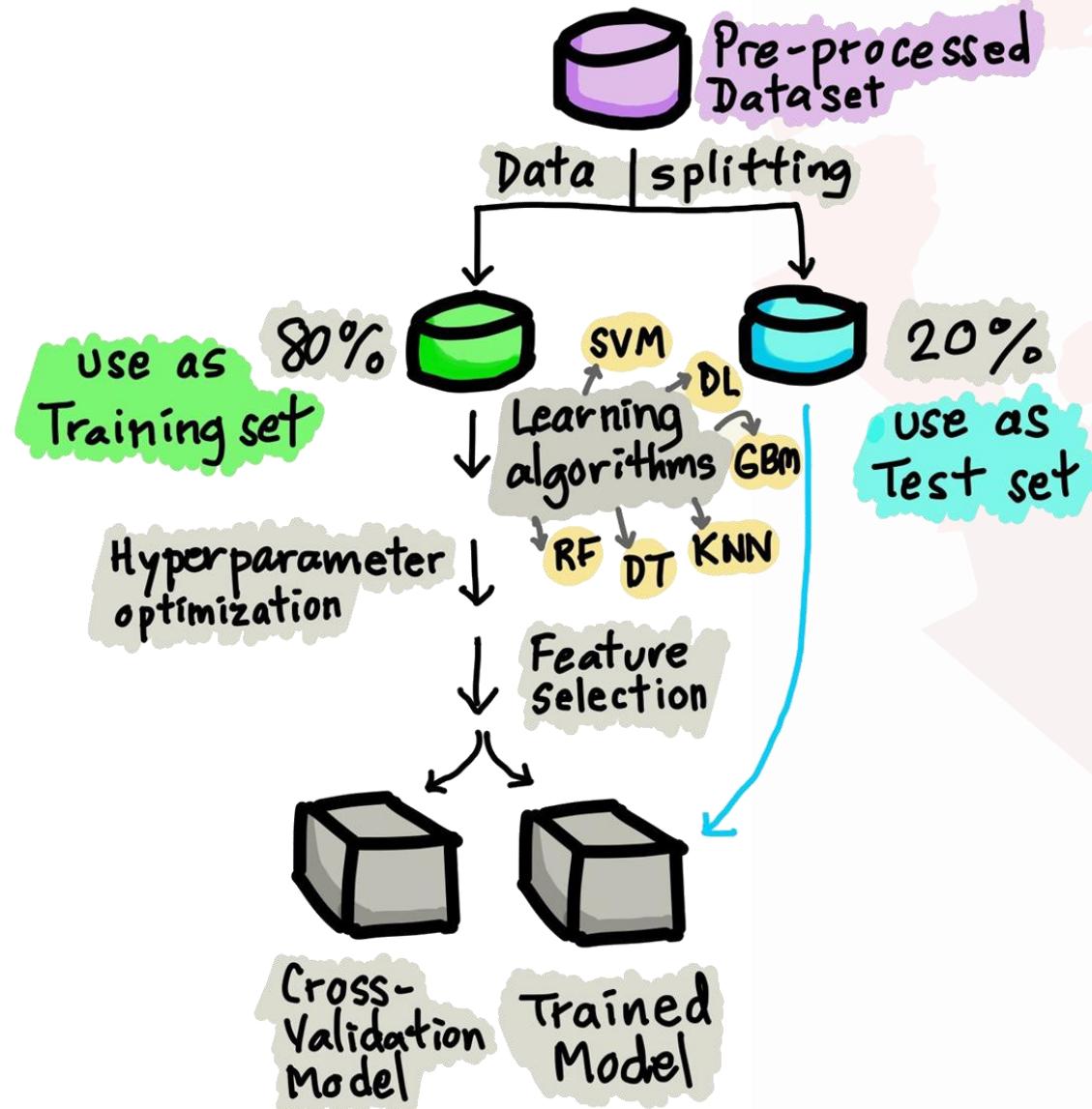
Criando o Dataset



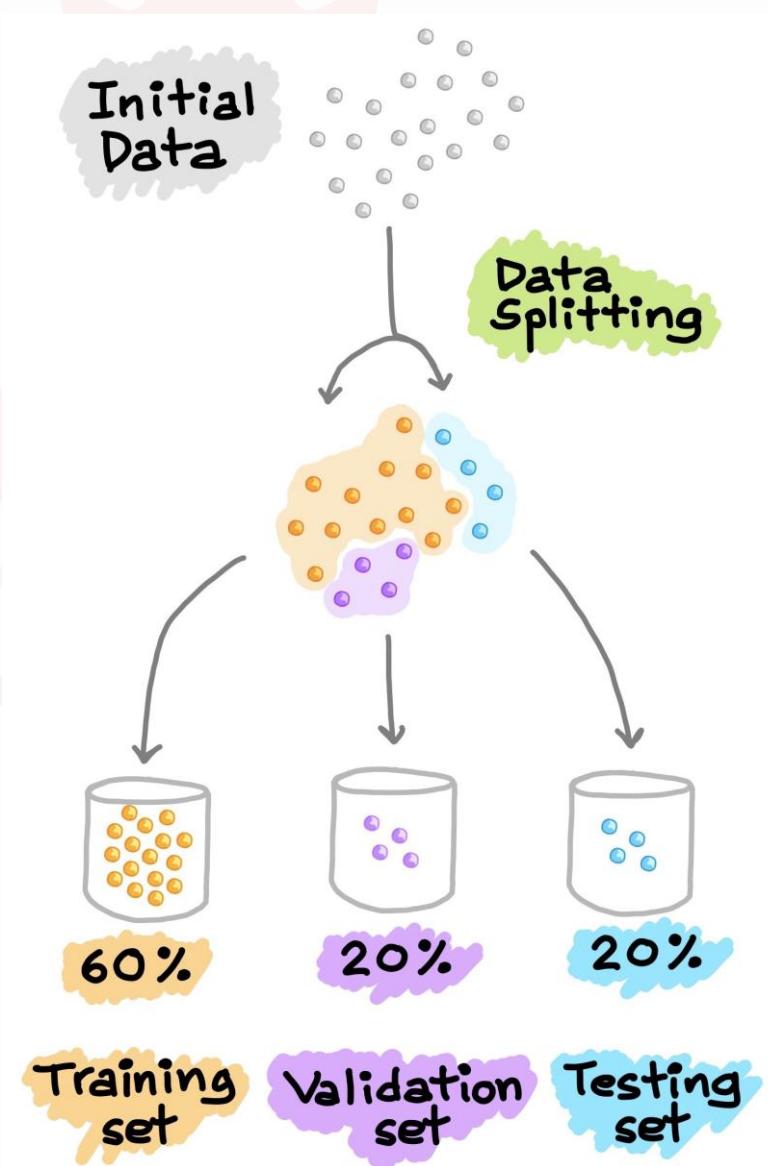
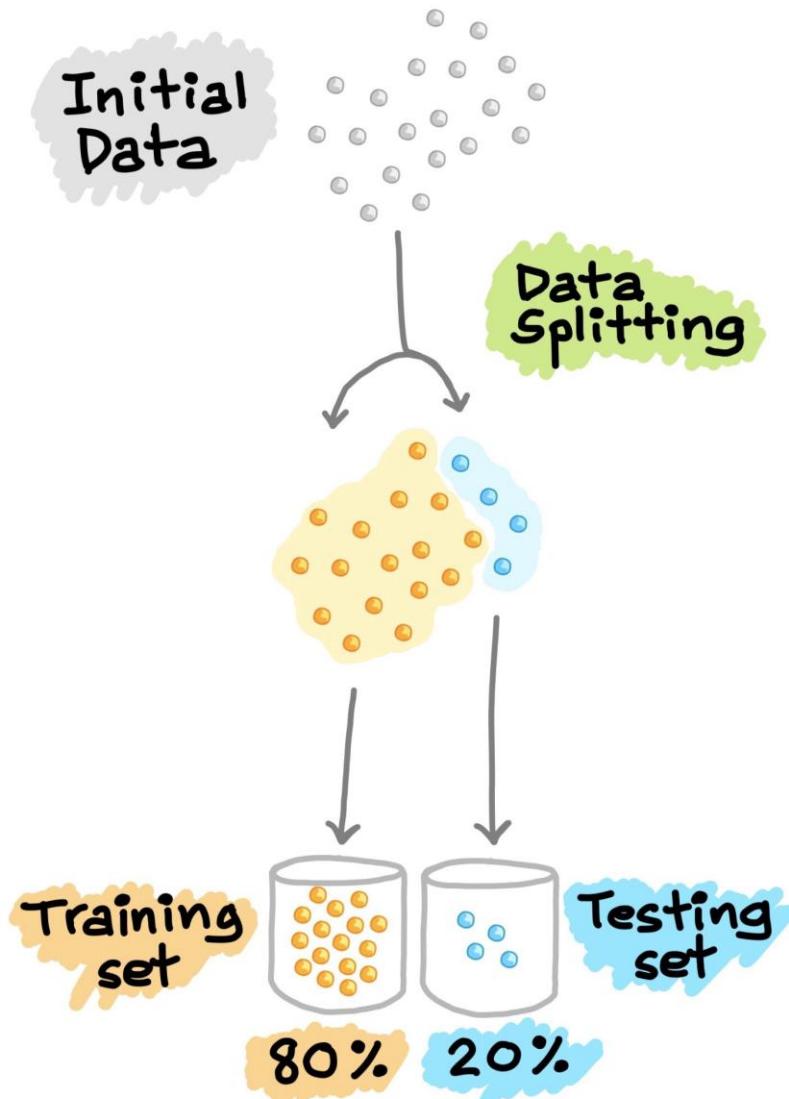
Tratando dados ausentes



Criando o modelo de ML

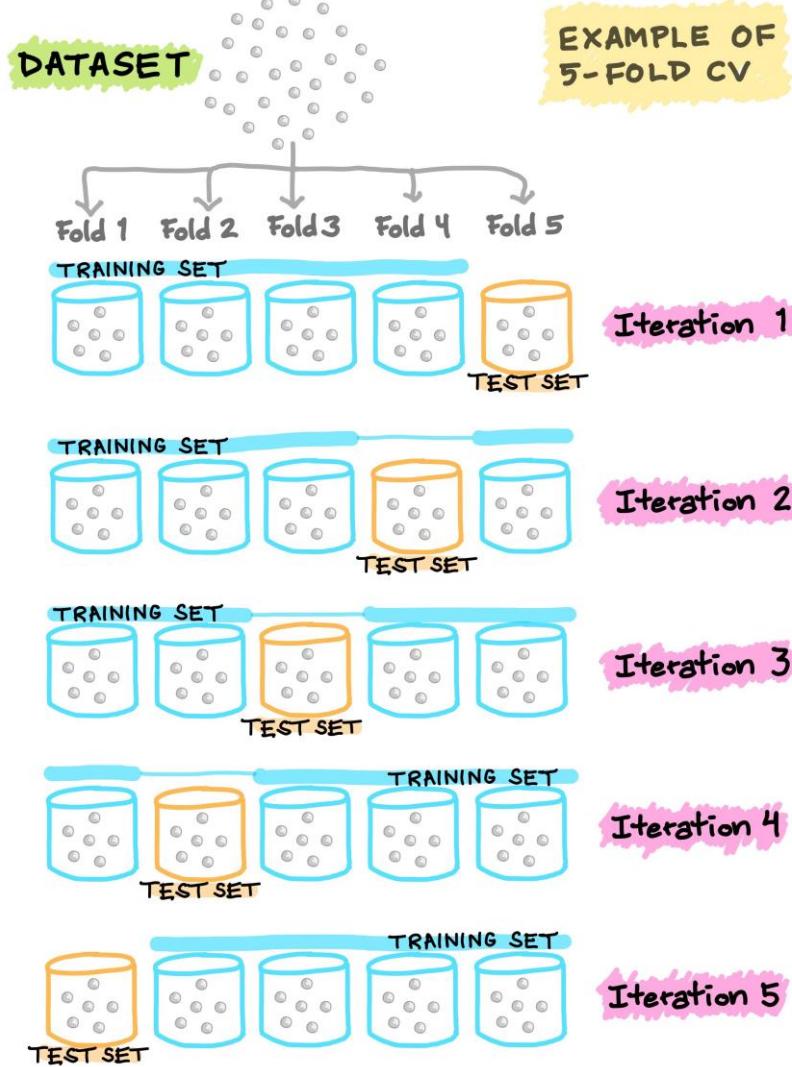


Preparando os Datasets

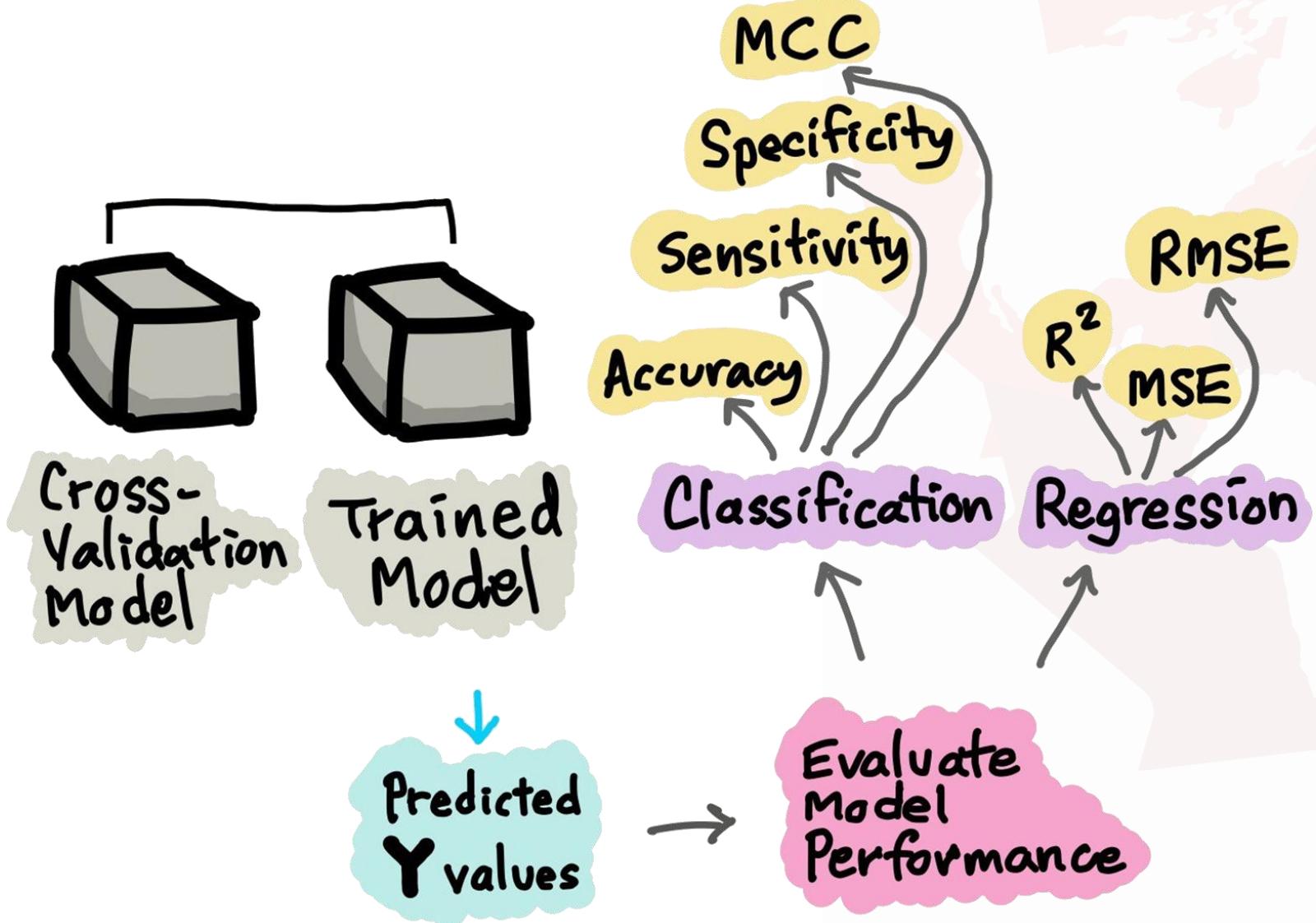


Cross-Validation

CROSS-VALIDATION



Fase 4 – Criando o modelo



Machine Learning: Tipos de Treinamento



Supervised Learning

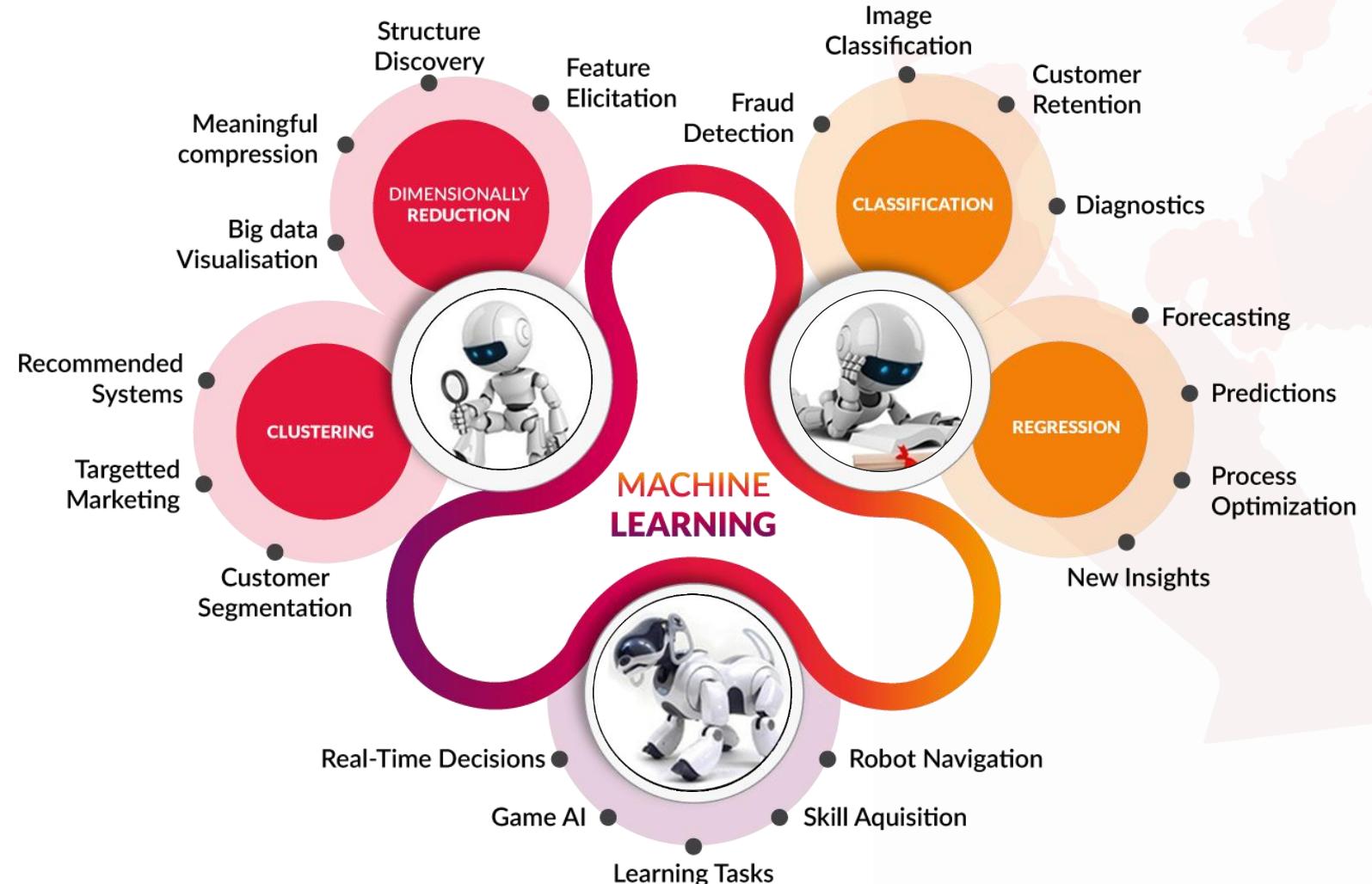


Unsupervised Learning

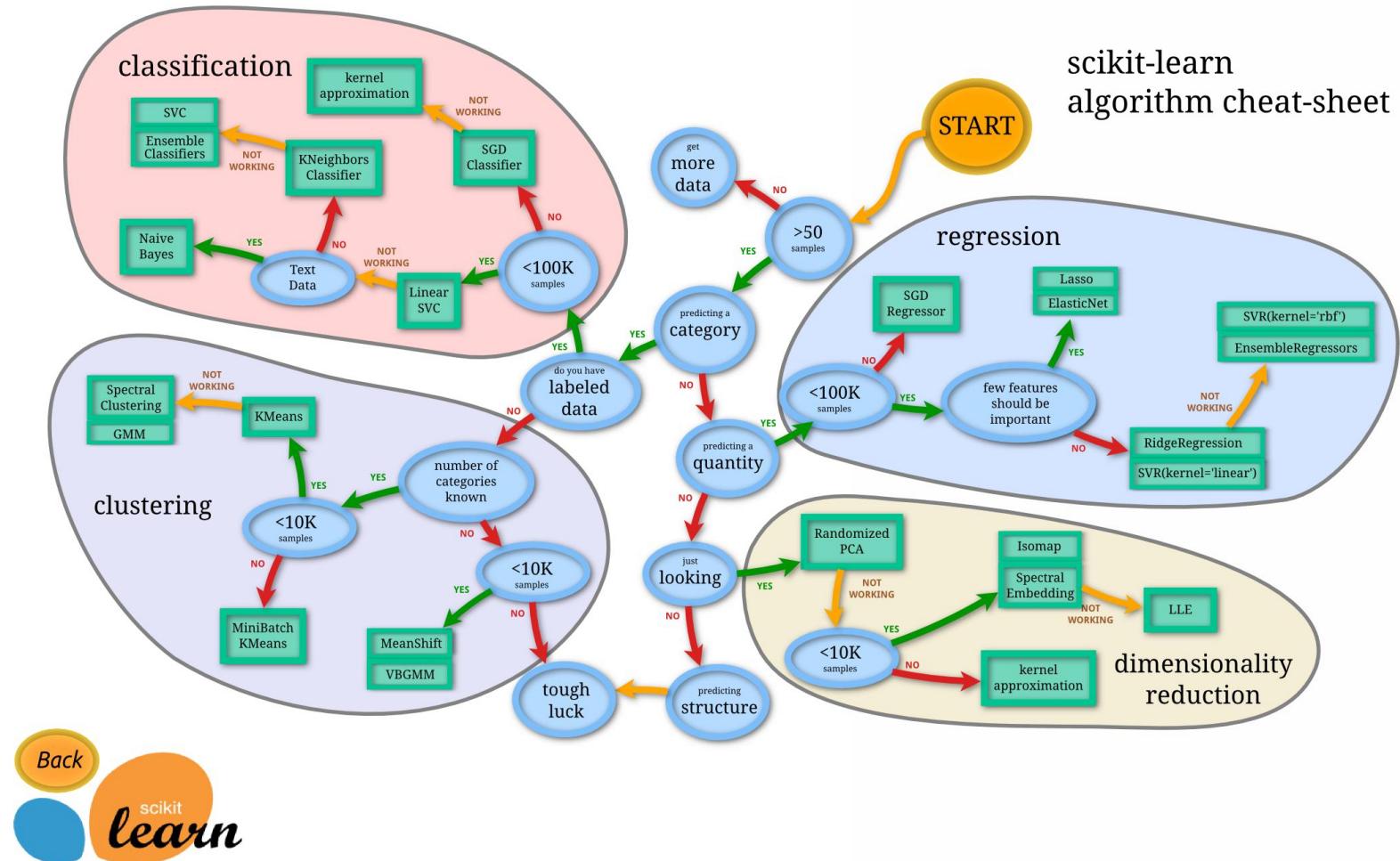


Reinforcement Learning

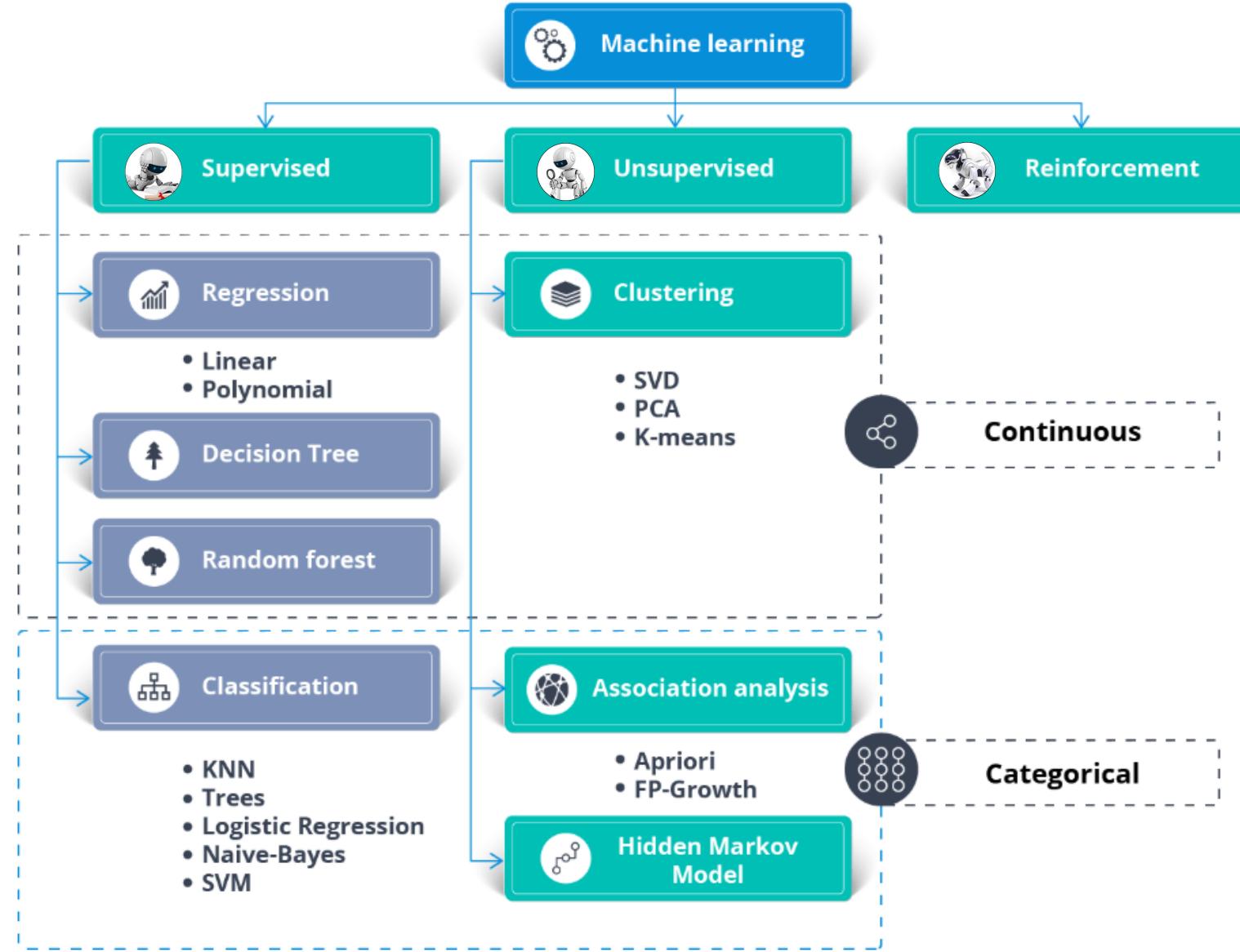
Quando usar cada treinamento?



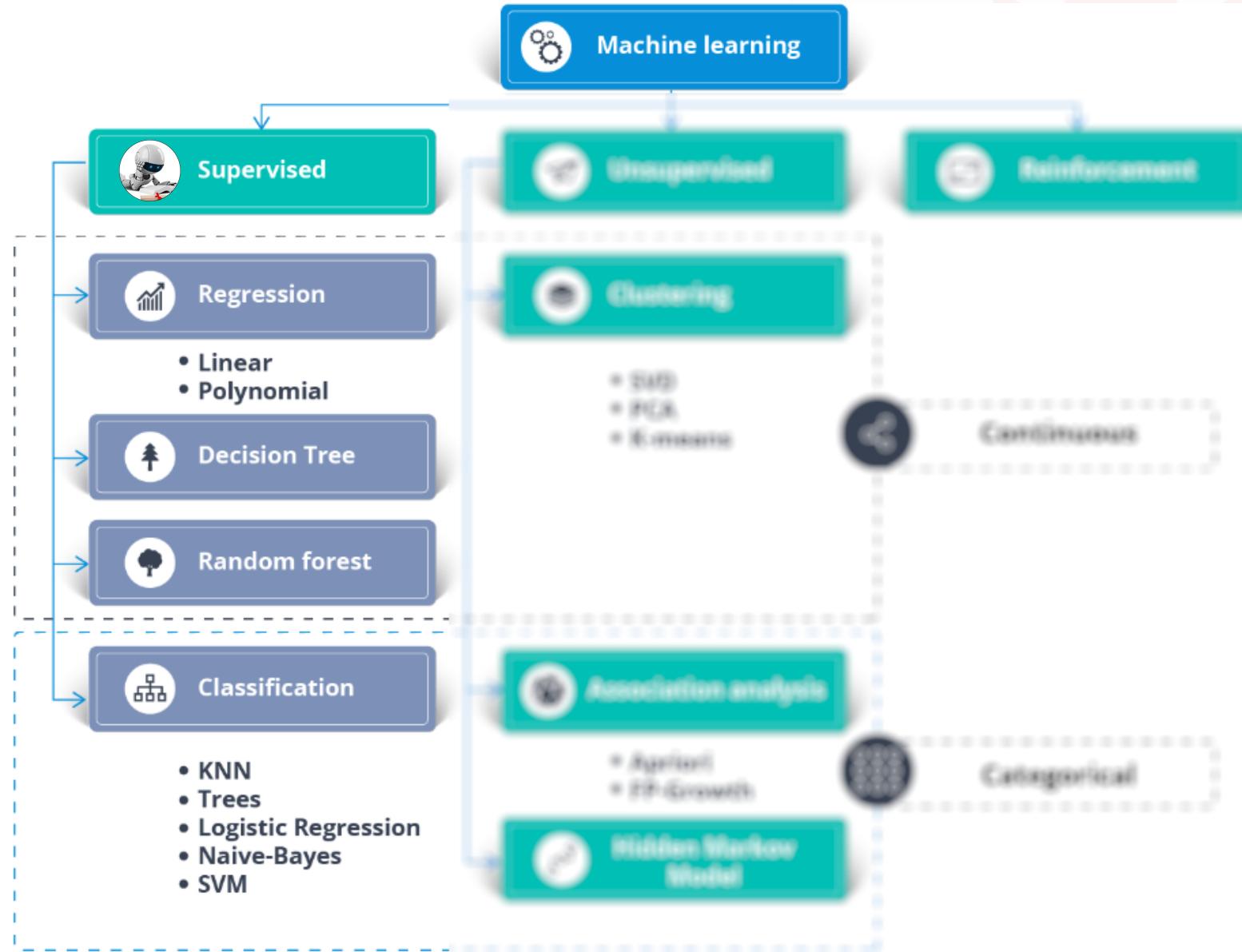
O Scikit-learn



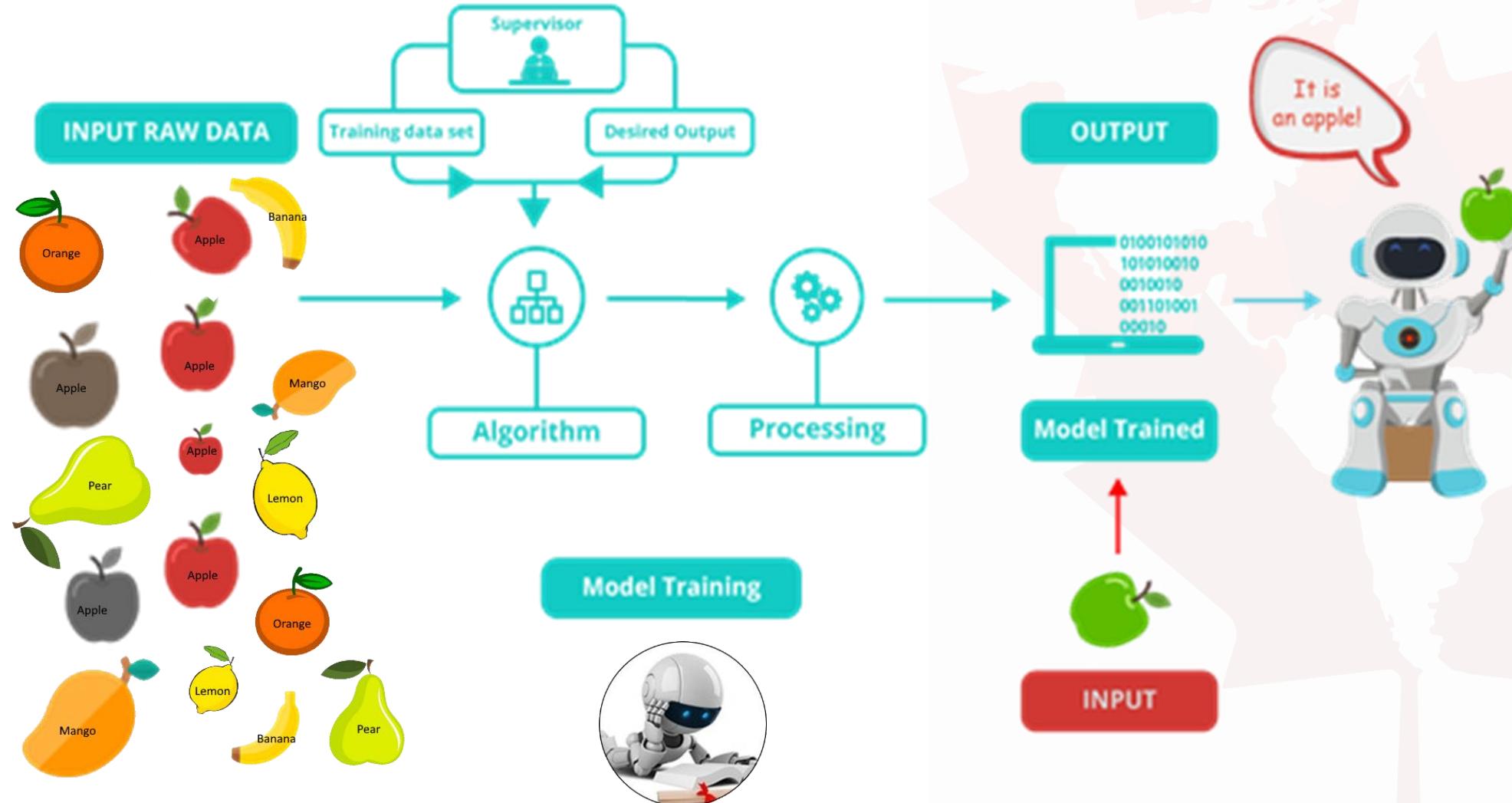
Alguns algoritmos mais comuns



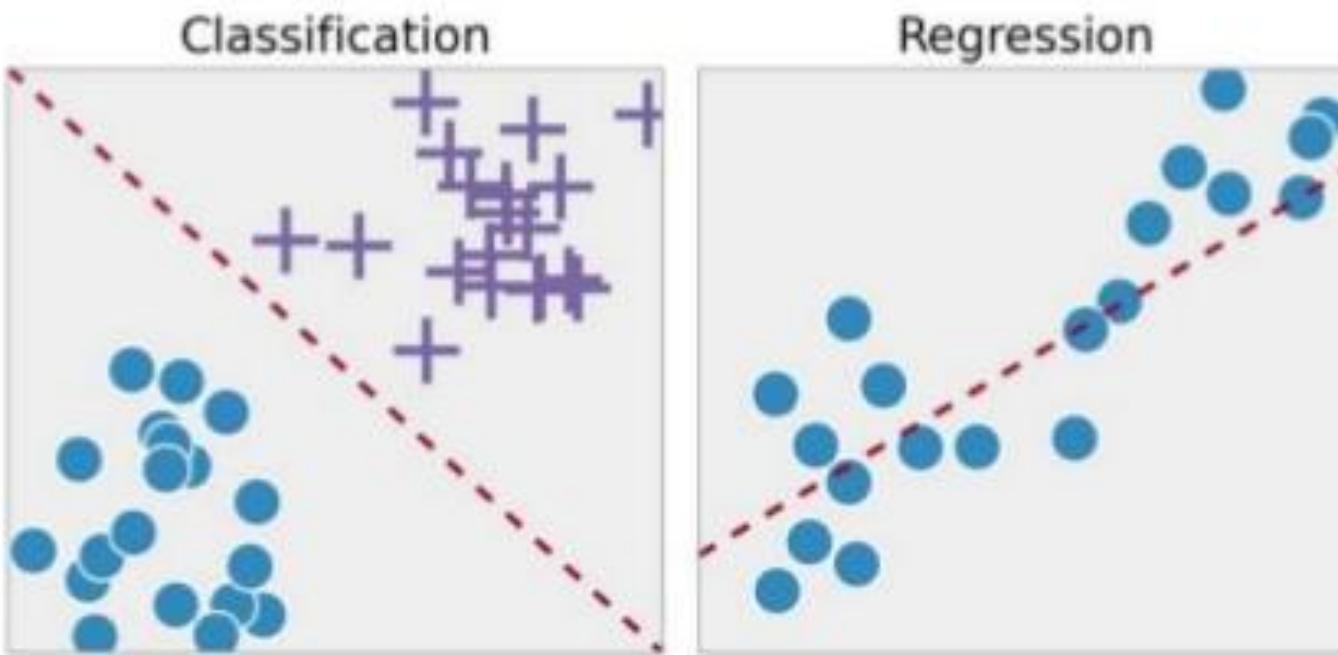
Aprendizado Supervisionado



Aprendizado Supervisionado

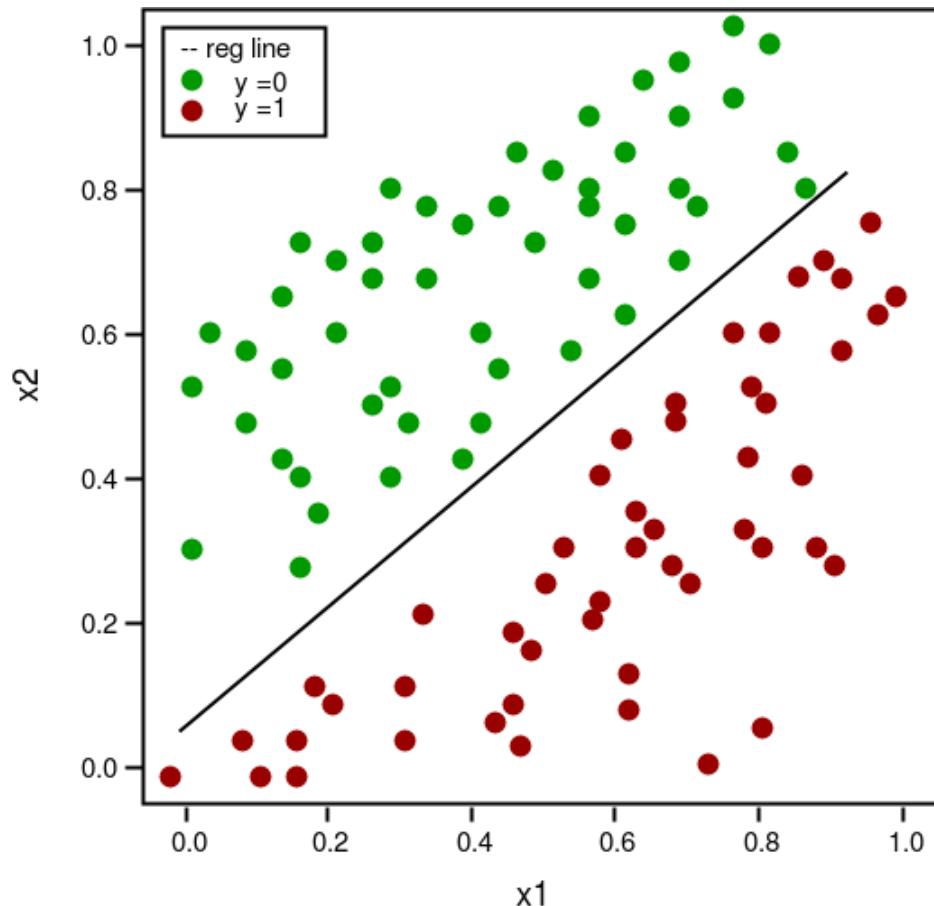


Classificação e Regressão



Classificação

Nesta classe de algoritmos, estão aqueles que usamos para **posicionar** um **valor** em uma **classe** previamente conhecida dentre várias outras.



Os tipos mais comuns são:

1. KNN
2. Arvores de Decisão (Trees)
3. Regressão Logística
4. Naive-Bayes
5. SVM
6. Multilayer Perceptrons

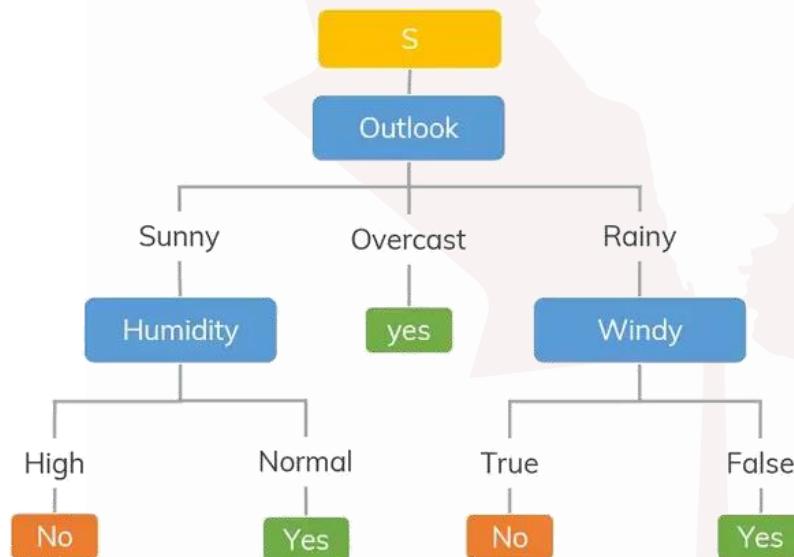
Árvores de Decisão

Nesta classe de algoritmos, estão aqueles que usamos para **decidir** através de inferência indutiva a **classe** a qual pertence um **conjunto de valores**.

Outlook	Temperature	Humidity	Windy	Play Golf
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Alguns modelos deste tipo:

1. **Decision Trees**
2. **Random Forest**



O que é Underfit e Overfit?

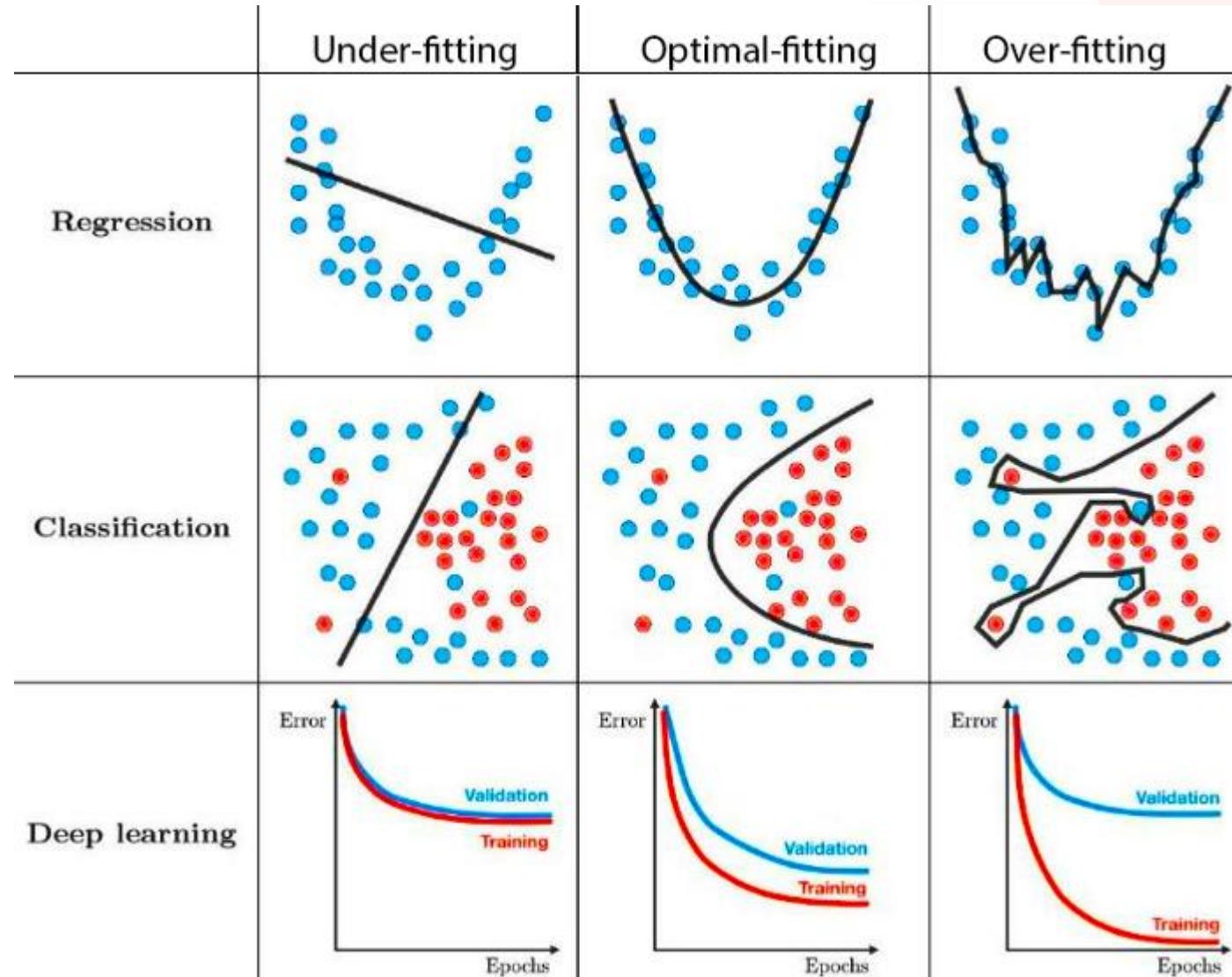
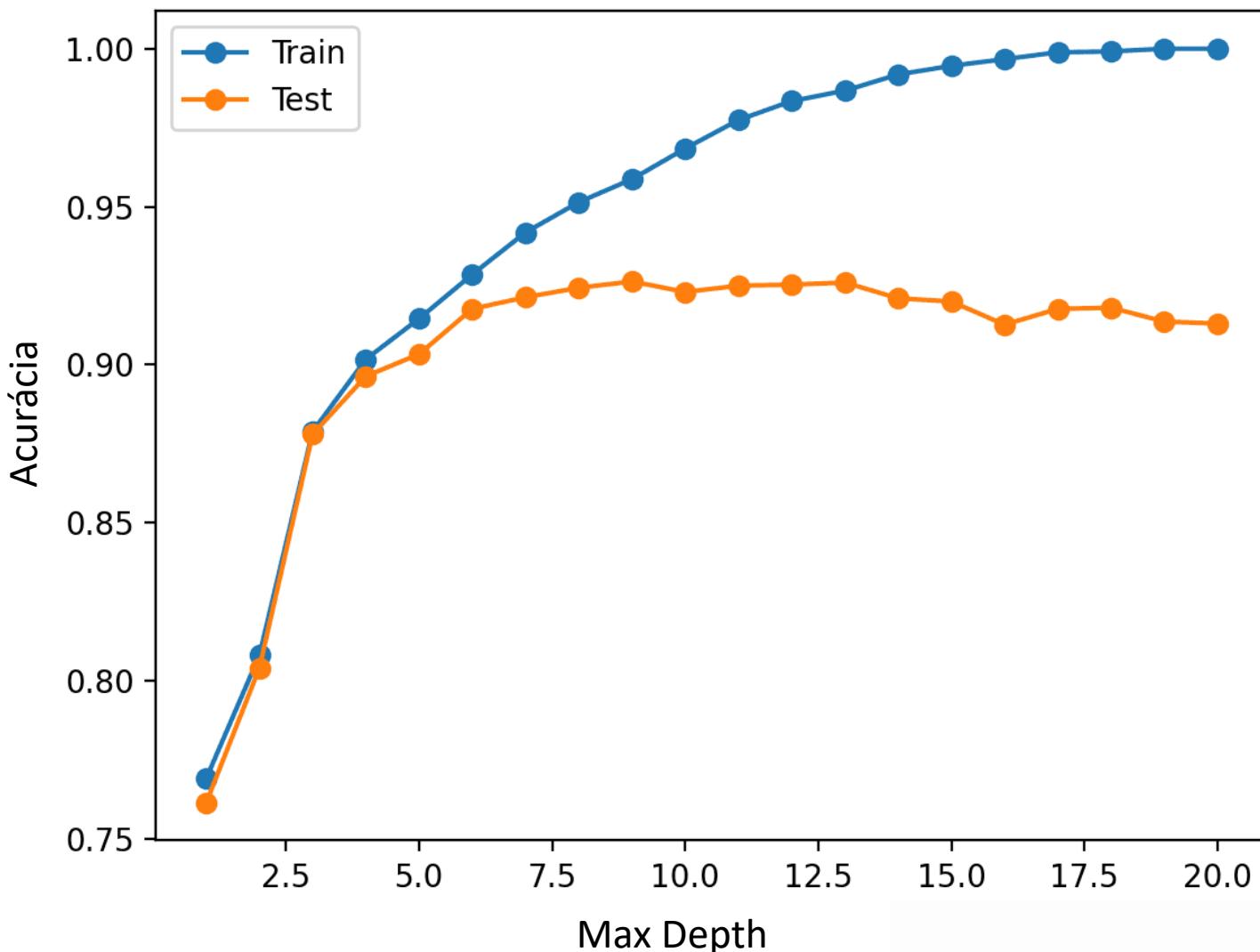


Gráfico de Aprendizado

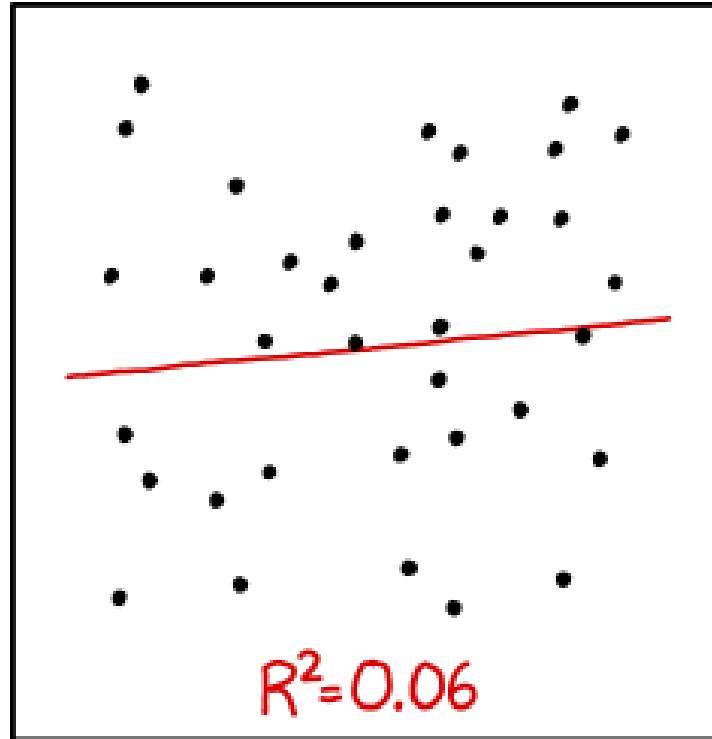


Remédios

- Underfitting
 - Obtenha mais dados
 - Diminua a regularização
 - Aumente a complexidade do modelo
 - Aumente o tempo de treinamento
- Overfitting
 - Obter mais dados ou Data augmentation
 - Aumentar a Regularização
 - Early stopping
 - Pruning
 - Dropout
 - Remover features
 - Cross validation



Vamos falar de regressão



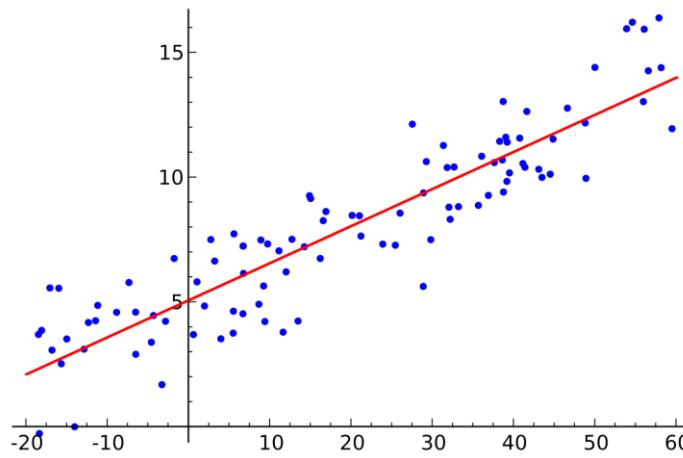
I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER
TO GUESS THE DIRECTION OF THE CORRELATION FROM THE
SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

Regressão

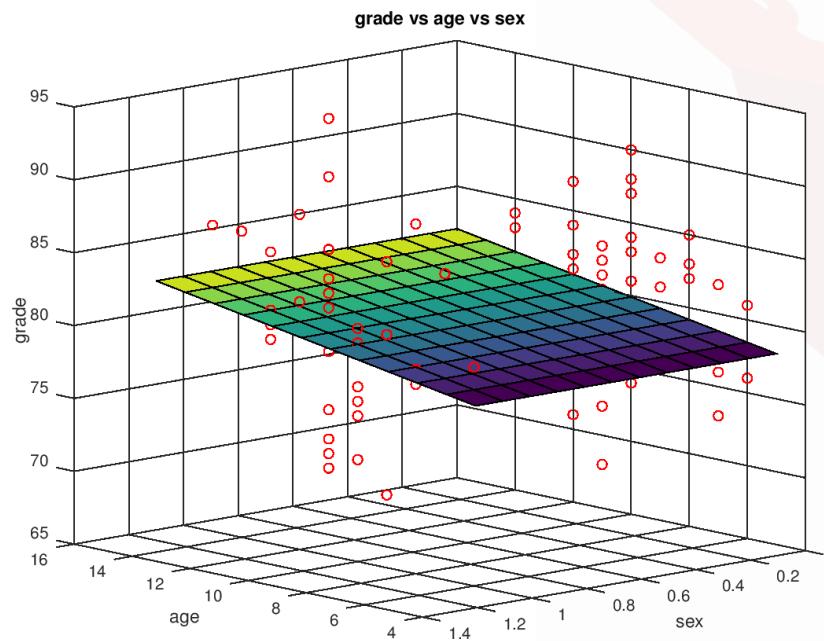
Utilizada quando queremos **estimar** algum **valor**.

Estabelecer a **relação de dependência** entre a variável **dependente** e a variável **independente**.

O modelo mais simples de regressão é a regressão linear, sendo representada pela equação $Y=aX+B$, onde Y é a variável dependente e X é a variável independente.



Regressão Linear Simples



Regressão Linear Múltipla

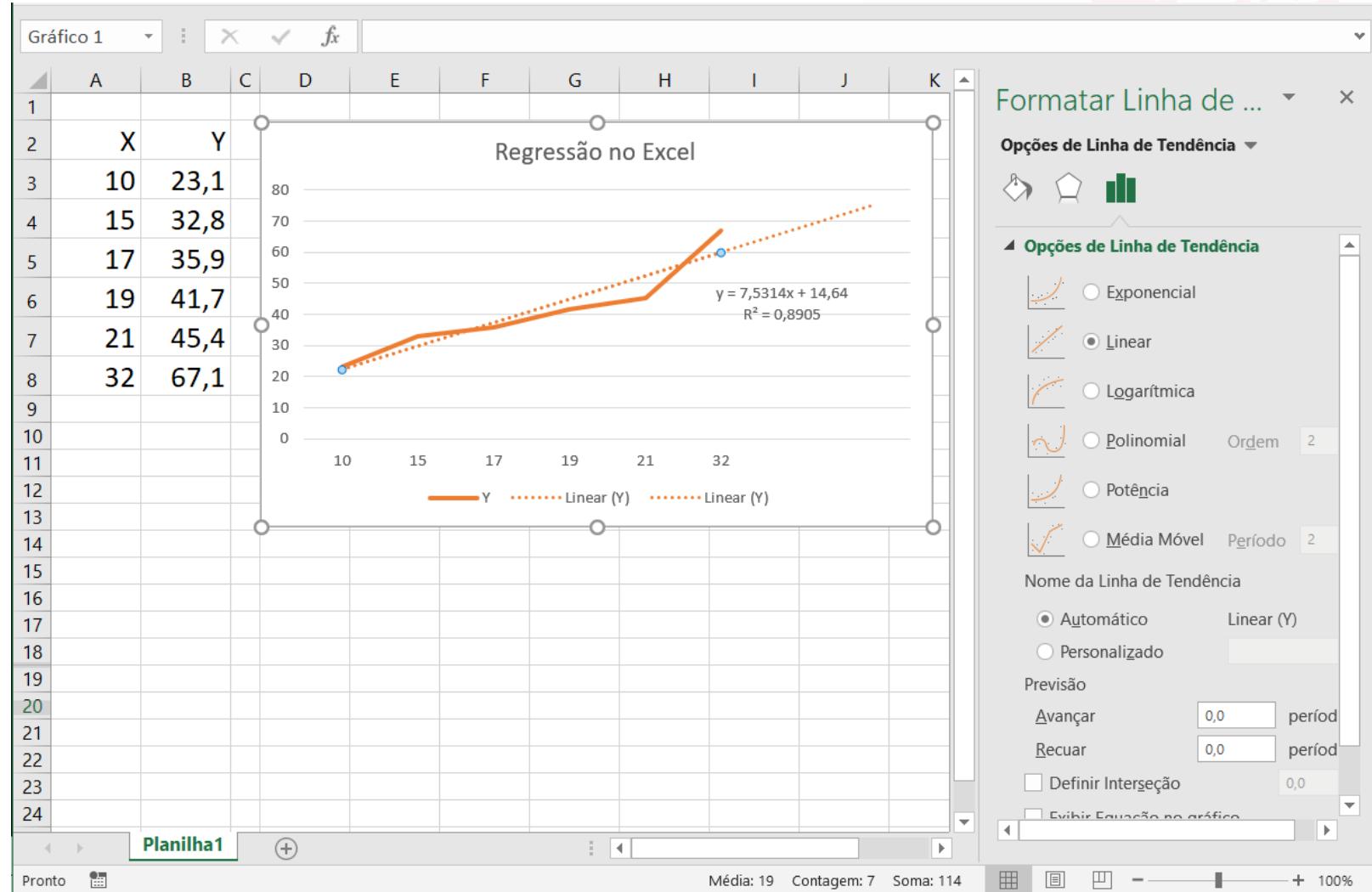
Os tipos mais comuns são:

1. Regressão linear
2. Regressão polinomial

Outros algoritmos mais sofisticados:

1. SGD for regression
2. Lasso
3. Elasticnet
4. Ridge regression
5. SVR
6. MLP for regression
7. LSTM

Regressão no Excel



Regressão no Weka

The image shows the Weka software interface with three main windows:

- Weka GUI Chooser:** A window displaying the WEKA logo (a kiwi bird) and the text "WEKA THE UNIVERSITY OF WAIKATO 30 Years Research in Machine Learning NEW ZEALAND". It also shows the version information: "Waikato Environment for Knowledge Analysis Version 3.8.6 (c) 1999 - 2022 The University of Waikato Hamilton, New Zealand".
- Weka Explorer (Preprocess tab):** This window is active. It shows the "Current relation" section with "Relation: boston", "Attributes: 14", "Instances: 506", and "Sum of weights: 506". Below this, the "Attributes" section lists attributes CRIM, ZN, INDUS, CHAS, NOX, and RM. A histogram is shown for the MEDV attribute.
- Weka Explorer (Classify tab):** This window shows a "Classifier" configuration with "Choose" set to "LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4". Under "Test options", "Cross-validation" is selected with "Folds: 10". The "Classifier output" pane displays the generated Linear Regression Model:

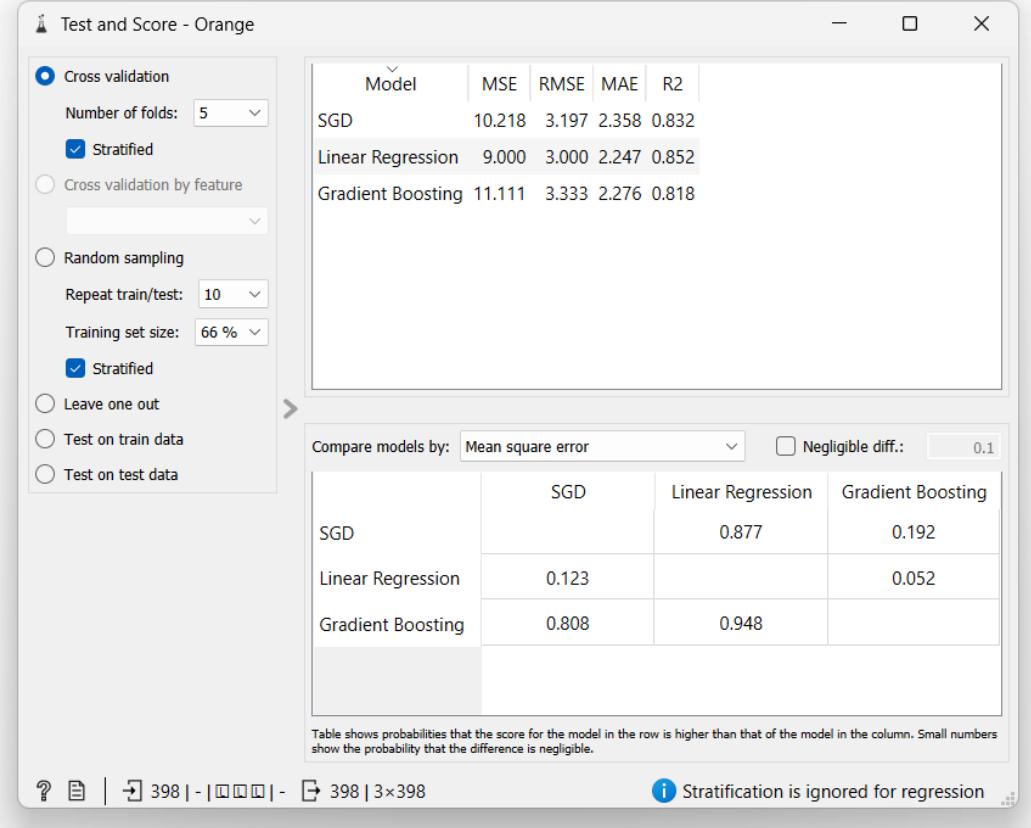
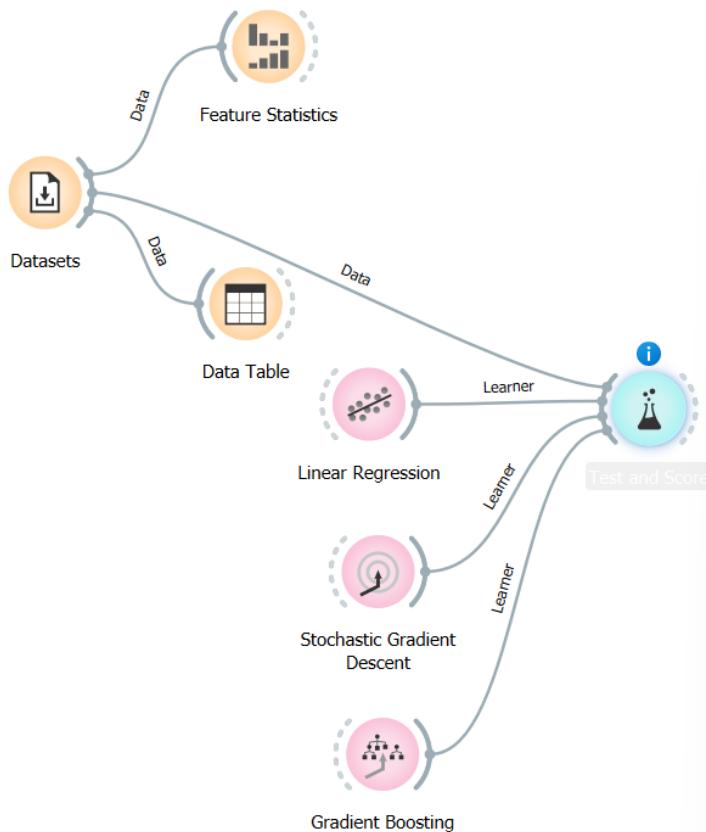
```
MEDV =  
-0.1078 * CRIM +  
0.0523 * ZN +  
2.6214 * CHAS=1 +  
-16.4618 * NOX +  
3.6229 * RM +  
-1.5394 * DIS +  
-5.948 * RAD=6,4,1,5,2,7,3,8 +  
1.5269 * RAD=4,1,5,2,7,3,8 +  
-2.5869 * RAD=1,5,2,7,3,8 +  
2.6026 * RAD=5,2,7,3,8 +  
2.1579 * RAD=7,3,8 +  
-0.0073 * TAX +  
-0.9866 * PTRATIO +  
0.0093 * B +  
-0.5333 * LSTAT +  
42.0712
```

Time taken to build model: 0.12 seconds

==== Cross-validation ====
==== Summary ====

Metric	Value
Correlation coefficient	0.8469
Mean absolute error	3.365
Root mean squared error	4.8892
Relative absolute error	50.4694 %
Root relative squared error	53.0339 %
Total Number of Instances	506

Regressão no Orange



Jupyter Notebooks

285 lines (285 sloc) | 3.79 KB

Simple linear regression

Import the relevant libraries

```
In [ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
import seaborn as sns
sns.set()
```

Load the data

```
In [ ]: data = pd.read_csv('1.01. Simple linear regression.csv')
```

```
In [ ]: data
```

```
In [ ]: data.describe()
```

Create your first regression

Define the dependent and the independent variables

```
In [ ]: y = data['GPA']
x1 = data['SAT']
```

Explore the data

```
In [ ]: plt.scatter(x1,y)
plt.xlabel('SAT', fontsize = 20)
plt.ylabel('GPA', fontsize = 20)
plt.show()
```

Regression itself

```
In [ ]: x = sm.add_constant(x1)
results = sm.OLS(y,x).fit()
results.summary()
```

```
In [ ]: plt.scatter(x1,y)
yhat = 0.0017*x1 + 0.275
fig = plt.plot(x1,yhat, lw=4, c='orange', label='regression line')
plt.xlabel('SAT', fontsize = 20)
plt.ylabel('GPA', fontsize = 20)
plt.show()
```

Import the relevant libraries

```
In [ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
import seaborn as sns
sns.set()
```



Regression itself

```
In [ ]: x = sm.add_constant(x1)
results = sm.OLS(y,x).fit()
results.summary()
```

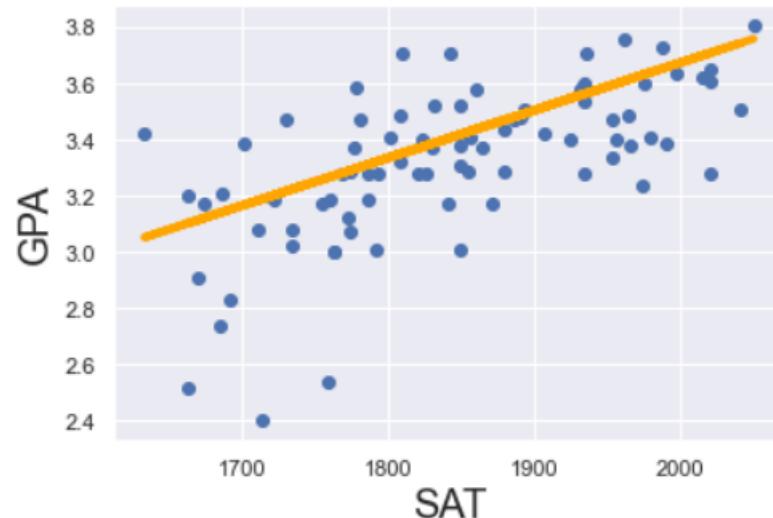
Jupyter Notebooks

Out[7]: OLS Regression Results

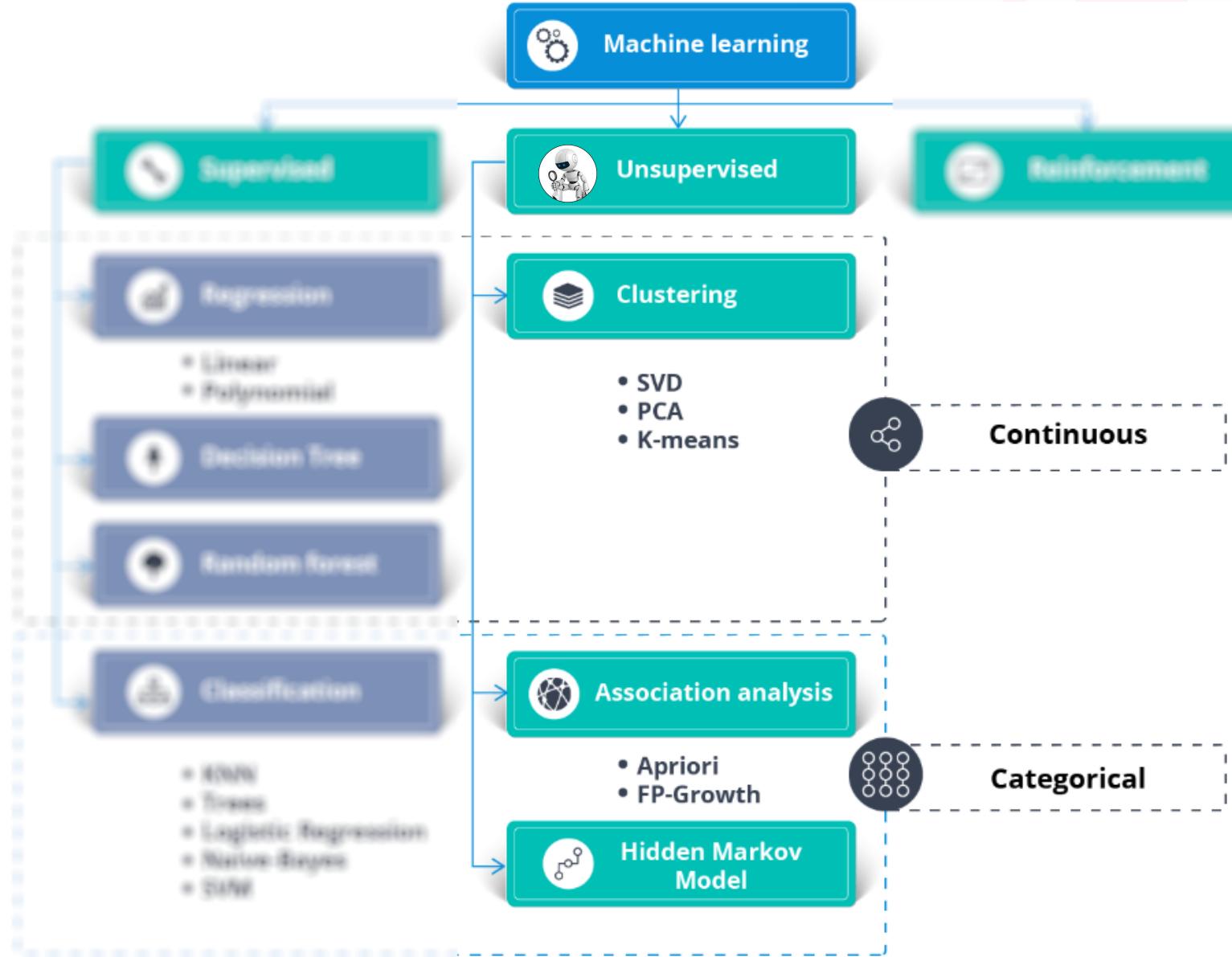
Dep. Variable:	GPA	R-squared:	0.406			
Model:	OLS	Adj. R-squared:	0.399			
Method:	Least Squares	F-statistic:	56.05			
Date:	Sat, 23 Oct 2021	Prob (F-statistic):	7.20e-11			
Time:	07:45:28	Log-Likelihood:	12.672			
No. Observations:	84	AIC:	-21.34			
Df Residuals:	82	BIC:	-16.48			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	0.2750	0.409	0.673	0.503	-0.538	1.088
SAT	0.0017	0.000	7.487	0.000	0.001	0.002
Omnibus:	12.839	Durbin-Watson:	0.950			
Prob(Omnibus):	0.002	Jarque-Bera (JB):	16.155			
Skew:	-0.722	Prob(JB):	0.000310			
Kurtosis:	4.590	Cond. No.	3.29e+04			

In [8]:

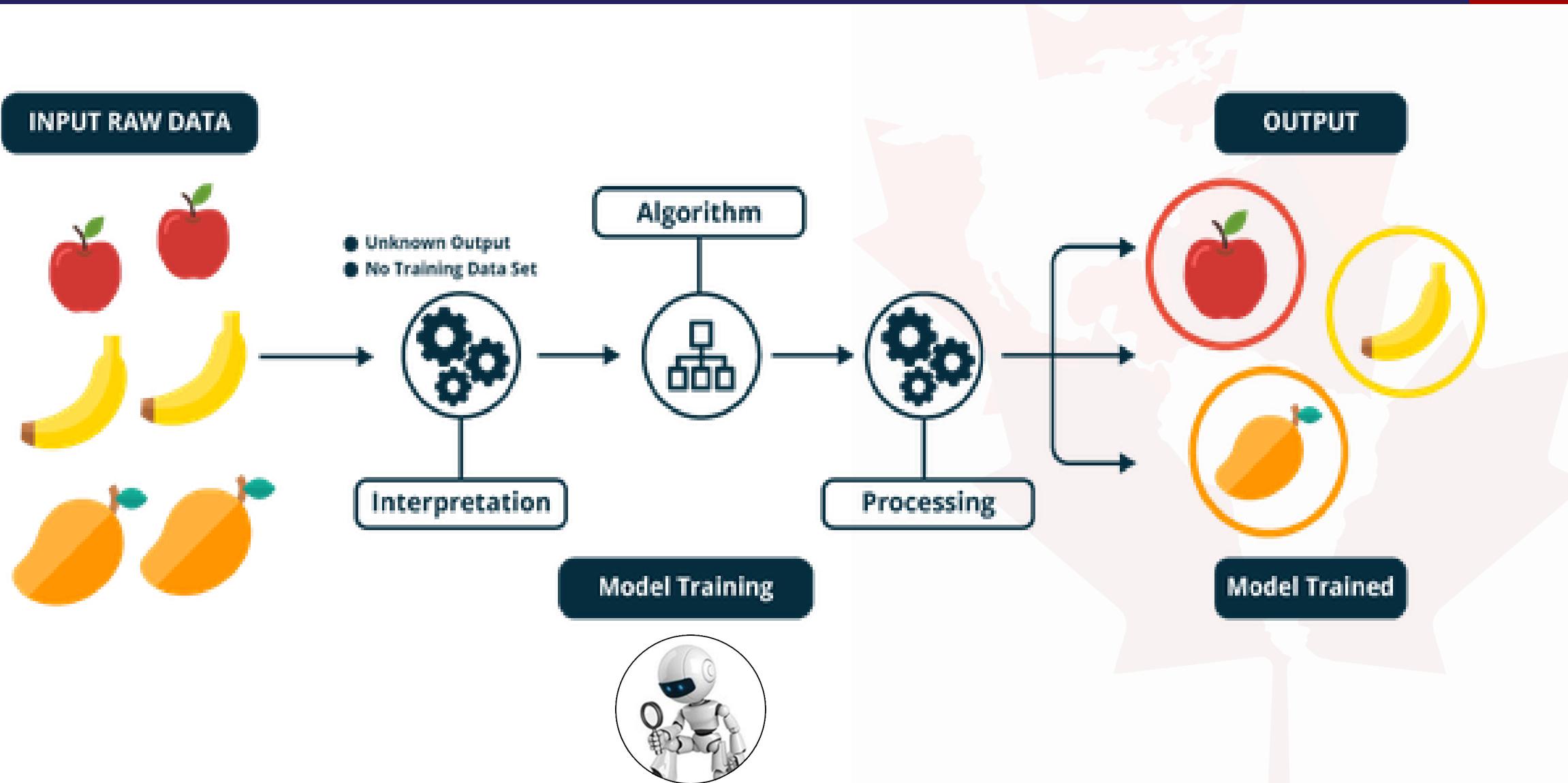
```
# Create a scatter plot
plt.scatter(x1,y)
# Define the regression equation, so we can plot it later
yhat = 0.0017*x1 + 0.275
# Plot the regression line against the independent variable (SAT)
fig = plt.plot(x1,yhat, lw=4, c='orange', label ='regression line')
# Label the axes
plt.xlabel('SAT', fontsize = 20)
plt.ylabel('GPA', fontsize = 20)
plt.show()
```



Aprendizado Não Supervisionado

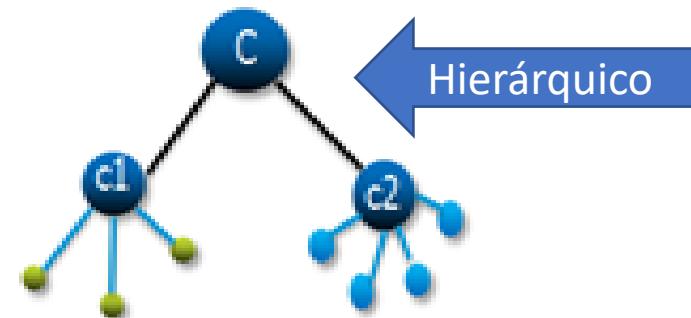
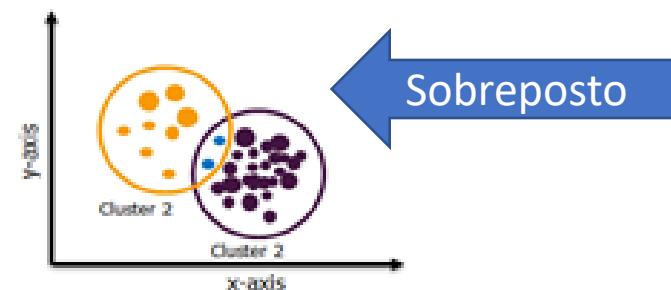
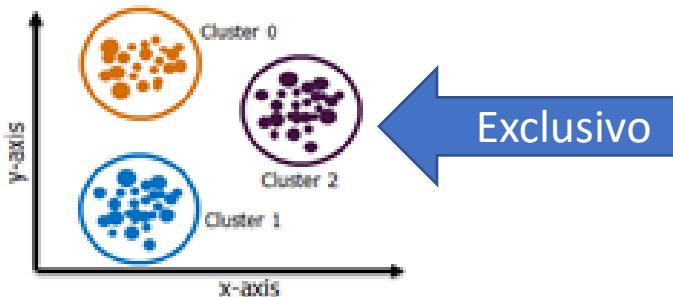


Aprendizado Não Supervisionado



Clusterização e associação

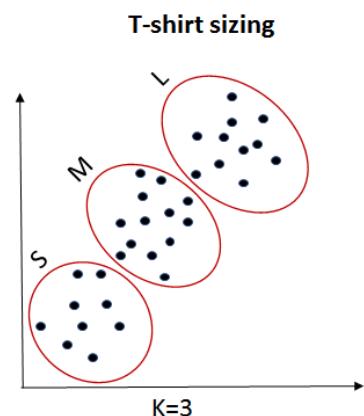
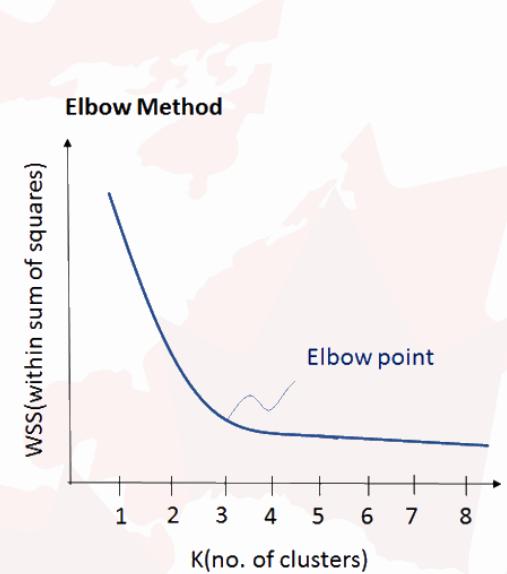
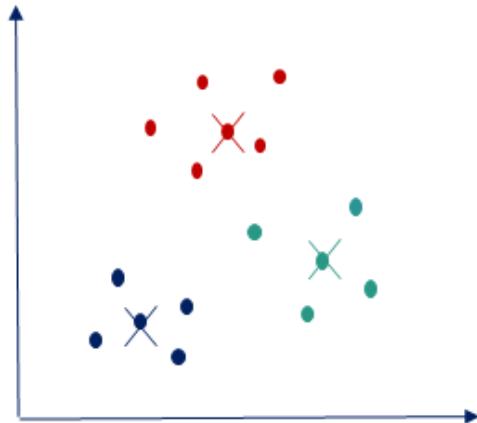
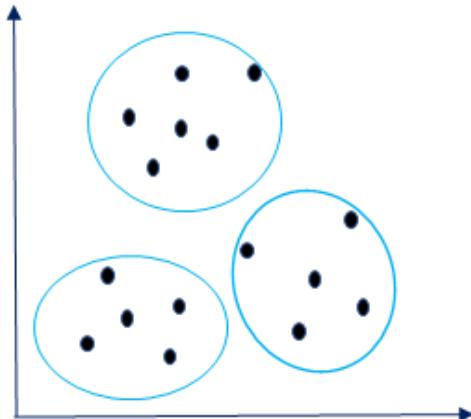
Utilizamos algoritmos de clusterização e associação quando queremos **descobrir** se existe algum **padrão** nos dados que possa dividi-los em **classes** que ainda não mapeamos.



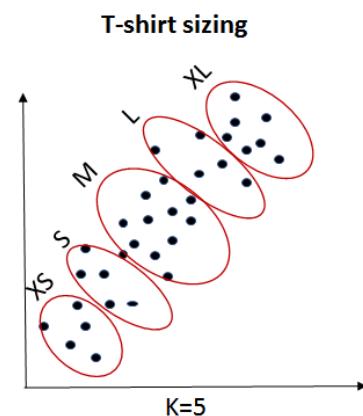
Os tipos mais comuns são:

1. K-means
2. Apriori algorithm for association rule learning problems
3. Principal Component Analysis
4. Singular Value Decomposition
5. Independent Component Analysis

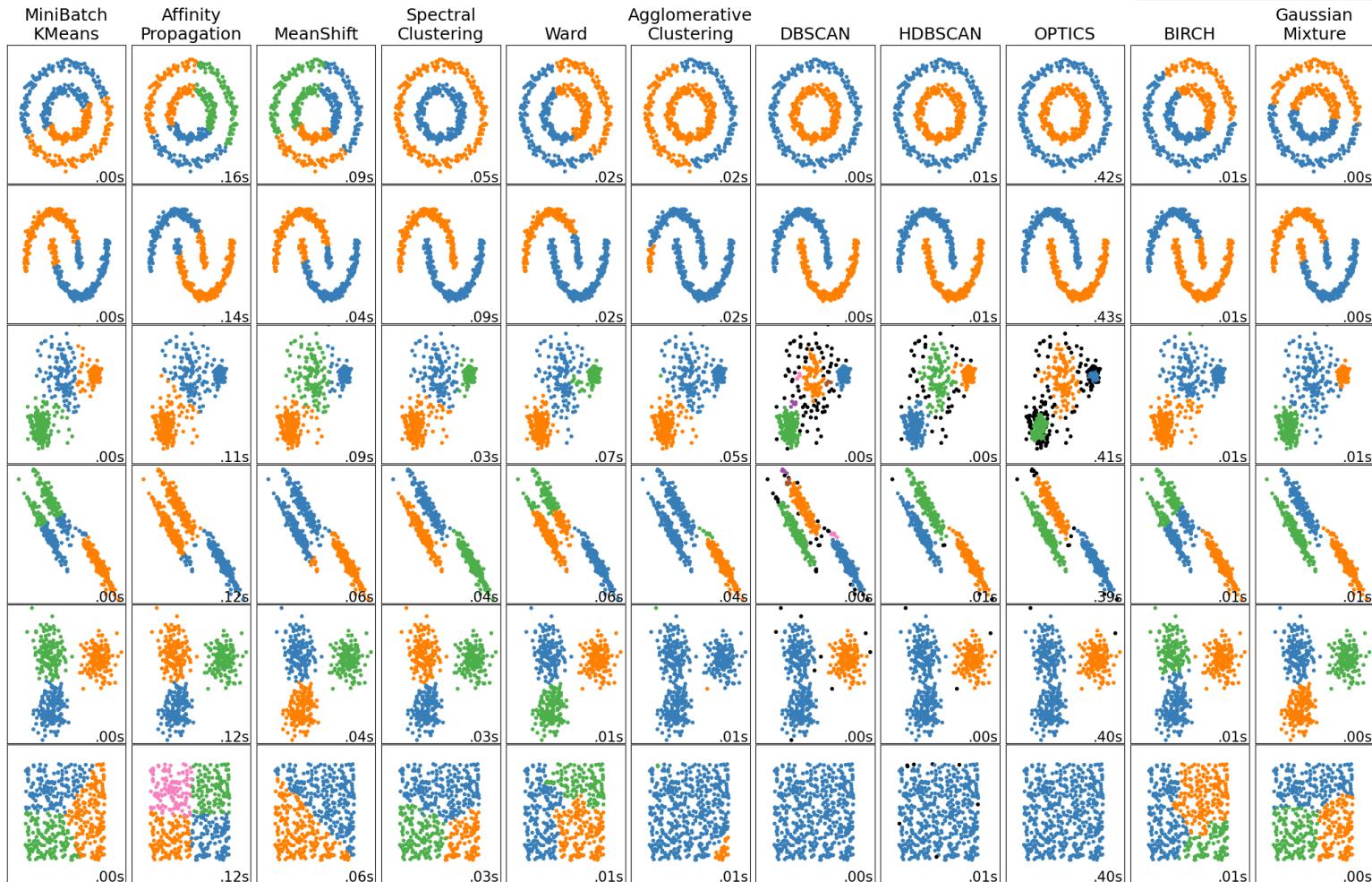
Entendendo o K-Means



Purpose
Based

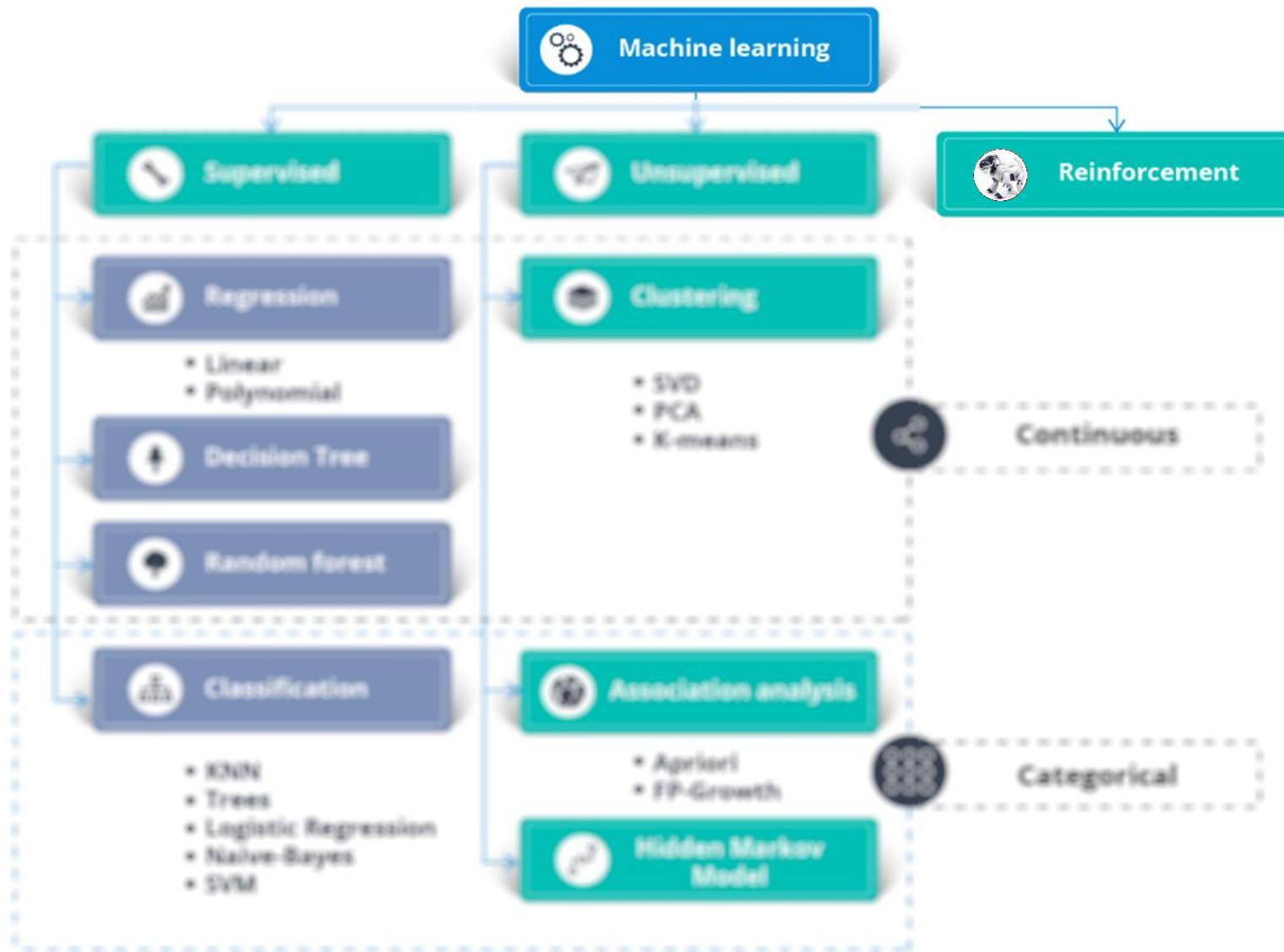


Algoritmos de Clusterização

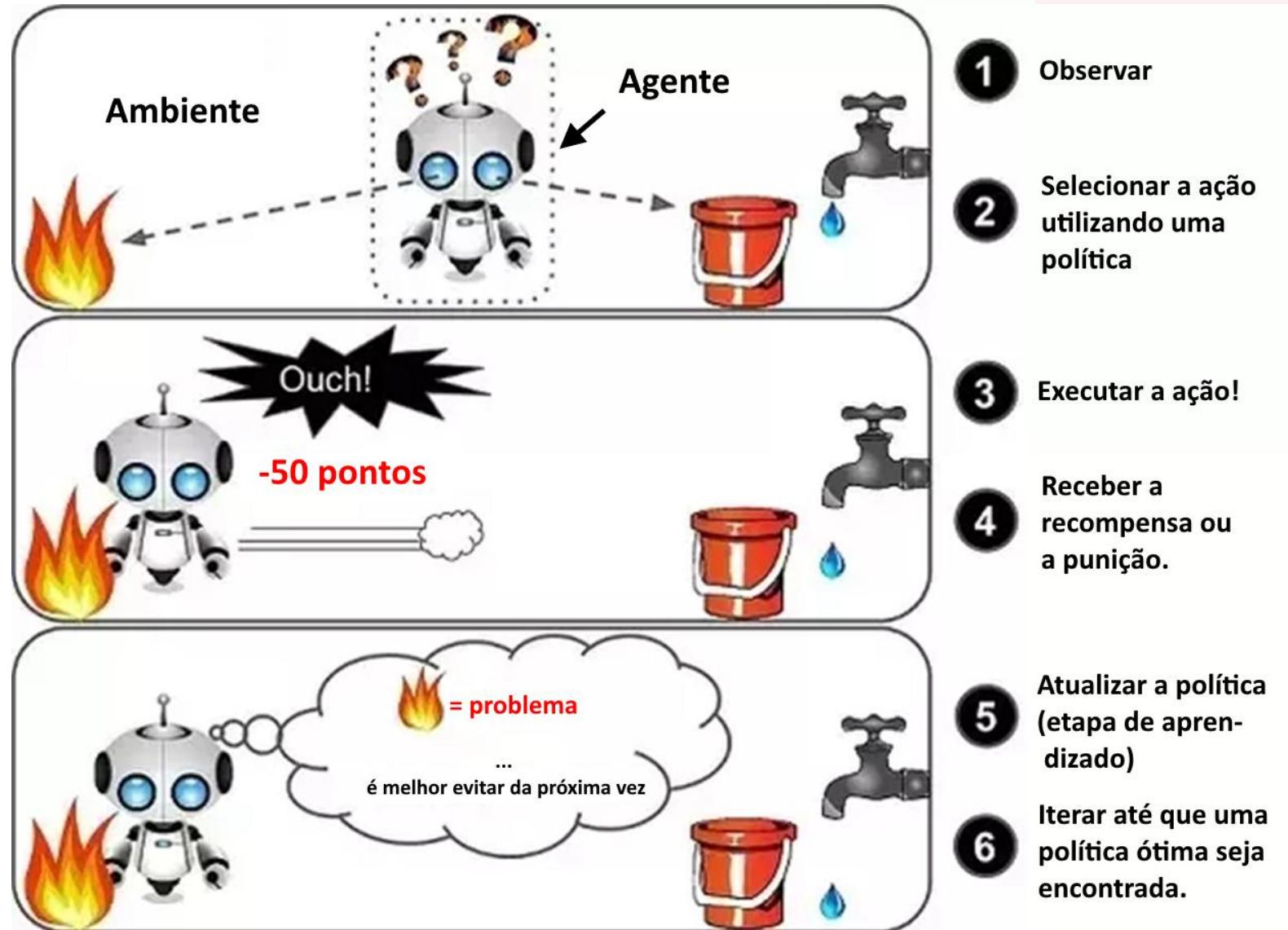


https://scikit-learn.org/stable/_downloads/803ca1bcd8dd2c364836a6784144355b/plot_cluster_comparison.ipynb

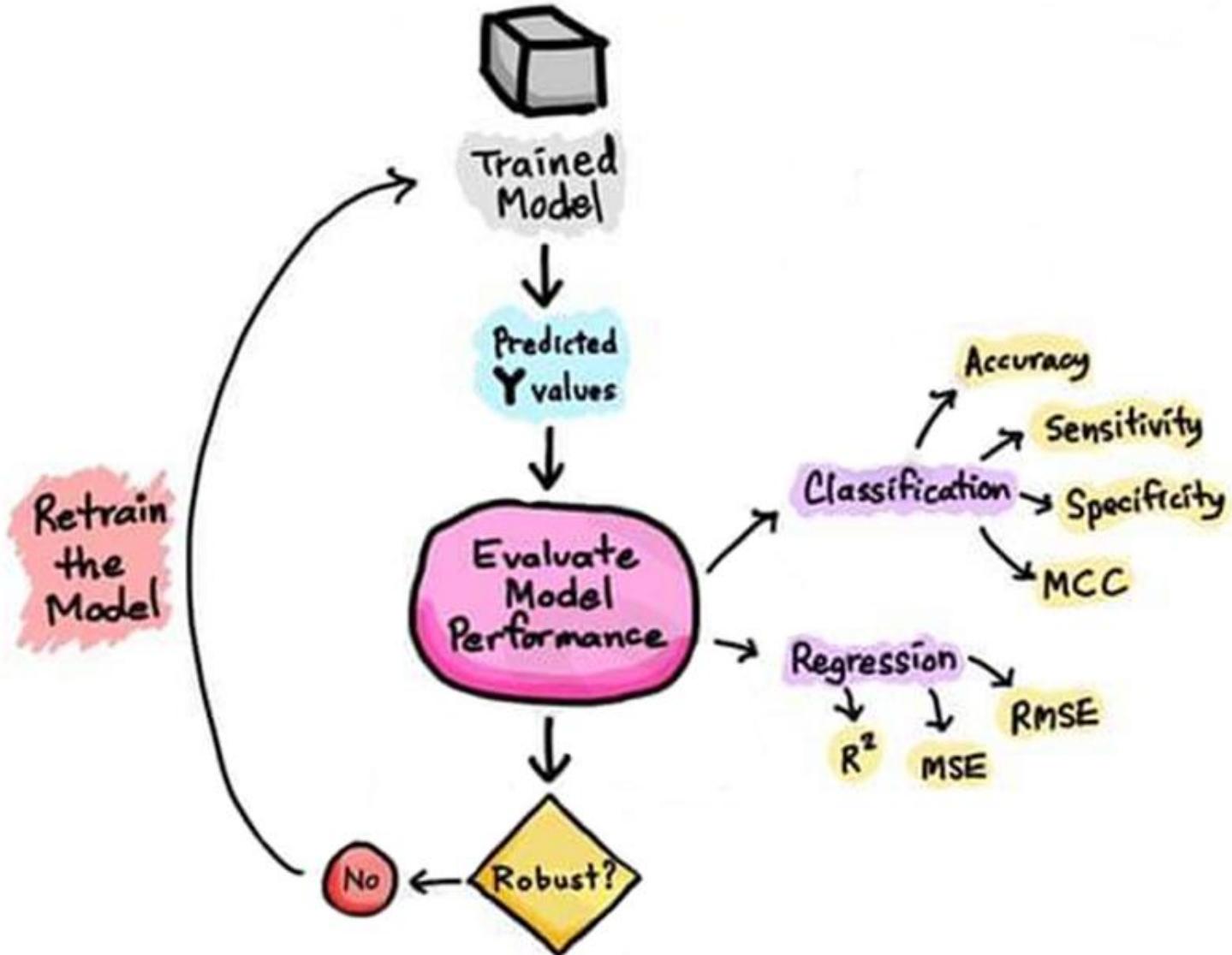
Aprendizado por Reforço



Aprendizado por Reforço



Fase 5 - Avaliação do Modelo



Classificação - Métricas

Matriz de Confusão

		Preditos	
		Positivo	Negativo
Observado	Positivo	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Negativo	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Classificação - Métricas

		Predito	
		Positivo	Negativo
Observado	Positivo	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Negativo	Falso Positivo (FP)	Verdadeiro Negativo (TN)

$$Error\ Rate = \frac{(FN + FP)}{(TP + FP + FN + TN)}$$

$$Precision = \frac{TP}{(TP + FP)}$$

$$Negative\ Predicted\ Value = \frac{TN}{(TN + FN)}$$

$$F\beta = \frac{(1 + \beta^2) \times Precision \times Recall}{(\beta^2 \times Precision) + Recall}$$

$$Sensitivity/Recall = \frac{TP}{(TP + FN)}$$

$$Specificity = \frac{TN}{(TN + FP)}$$

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)}$$

Classificação - Métricas

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

Equation for calculating the Matthews Correlation Coefficient.

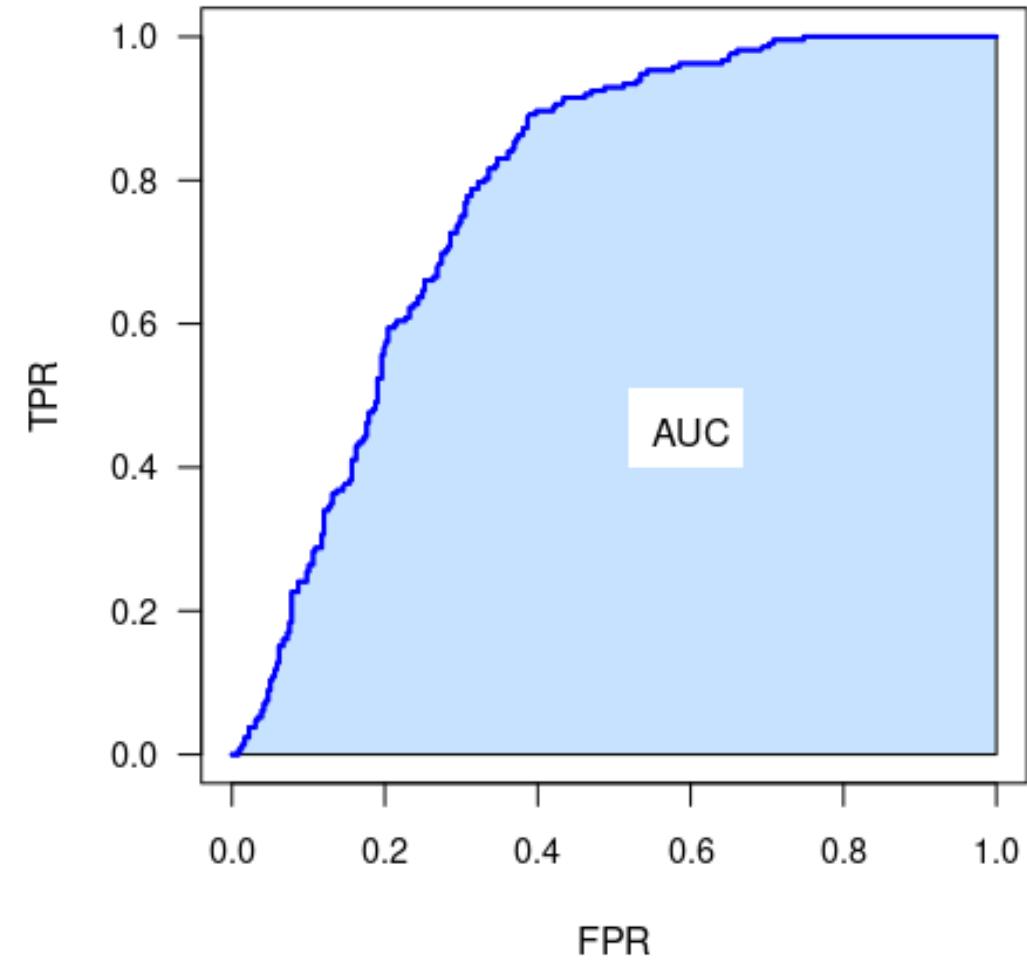
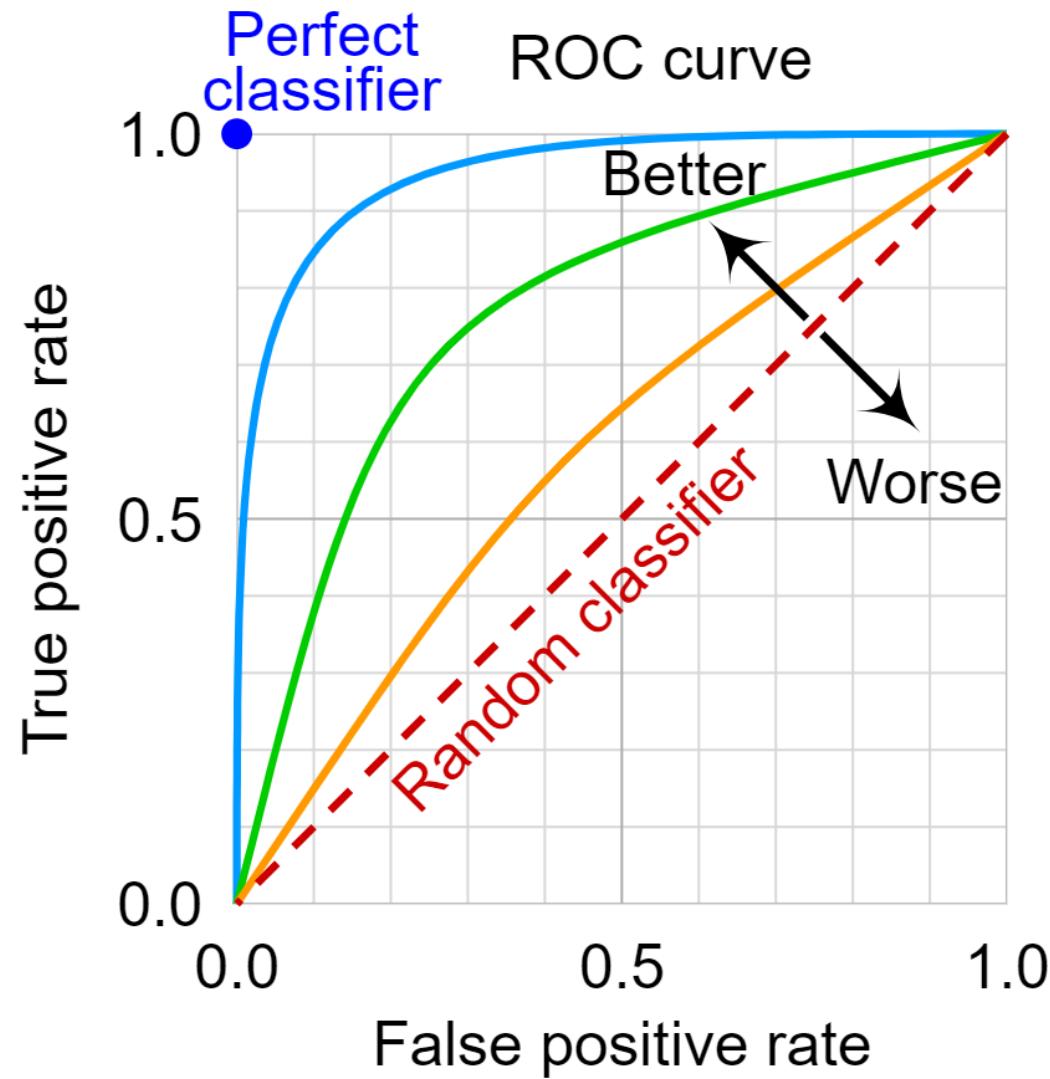
Classificação - Métricas

- Acurácia: o quanto o meu modelo está classificando corretamente.
- Precisão: O quanto estou classificando corretamente os positivos. Dos exemplos classificados como positivos, quantos realmente são positivos?
- Sensibilidade: dentre todas as situações de classe positivo como valor esperado, quantas estão corretas. De todos os exemplos que são positivos, quantos foram classificados corretamente como positivos?
- F1 Score: média harmônica entre precisão e sensibilidade

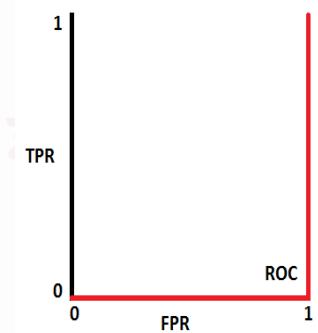
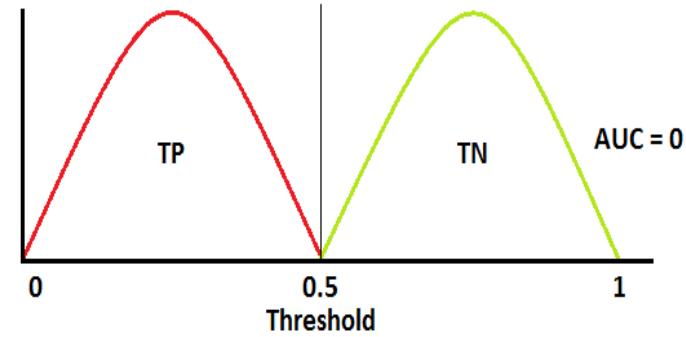
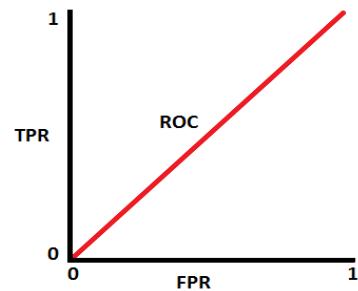
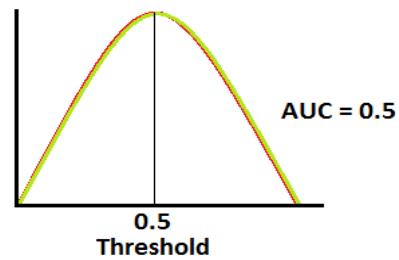
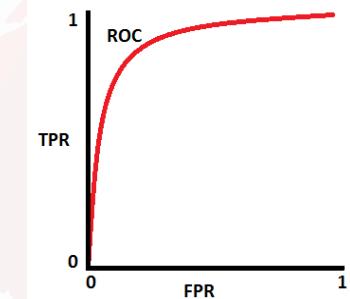
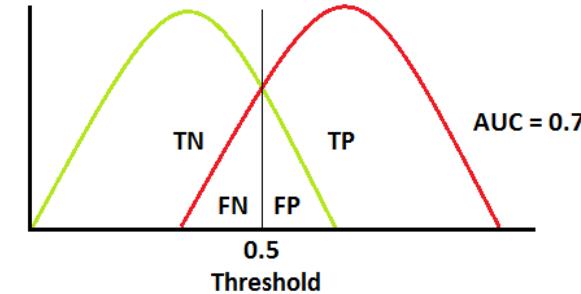
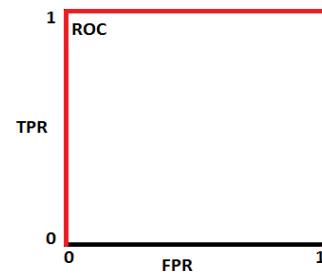
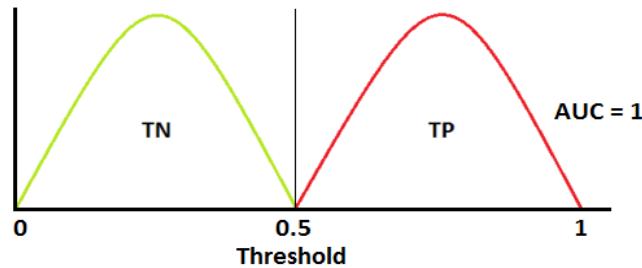
Classificação - Métricas

- Acurácia: mede o desempenho geral do modelo. Cuidado... Pode ser enganosa.
- Precisão: Falsos positivos são mais prejudiciais que falsos negativos (ex. investimentos).
- Sensibilidade: Falsos negativos são mais prejudiciais que falsos positivos (ex. doenças).
- F1 Score: Avaliar Precisão e Sensibilidade juntos. É alto se ambos forem altos.

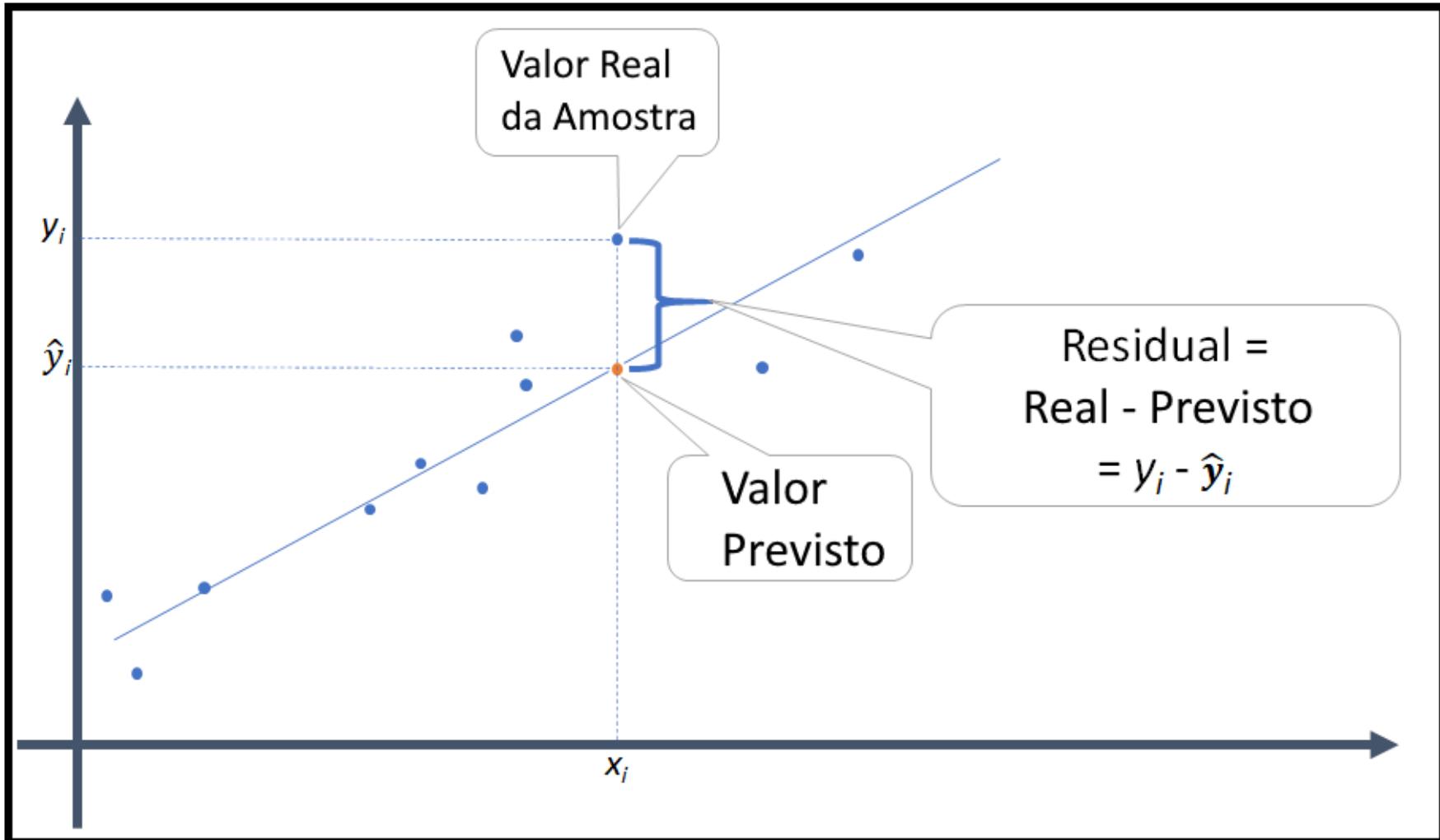
Classificação - Curva ROC e AUC



Classificação - Curva ROC e AUC

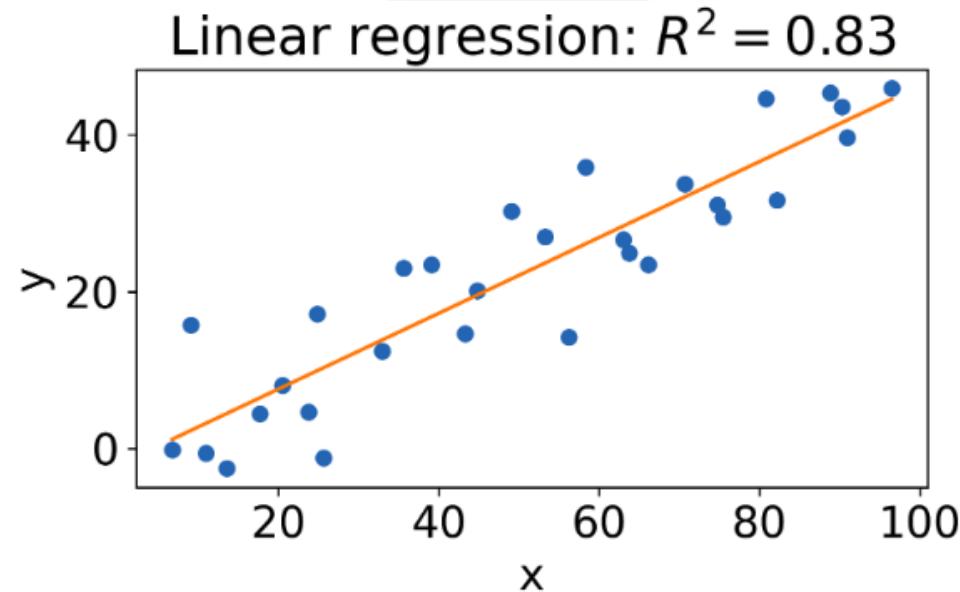
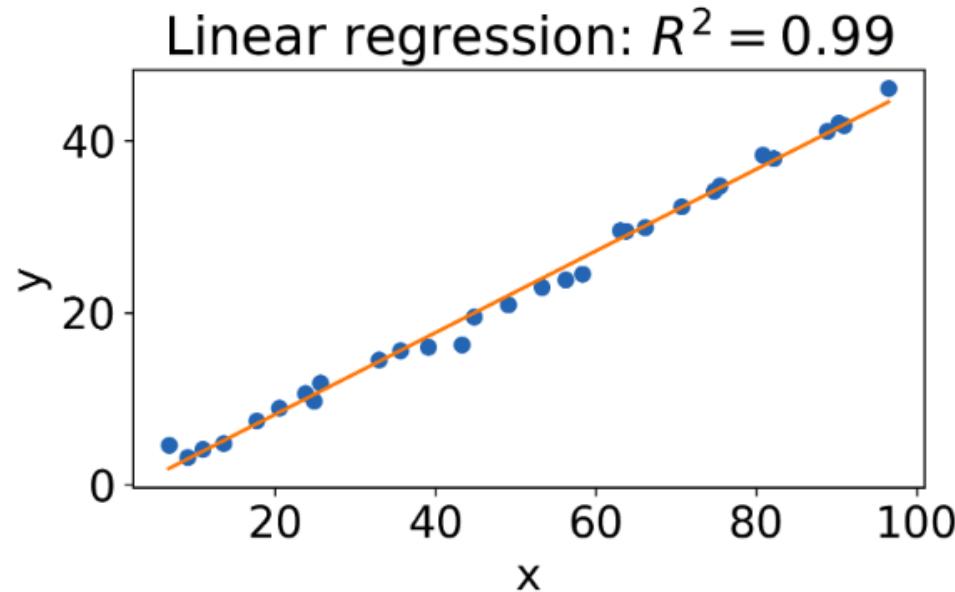


Regressão - Erro Residual



Regressão - Métricas de avaliação

$$R^2 = 1 - \frac{\text{Variancia Residual}}{\text{Varianca Total}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$



Regressão - Métricas de avaliação

Várias desvantagens e Limitações do R²

- Por definição é enviesado;
- Só pode ser aplicado perfeitamente em modelos univariados;
- Em casos de Overfitting, o valor dessa métrica continua alta;
- Apenas o R-Quadrado não consegue indicar se um modelo de regressão é eficiente ou não;

Regressão - Métricas de avaliação

$$R_a^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Entrada de mais variáveis -> mais vantajoso que o R^2

- 👉 Avaliar modelos com mais precisão e segurança
- 👉 É aplicável em modelos multivariados
- 👉 Não apresenta um viés dependente dos dados inseridos

Regressão - Métricas de avaliação

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

- 👍 Excelente para problemas nos quais grandes erros não são tolerados.
- 👎 Carece de interpretabilidade direta.

Regressão - Métricas de avaliação

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

- 👍 Melhora a interpretabilidade do MSE.
- 👎 Sensível a previsões muito distantes.

Regressão - Métricas de avaliação

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

- 👍 Não pune severamente outliers.
- 👍 Interpretação mais intuitiva.
- 👎 Não é adequado para problemas delicados.

Regressão - Métricas de avaliação

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- 👍 Interpretação intuitiva (%).
- 👎 Sensível a grandes variações.

Regressão - Métricas de avaliação

$$\text{RMSLE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(y_i + 1) - \log \hat{y}_i + 1))}$$

Evita penalização de diferenças elevadas quando **ambos** os valores são muito grandes

Clusterização - Métricas de Avaliação

Método do Cotovelo:

$$\text{WCSS} = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

onde k = o número de clusters, n = o número de objetos, x_i = $i^{\text{ésimo}}$ elemento no cluster e c_j = o centroide do $j^{\text{ésimo}}$ cluster.

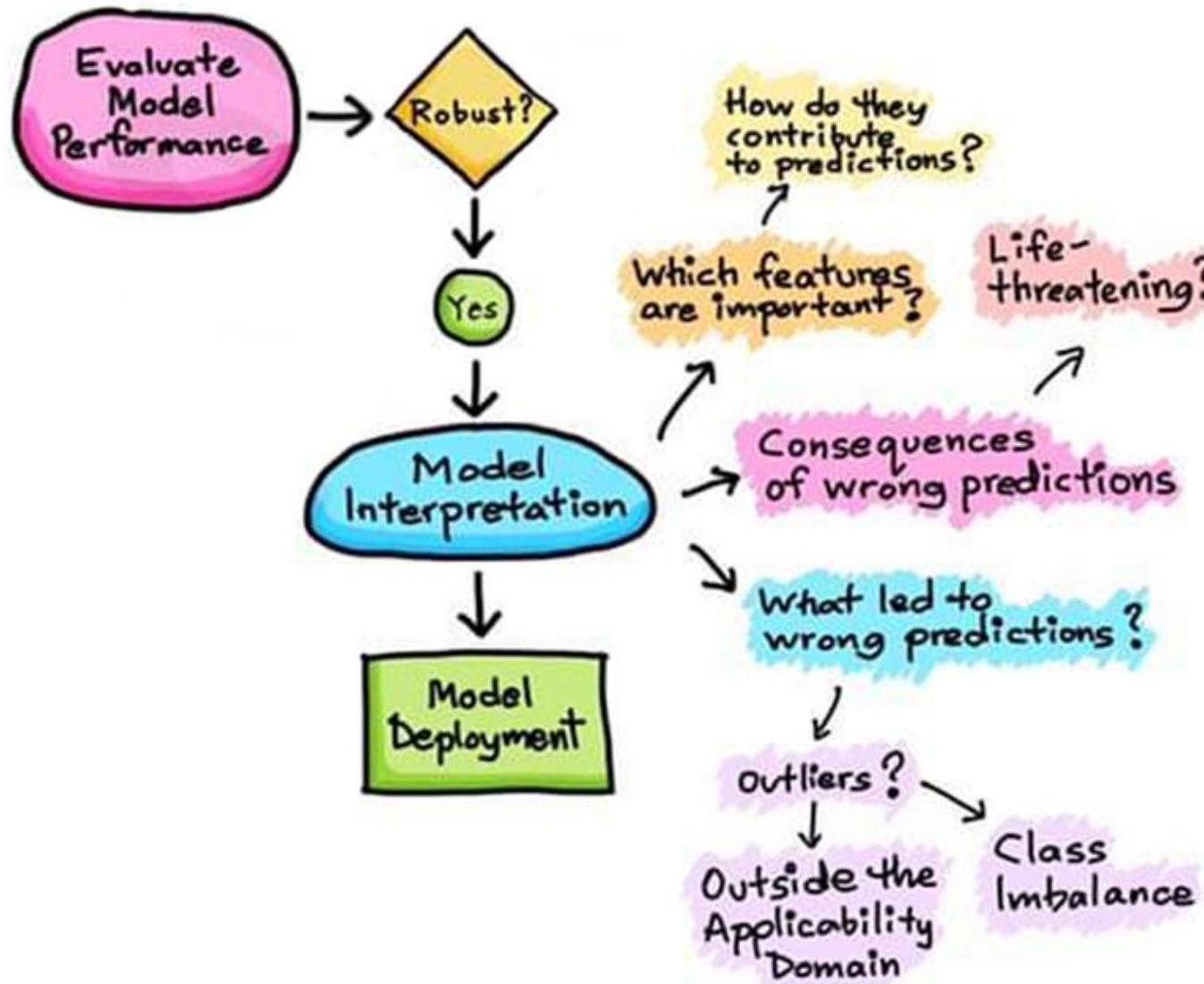
Índice de Silhueta:

$$s(o) = \frac{b(o) - a(o)}{\max\{a(o), b(o)\}}$$

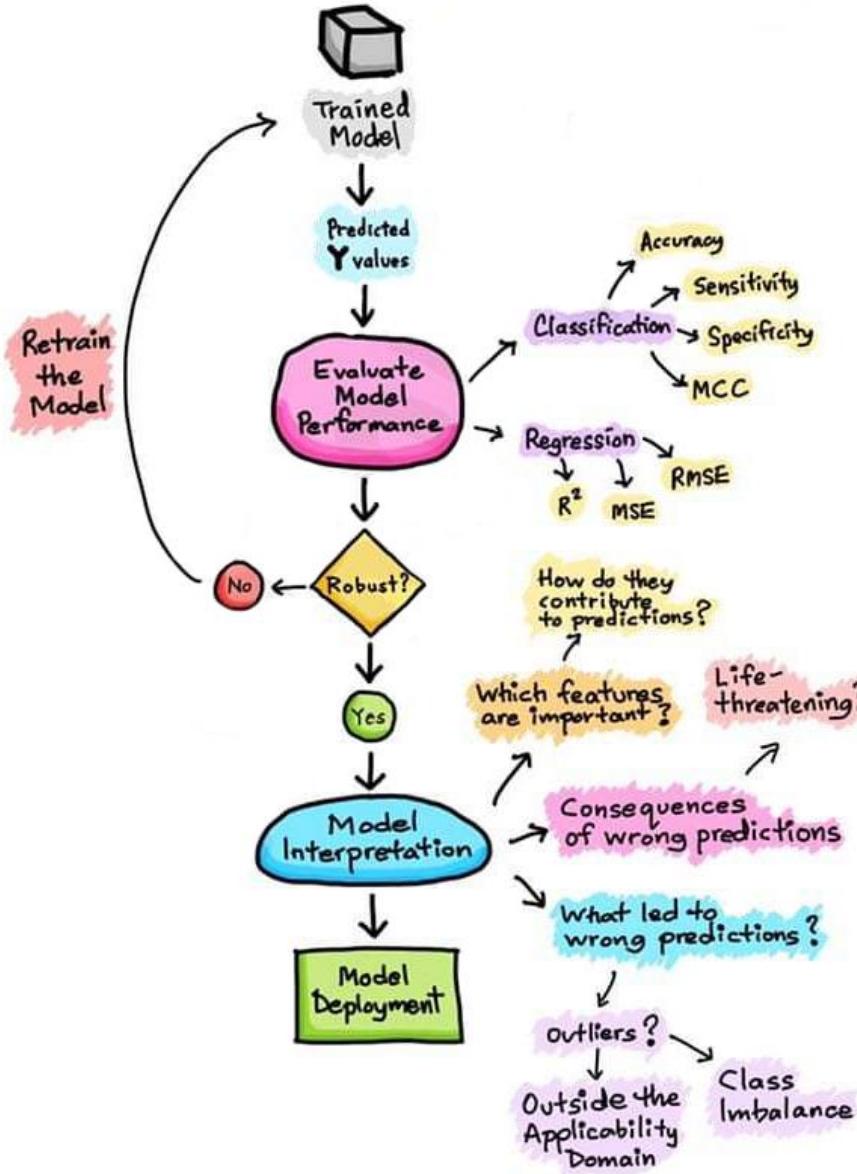
onde $s(o)$ = representa o coeficiente de silhueta para o ponto o ,
 $a(o)$ é a distância média entre o e todos os demais pontos do seu
cluster e
 $b(o)$ é a distância média mínima de o a todos os clusters que ele não
faz parte.

O resultado é um número entre 1 e -1, sendo
1 a melhor silhueta, -1 significa que o dado está no cluster errado e 0
significa que não existem clusters significativos.

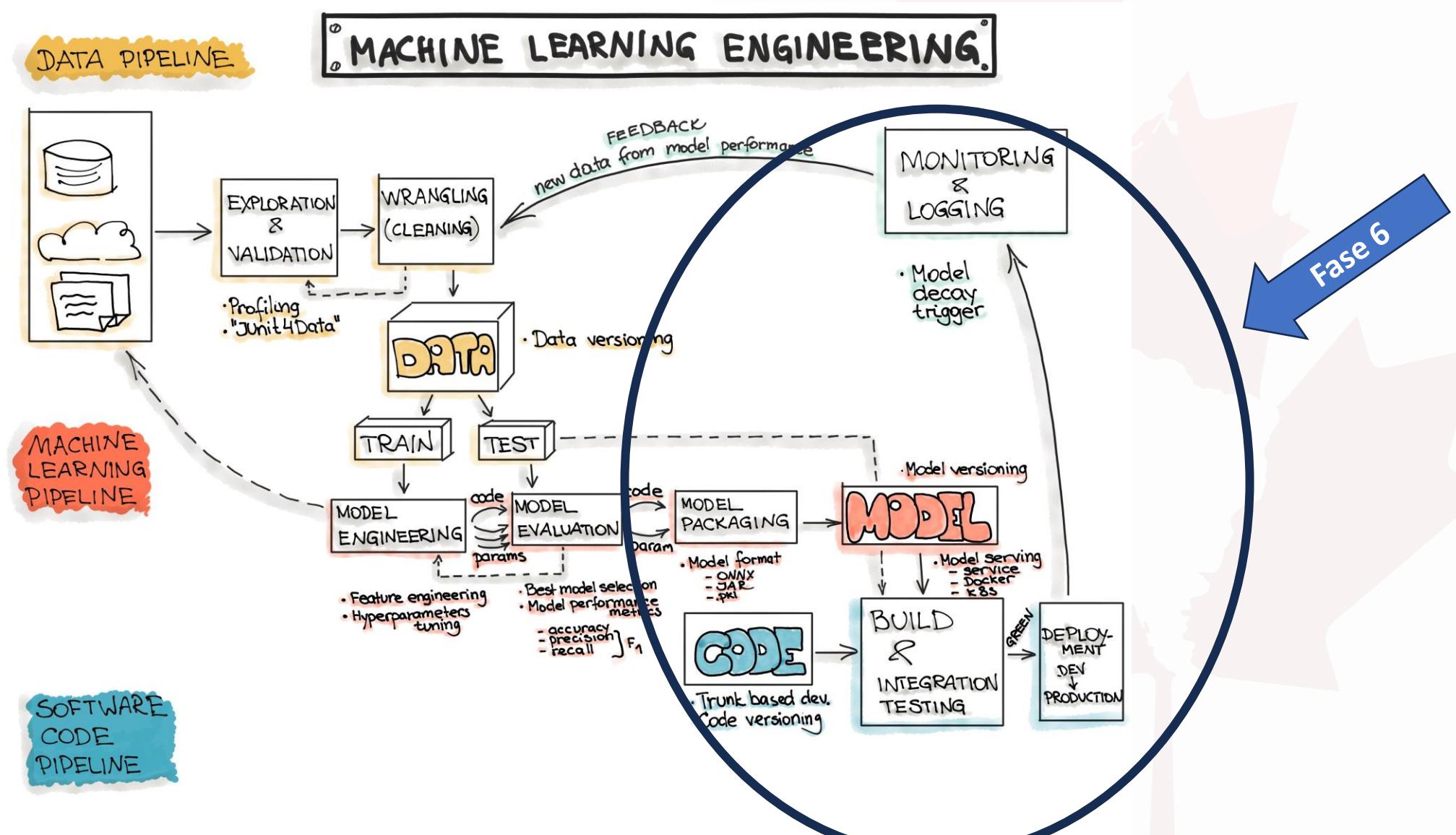
Avaliando o Modelo



A finalização do processo



O ciclo completo do modelo de ML



Extra – Vibe Coding

Aulas do Prof. Onédio Seabra no curso de IA para Projetos Sustentáveis sobre Vibe Coding:

<https://www.youtube.com/watch?v=pL38Lb66qy4>

<https://www.youtube.com/watch?v=Kd4Nj-NeB1Q>

<https://www.youtube.com/watch?v=VgFeF7JJzXo>

Dúvidas?





Dall.e

<https://labs.openai.com/e/xYvsdZh7P3kY7OTbeD4RqdD6>

Desafio 4

“Um caso real”

Desafio 4 – Um caso real

O objetivo

- O objetivo desta atividade é gerar uma planilha de compra de vales refeição que deve ser enviada mensalmente para a operadora.
- A descrição do processo se encontra no arquivo “Desafio 4 - Descrição.pdf” que foi postado na nossa área de arquivos compartilhados.
- Nesta área também existe um arquivo compactado chamado “Desafio 4 – Dados.zip”, o qual contém as planilhas Excel referenciadas no texto da descrição do desafio.

O que deve ser feito

- Cada grupo deverá ler detalhadamente a descrição do desafio, analisar os arquivos disponíveis, obter as convenções coletivas de trabalho vigentes de cada um dos sindicatos.
- De posse de todas estas informações, o grupo deve construir os agentes que executam os passos do processo de modo a gerar a planilha final desejada, observando todas as regras descritas.

Desafio 4 – Um caso real

- **A entrega será por e-mail**
 - O responsável pelo grupo deverá enviar uma mensagem para o endereço challenges@i2a2.academy onde deve anexar a planilha final gerada e todos os códigos fonte criados (sejam notebooks Jupyter, scripts Python, arquivos .json do N8N, modelos, diagramas, etc.), com cópia para si mesmo.
 - O título do e-mail deverá ser “**Agentes Autônomos – Desafio 4**”.
 - Opcionalmente, o responsável pelo grupo poderá copiar os demais membros do grupo e desta forma gerar um “protocolo de entrega” adicional.
- O limite para entrega é **20/08/2025 às 23h59**.
- **Esta atividade É ELIMINATÓRIA. A ausência da entrega elimina todos os participantes do GRUPO**
- No dia 21/08/2025 teremos a apresentação de alguns grupos.
- Aguardamos que voluntários se manifestem.
- Quem desejar fazer sua apresentação, informe no corpo do e-mail.

Perguntas?





PRACTICE
MAKES
PERFECT

Obrigado!

i2a2.academy



+55 16 99213-2650



celso@i2a2.academy



/in/celso-augusto-morato-azevedo