



Machine Learning com R

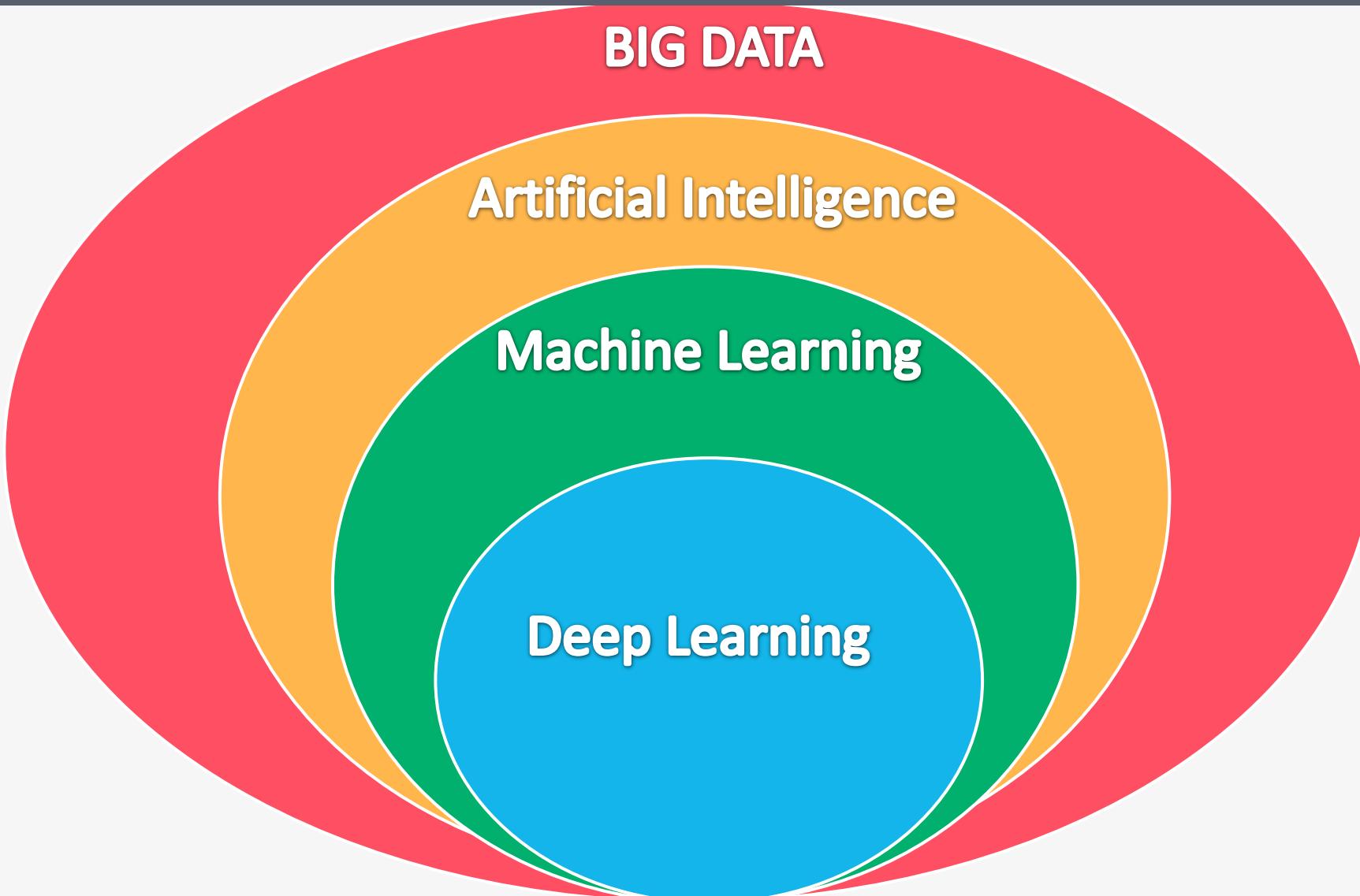
Seja bem-vindo!!!

Machine Learning

Machine Learning

- Algoritmos com Cálculos estatísticos/probabilidade;
- Aprende a partir de um conjunto de dados;
- Apresenta a probabilidade de tal evento acontecer com base dados apresentados;
- Agrupa dados semelhantes;
- o objetivo principal é fazer o algoritmo **GENERALIZAR** e NÃO decorar.

Machine Learning



Machine Learning



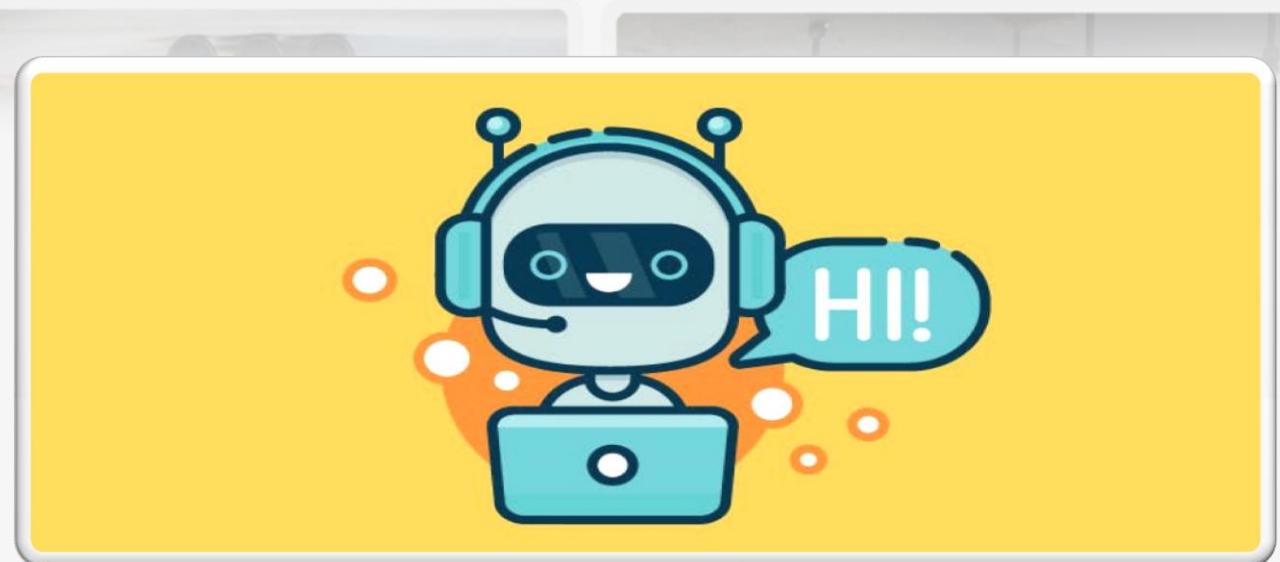
Machine Learning

- E O FUTURO???



Machine Learning

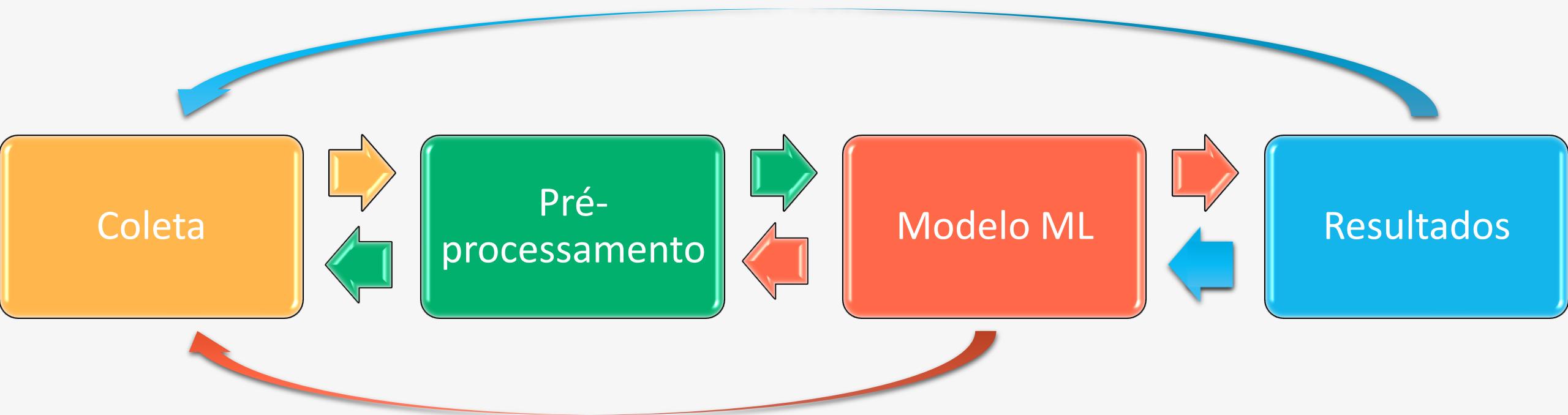
- E O FUTURO???



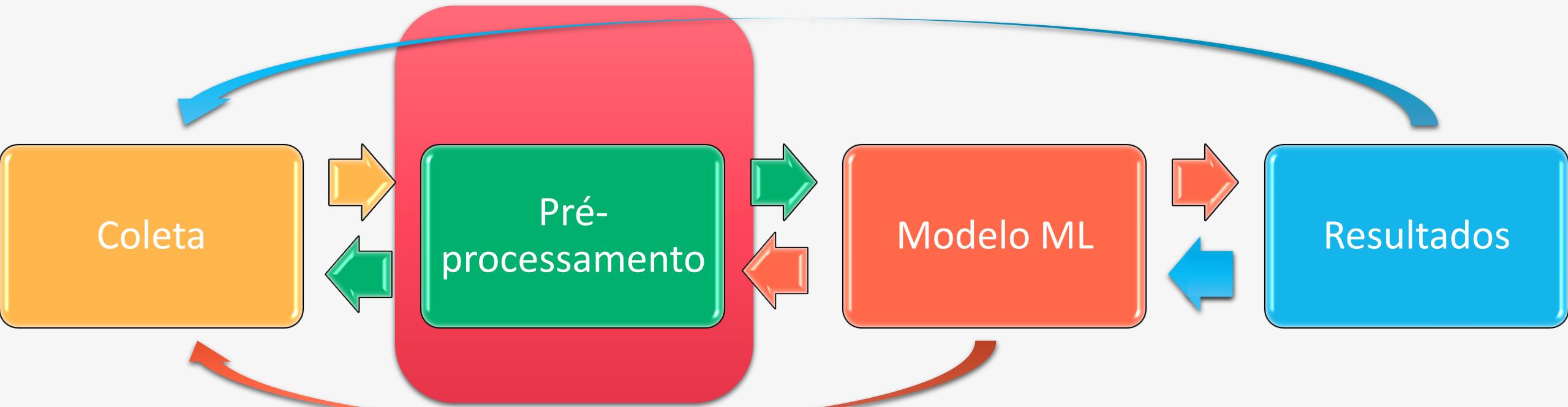
Machine Learning

- <https://www.youtube.com/watch?v=aFuA50H9uek>
- <https://www.youtube.com/watch?v=-Wnp-OOZB34>

Machine Learning - Processo

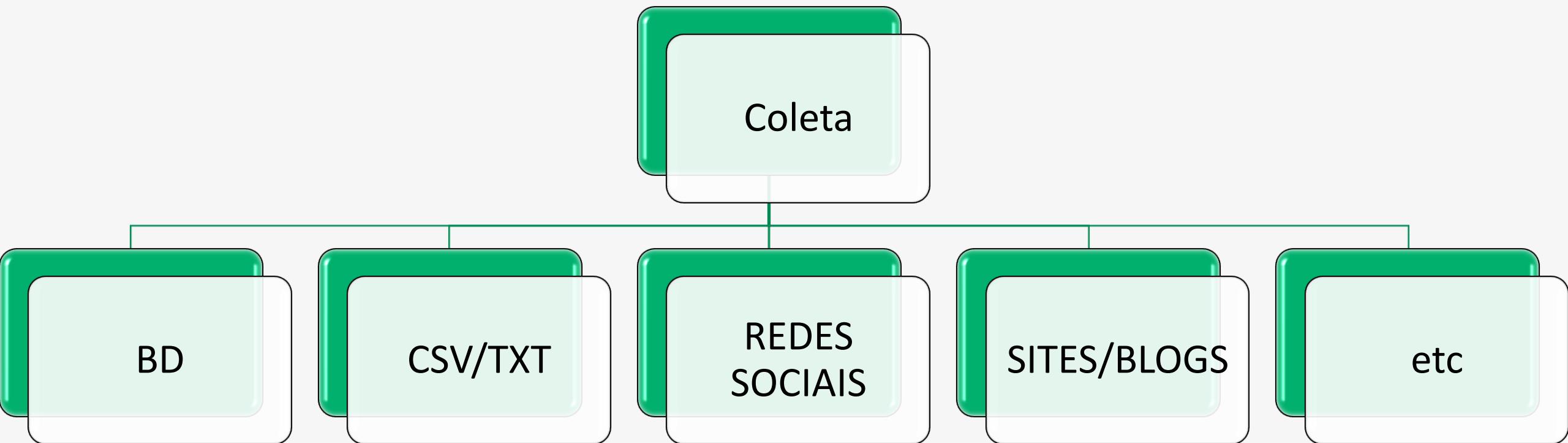


Machine Learning - Processo

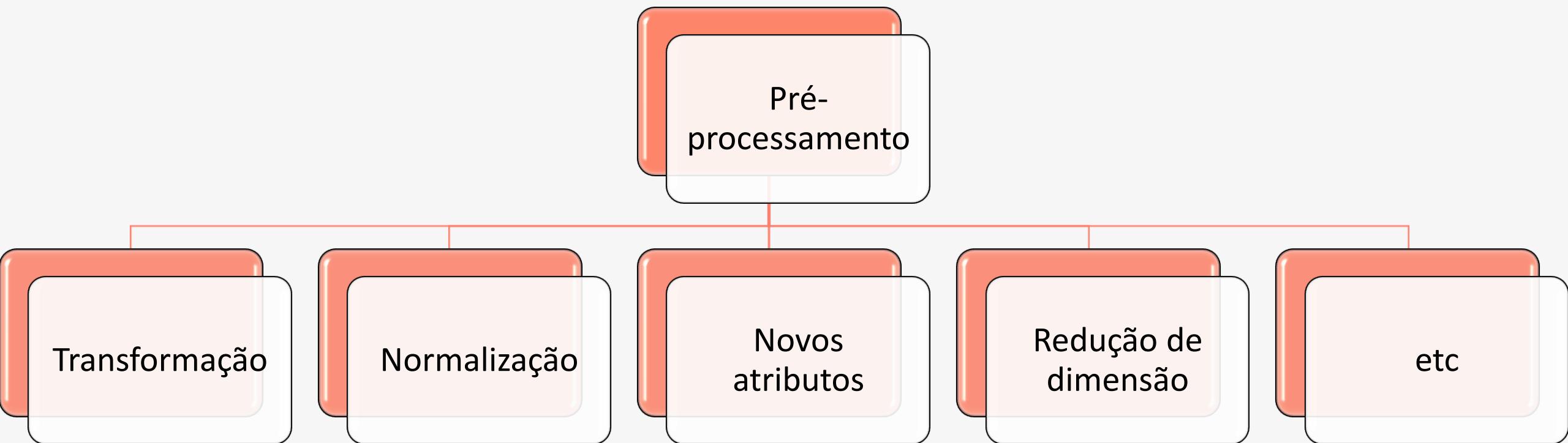


Até 80% tempo
do projeto de ML

Machine Learning - Processo



Machine Learning - Processo



Machine Learning - Processo

- **Pré-processamento:**
 - **Normalização:** dados homogêneos, por exemplo: letras minúsculas, números inteiros, números entre -1 e 1 ou 0 e 1 etc.
 - **Transformação:** mudar os dados, por exemplo: através da idade criar categorias como, criança, adolescente, adulto, idoso.

Machine Learning - Processo

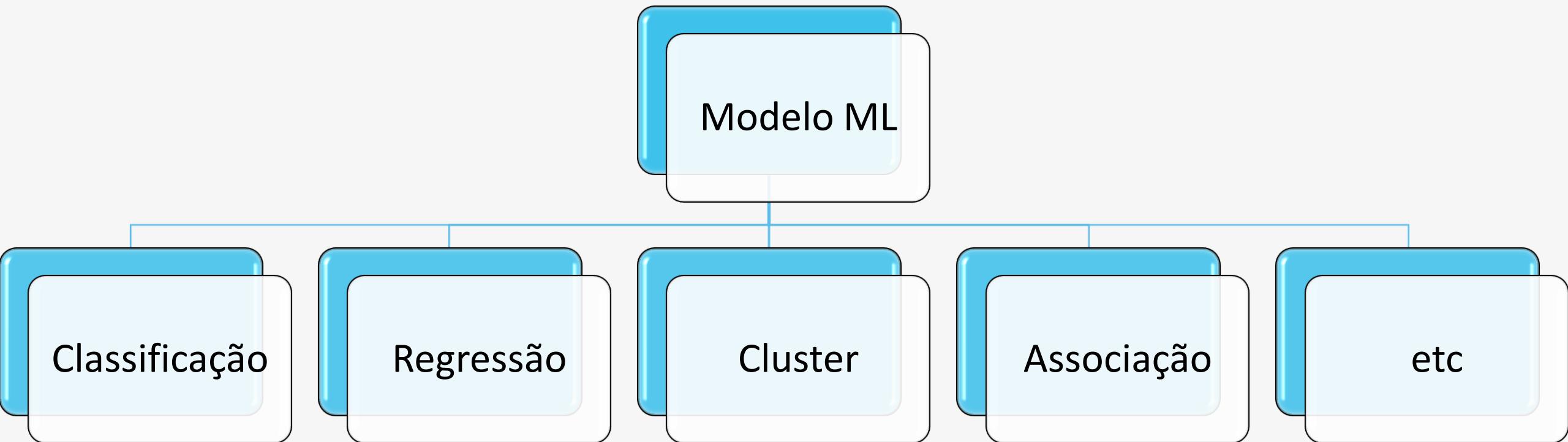
- **Pré-processamento:**
 - **Redução de dimensão:** diminuir a quantidade de atributos (colunas). Geralmente são utilizados algoritmos, como, PCA, para realizar essa redução de forma semi-automatizada.
 - **Novos atributos (colunas):** inserir um nova coluna através de outros dados, por exemplo, a partir da data de nascimento, calcular a idade e criar uma nova coluna com estes dados.

Machine Learning - Processo

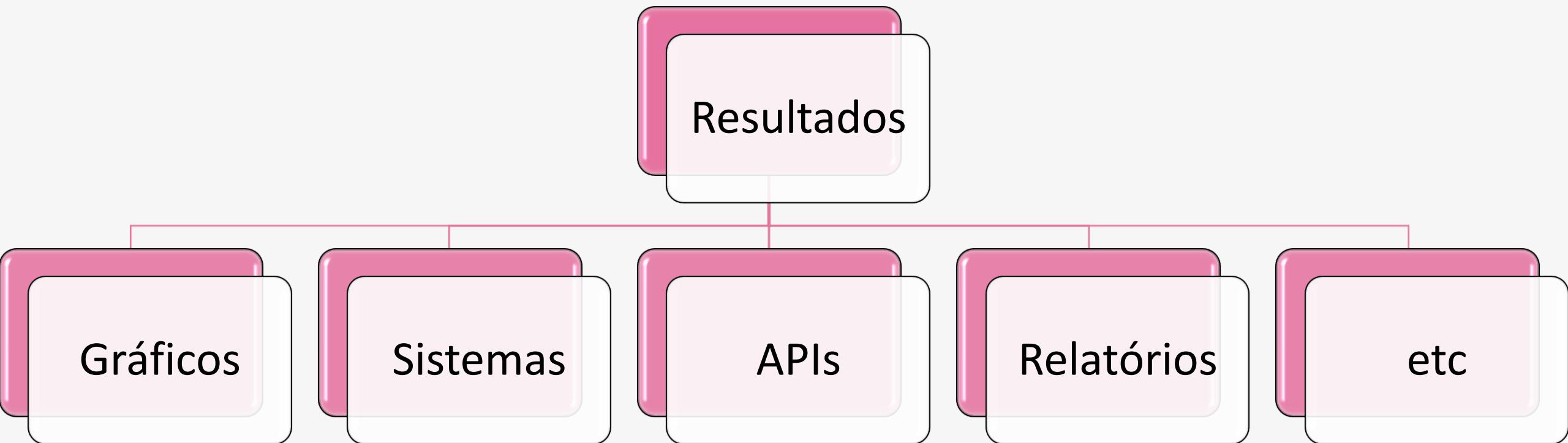
- Pré-processamento: ATENÇÃO!!!!!!!

**TODO PRÉ-PROCESSAMENTO REALIZADO
PARA CONSTRUIR O MODELO, TAMBÉM
DEVE SER FEITO NOS NOVOS DADOS.**

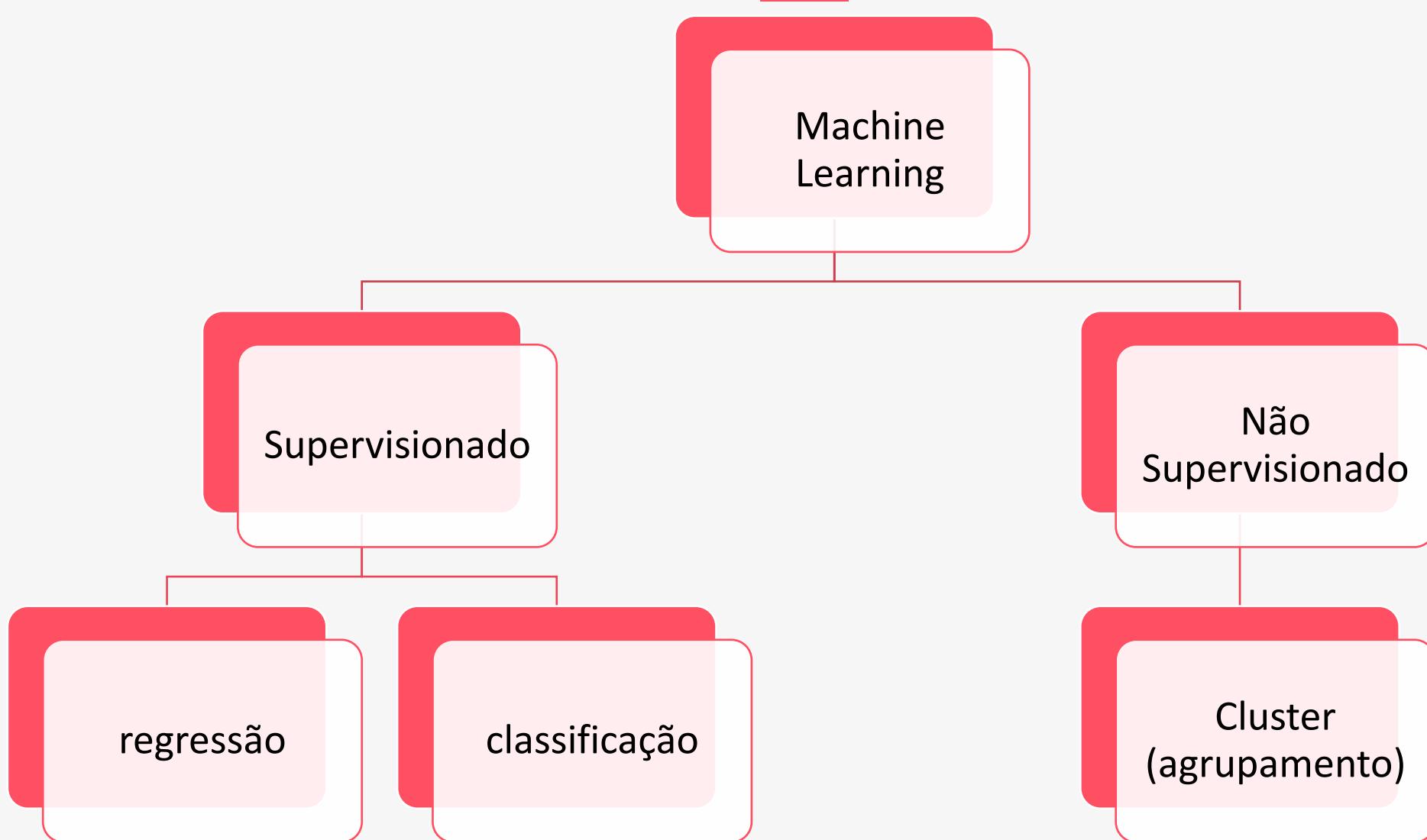
Machine Learning - Processo



Machine Learning - Processo



Machine Learning

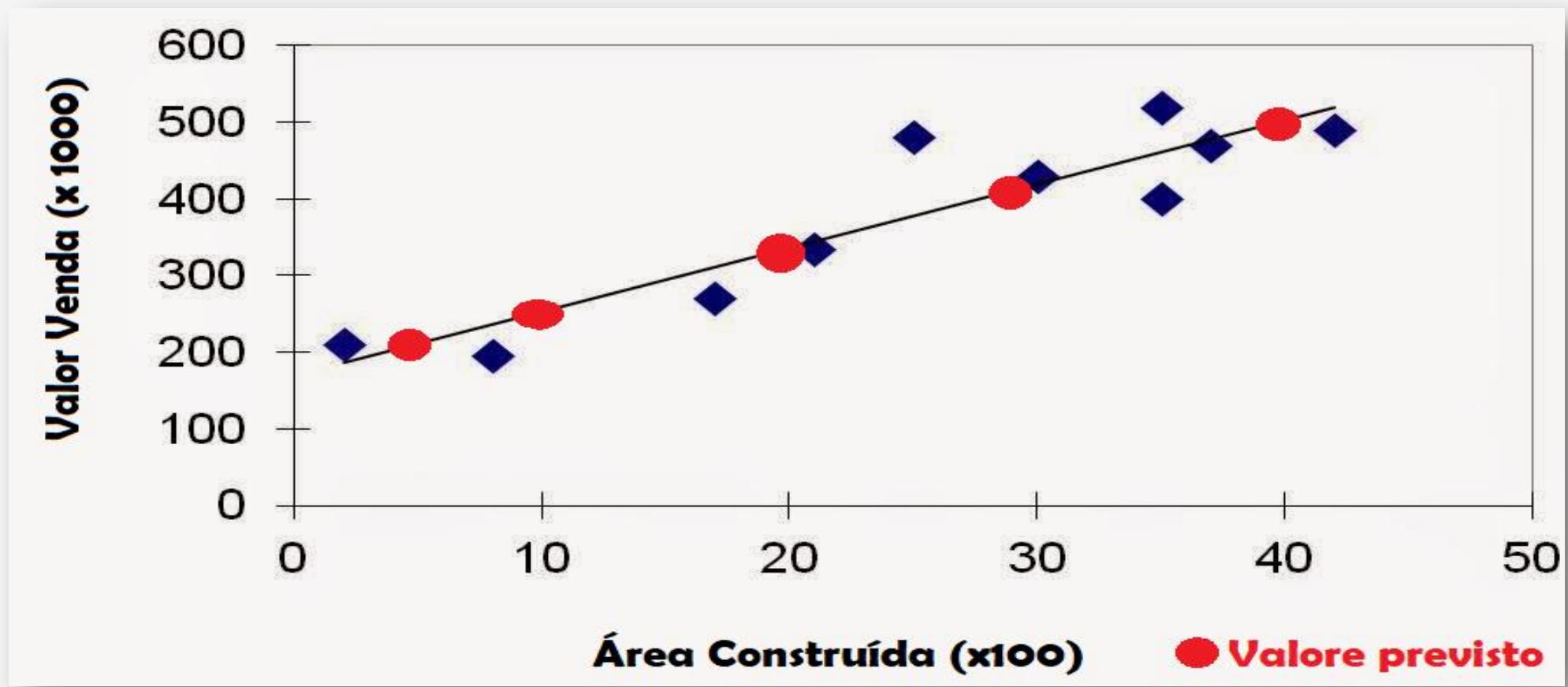


Machine Learning

- Aprendizado supervisionado:
 - deve ter uma variável classe(target);
 - com base nos dados, faz classificação para determinada classe (target);
- Exemplo:
 - prever o valor valor de uma casa baseada no tamanho dela;
 - classificar uma frutas com base nas características.

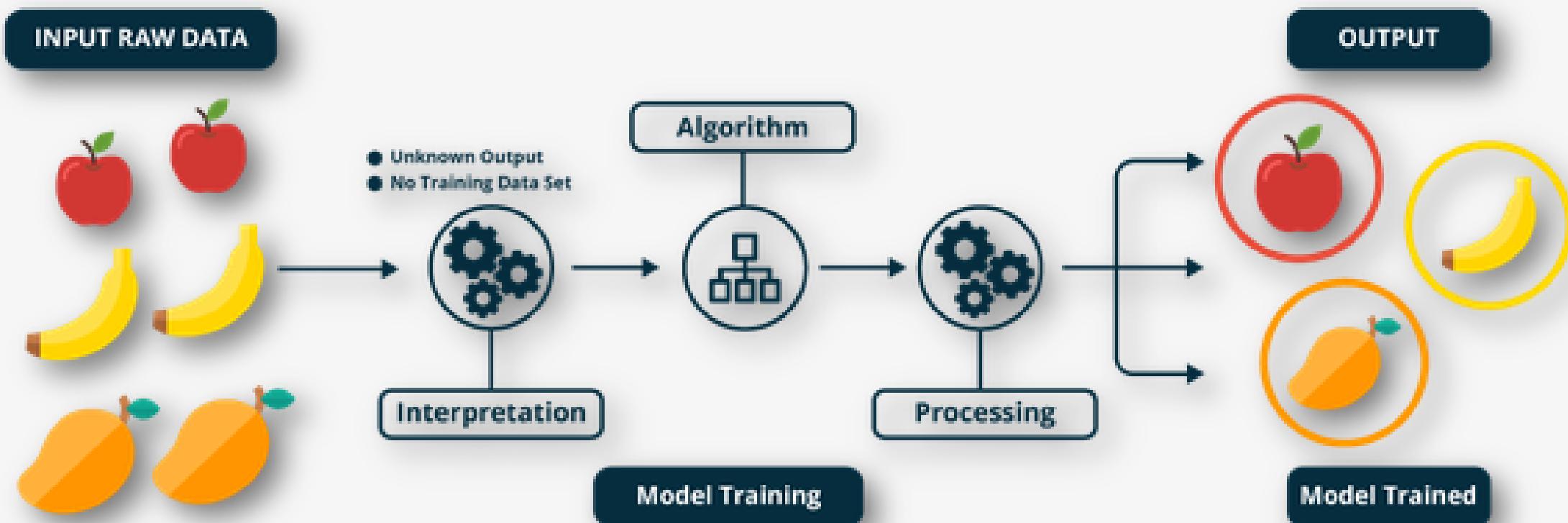
Machine Learning

- Aprendizado supervisionado: regressão



Machine Learning

- Aprendizado supervisionado: classificação

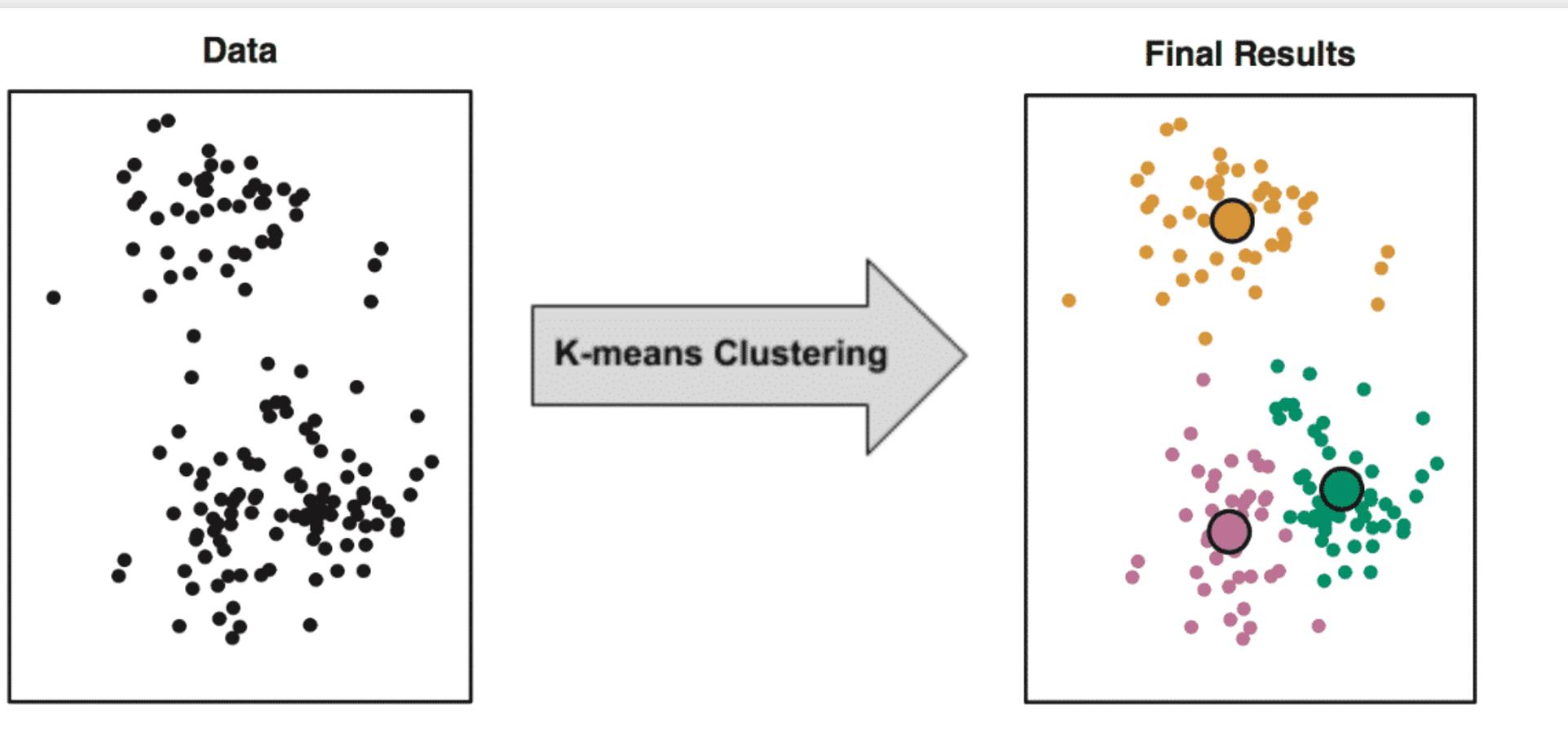


Machine Learning

- Aprendizado Não supervisionado:
 - Não precisa ter uma classe(target);
 - Exemplo:
 - agrupar clientes com perfis semelhantes;
 - identificar preferências de produtos com os cliente.

Machine Learning

- Aprendizado não supervisionado: cluster (agrupamento)



Machine Learning

- **Algoritmos:**

- **Regressão:** linear regression, polynomial regression, random forest, DL;
- **Classificação:** Naïve Bayes, Decision Trees, Random Forest, SVM, DL;
- **Cluster:** K-means, DBSCAN, HDBSCAN, DL.

Machine Learning

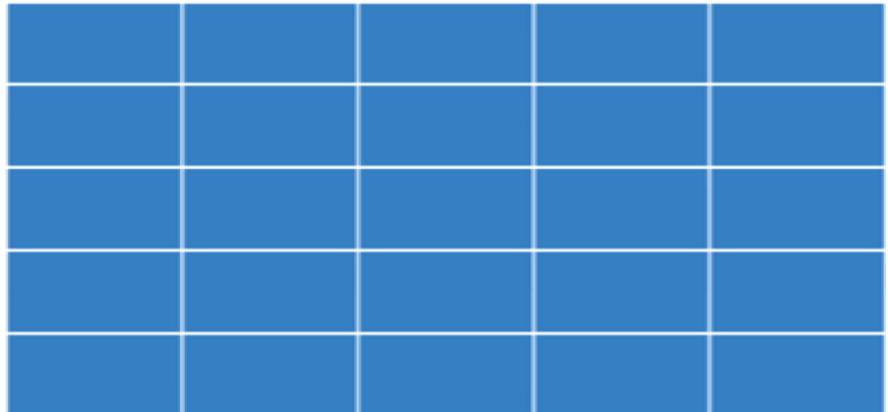
- **DADOS DE TREINO x DADOS DE TESTE**

- **Dados de treino:** dados utilizados para treinar o modelos. É utilizado entre 70% e 80% do total da base de dados.
- **Dados de teste:** dados utilizados para verificar se o modelo está ruim ou bom, se aprendendeu padrão ou decorou os valores.

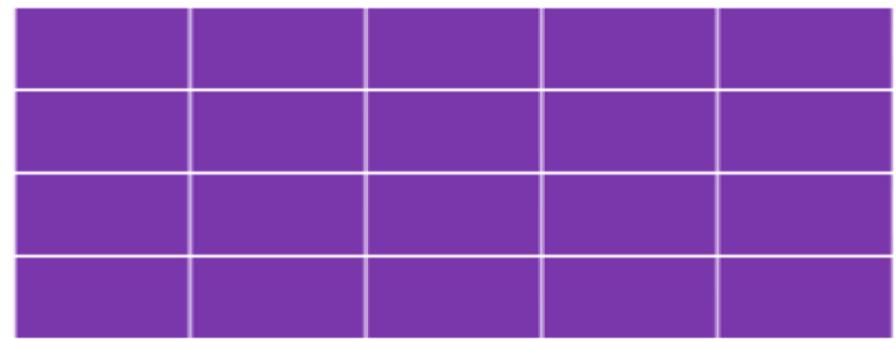
Machine Learning

- DADOS DE TREINO x DADOS DE TESTE

Original Data



Training Set



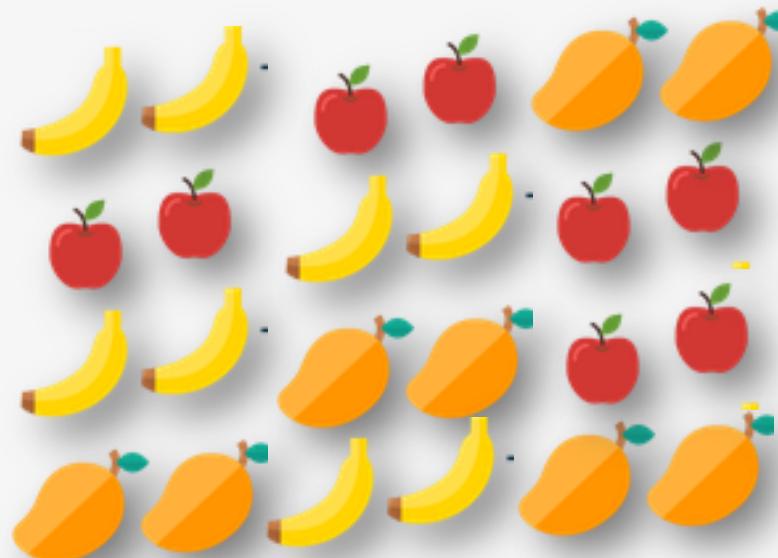
Test Set



Machine Learning

- DADOS DE TREINO x DADOS DE TESTE

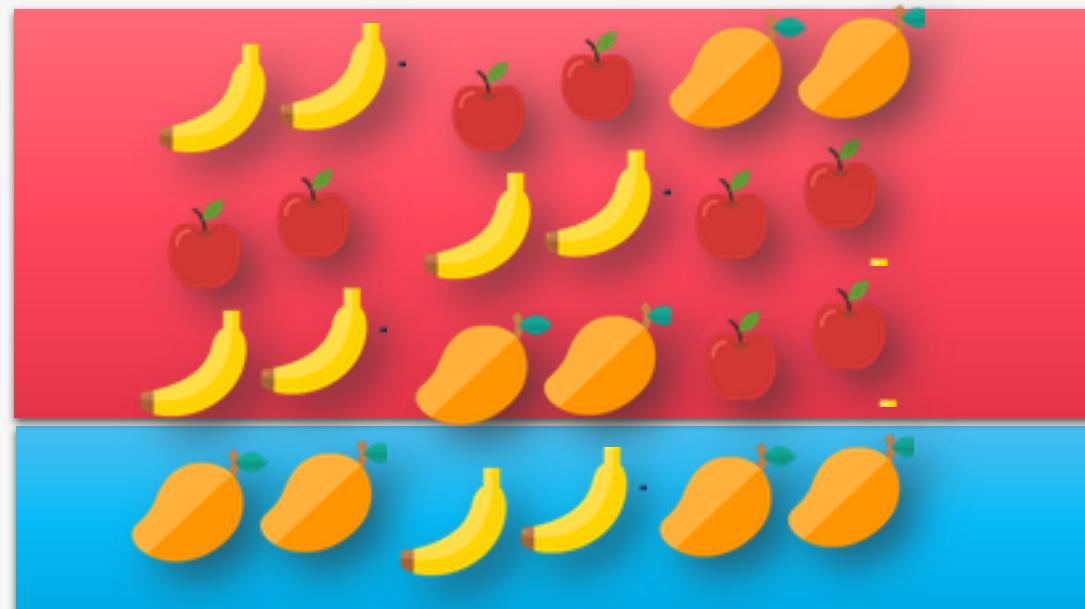
DADOS (POPULAÇÃO)



Machine Learning

- DADOS DE TREINO x DADOS DE TESTE

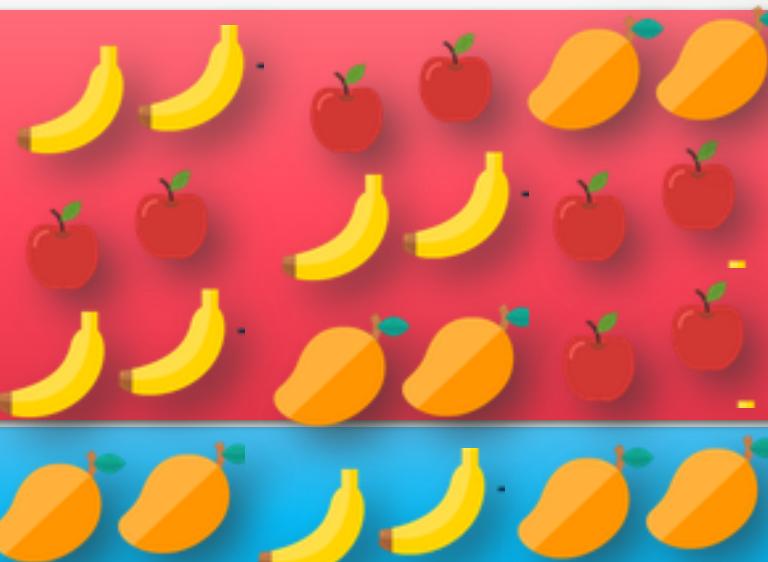
DADOS (POPULAÇÃO)



Machine Learning

- DADOS DE TREINO x DADOS DE TESTE

DADOS (POPULAÇÃO)



TREINO (AMOSTRA)



TESTE (AMOSTRA)



Machine Learning

- Avaliando o modelo de ML
 - Verificar o quanto o modelo está acertando/prevendo os resultados;
 - Utilizado em modelos de ML Supervisionado;
 - Pode ser utilizado: **Matrix Confusion, Accuracy, recall, precision, fscore.**

Machine Learning

OVERFITTING

X

UNDERFITTING

Machine Learning

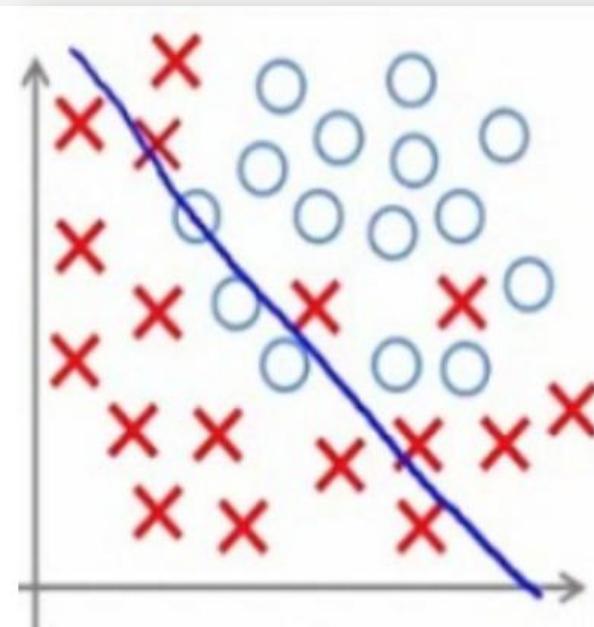
- **Overfitting:**
 - Indica que o modelo de ML não aprendeu os padrões nos dados, mas sim, **DECOROU** os dados;
 - Pode apresentar alguma métrica de validação alta. (**accuracy**);
 - Quando for inserido um dado diferente do que estava na base de treino, o modelo pode apresentar um resultado errado;
 - Pode acontecer por falta de registros, por falta de pré-processamento, colunas que podem ter alta correlação com a classe (target).

Machine Learning

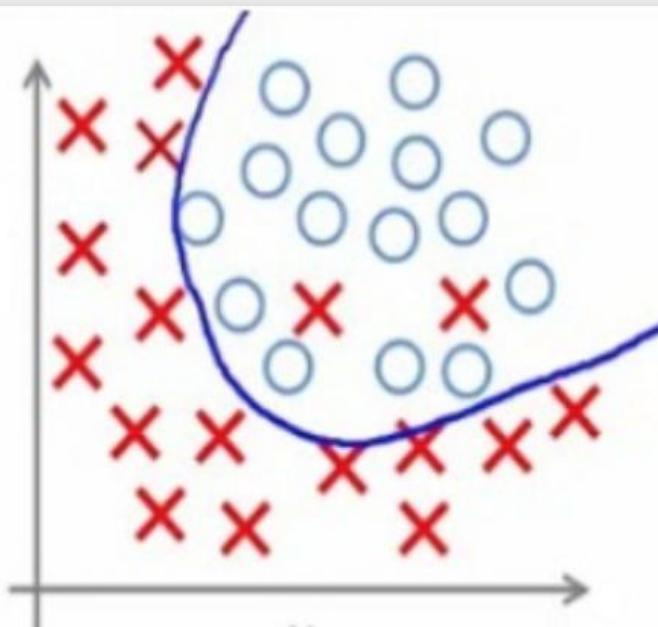
- **Underfitting:**
 - indica que o modelo de ML não conseguiu aprender os padrões nos dados;
 - As métricas de validação podem ser muito baixos;
 - Pode acontecer por falta de registro, falta de pré-processamento, colunas que não são importantes para o modelo.

Machine Learning

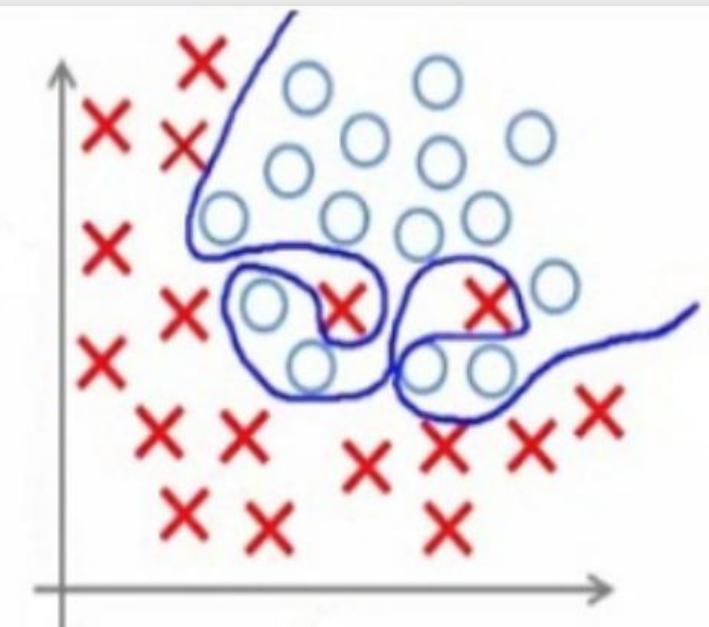
- Underfitting x Overfitting:



Under-fitting



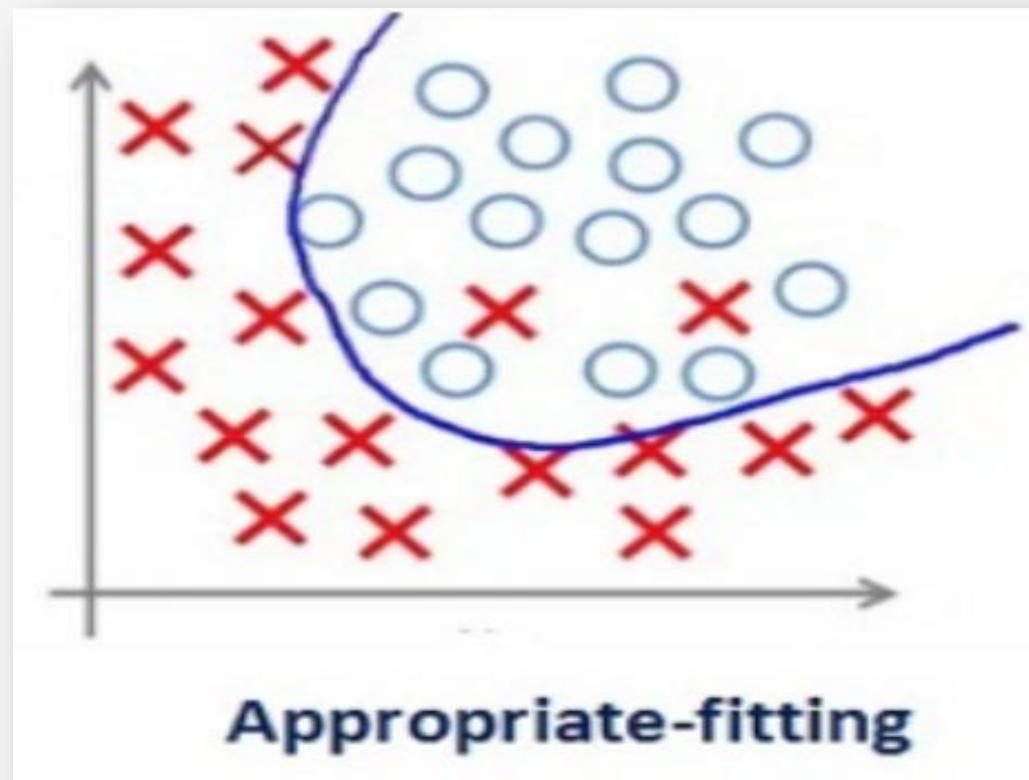
Appropriate-fitting



Over-fitting

Machine Learning

- O IDEAL (MODELO CONSEGUE GENERALIZAR)



Referências

- <https://towardsdatascience.com/what-is-machine-learning-891f23e848da>
- https://www.sas.com/pt_br/insights/analytics/machine-learning.html
- <https://medium.com/data-hackers/entendendo-o-que-%C3%A9-matriz-de-confus%C3%A3o-com-python-114e683ec509>
- <https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall>
- <https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124>

Referências

- <https://towardsdatascience.com/various-ways-to-evaluate-a-machine-learning-models-performance-230449055f15>
-