



Aprendizagem Automática em Sistemas Empresariais

PEDRO PEREIRA

AULA 7

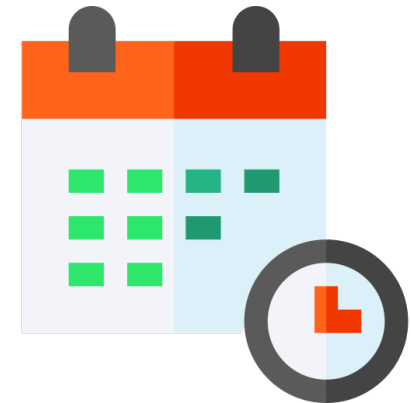


Agenda

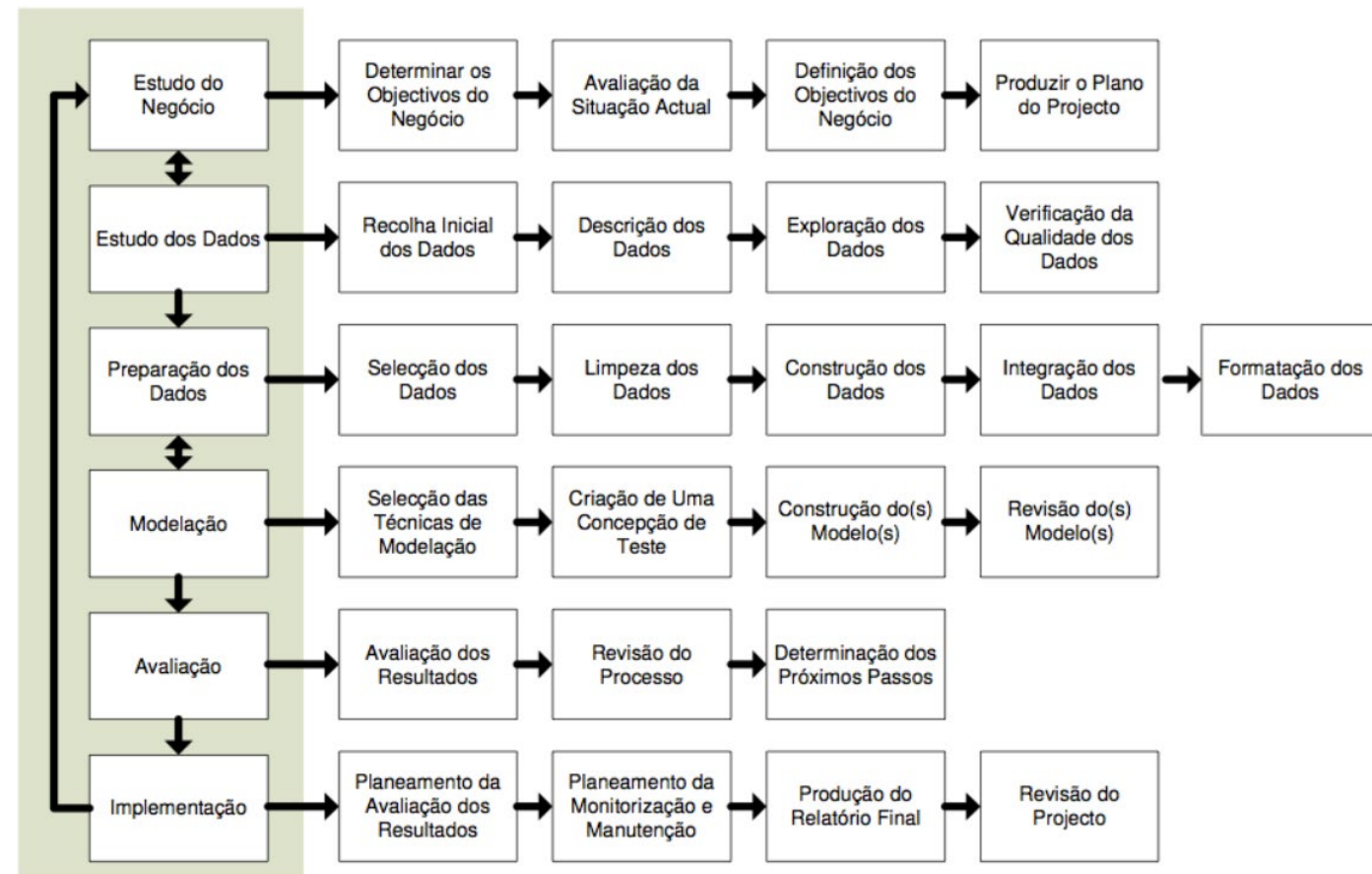
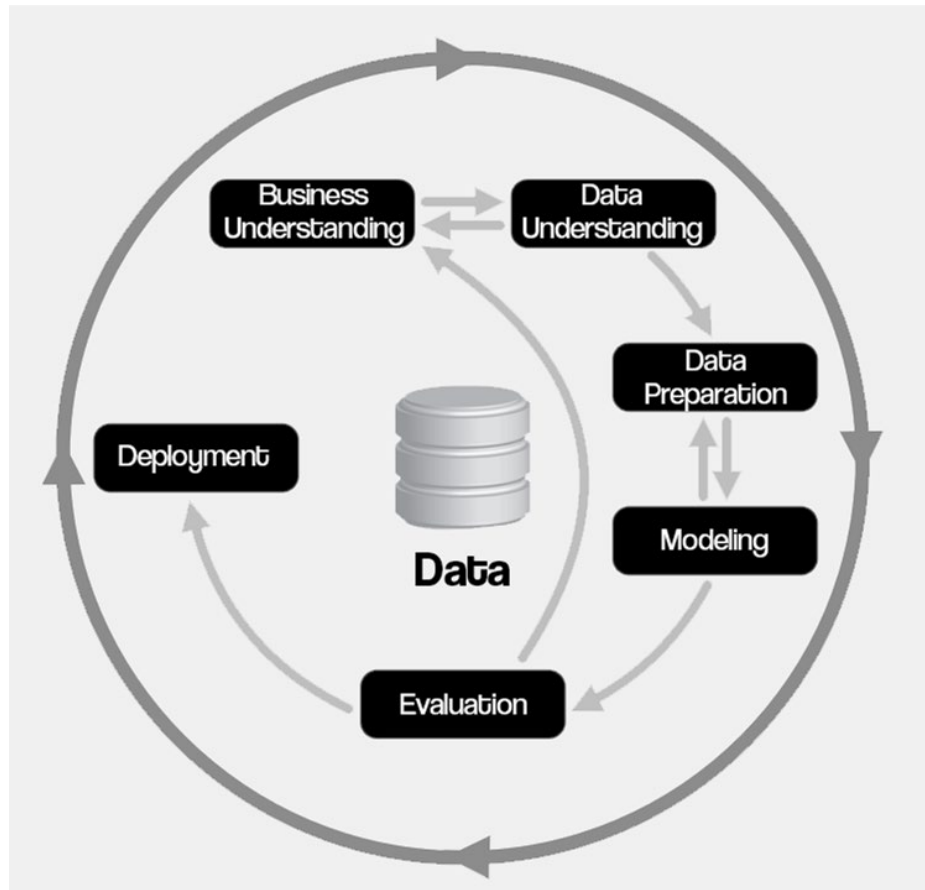
CRISP-DM: Avaliação

- Validação de Modelos
- Cross-validation

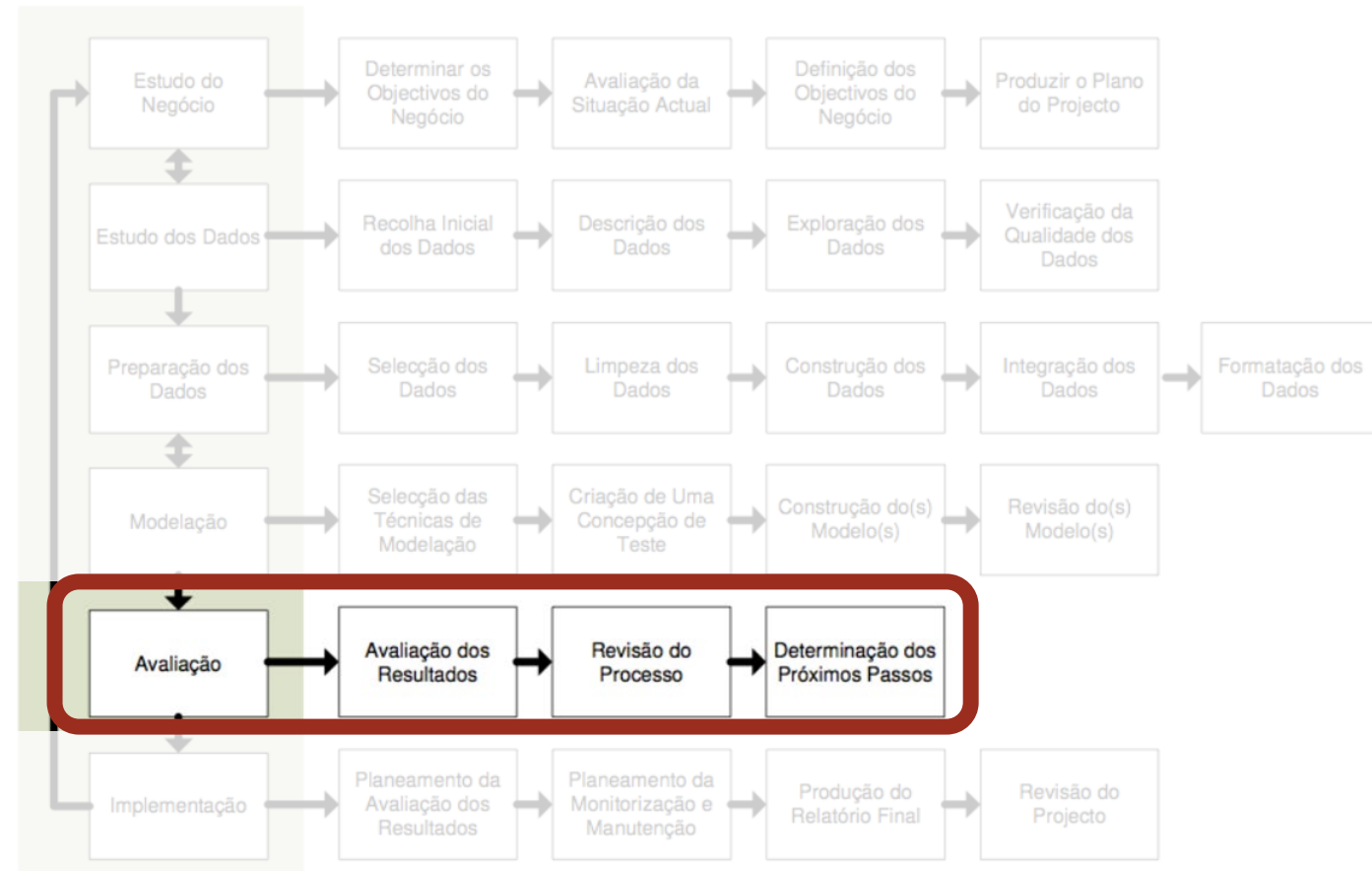
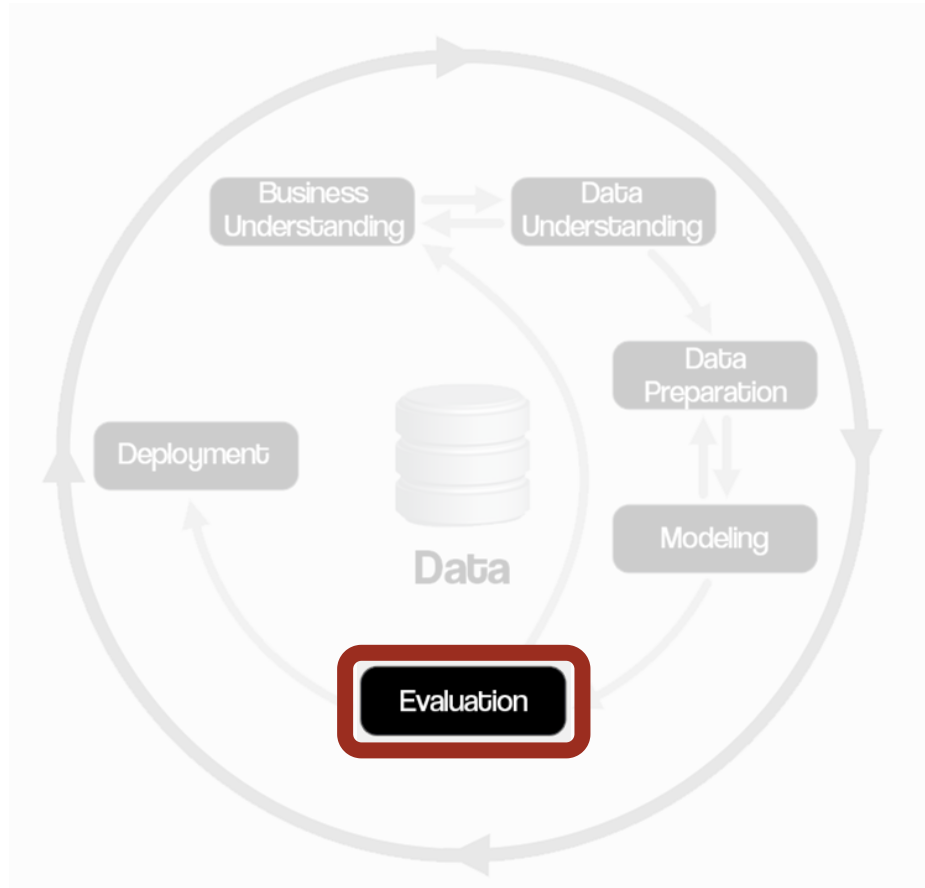
Acompanhamento ao projeto



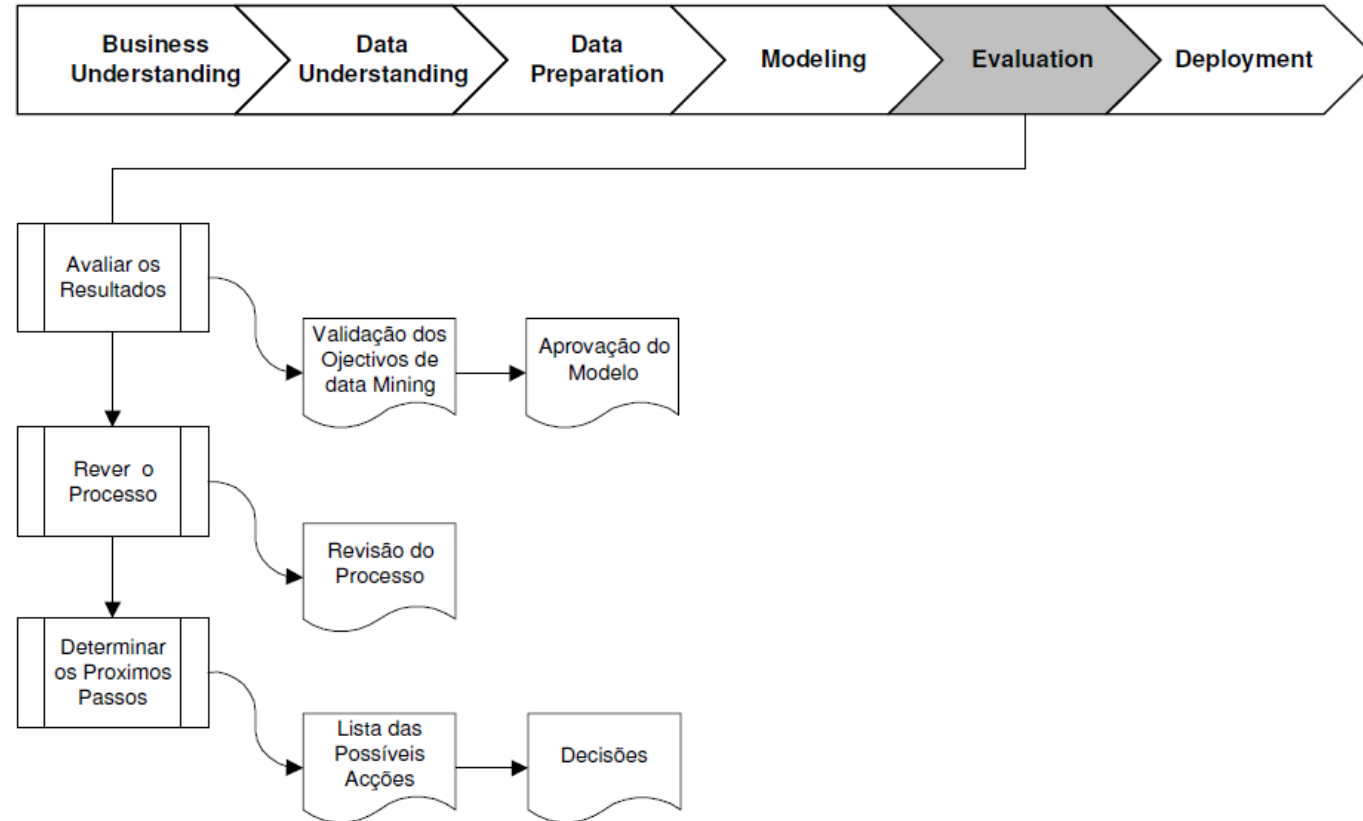
Cross Industry Process for Data Mining (CRISP-DM)



CRISP-DM – Avaliação



CRISP-DM – Atividades da Modelação e Avaliação





Avaliação: validação de modelos

A validação de modelos pretende “estimar” a sua capacidade de generalização, medindo a sua qualidade/desempenho.

Como tal, as métricas **não podem ser calculadas utilizando dados que o modelo já “viu”**.

Holdout: divisão dos dados em dois conjuntos exclusivos, através de uma amostragem aleatória.

- **Treino:** tipicamente **2/3 do conjunto dos dados**, usado para treinar modelos e tomar decisões (melhor modelo, melhores hiper-parâmetros, melhor pré-processamento,...). Por vezes, este conjunto é subdividido em 2 conjuntos (**treino** e **validação**) para verificar decisões internas do modelo.
- **Teste:** tipicamente **1/3 do conjunto dos dados**, é utilizado para avaliar as capacidades do modelo.

Treino

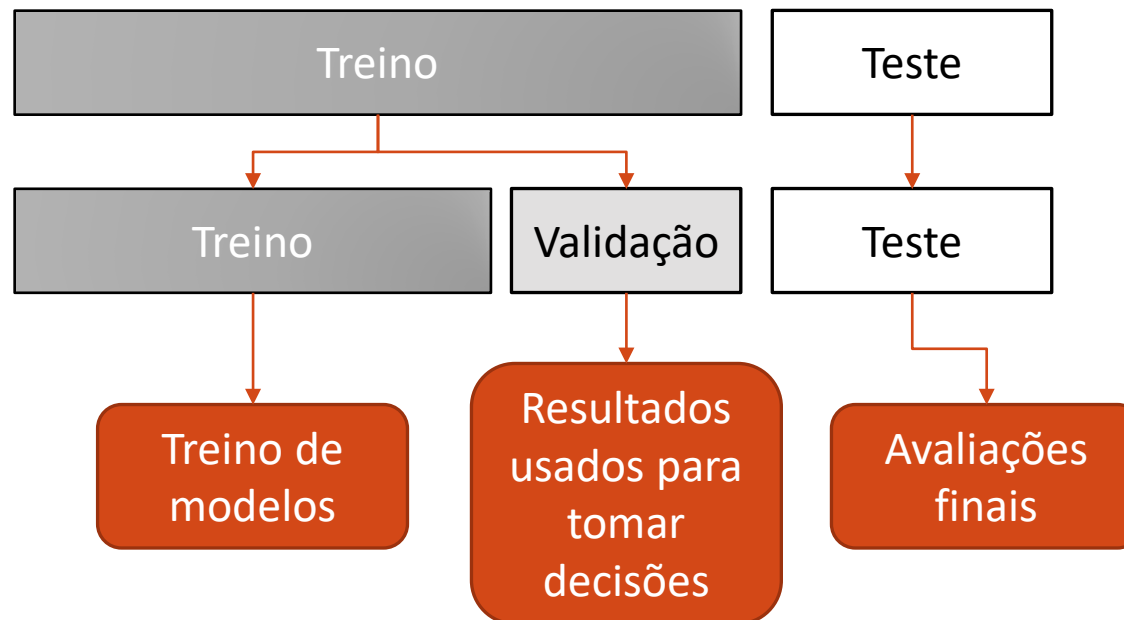
Teste

Avaliação: validação de modelos

Usando uma única divisão de dados, a credibilidade dos resultados é **reduzida**.

Quando têm de ser tomadas decisões sobre os dados (ex.: transformações, *oversampling*) ou sobre os modelos (ex.: hiper-parâmetros), usar um único *split* **não é viável!**

Nessas situações, faz-se uma nova divisão do conjunto de **treino** em **treino e validação**.





Avaliação: validação de modelos

Existem vários tipos de validações mais complexas e fiáveis → ***cross-validation***.

Consiste em várias divisões **diferentes** dos dados em treino e teste.

Várias iterações, diferentes conjuntos de treino e teste → processo mais **robusto** e **confiável**.

Em cada iteração, um modelo é treinado e avaliado, agregando-se os resultados de todas as iterações pela média (ou mediana).

Maior **custo computacional** (são treinados mais modelos).

Maior **credibilidade nos resultados** (menos influenciável pela aleatoriedade).

Avaliação: validação de modelos

K-fold: k partições e rotações dos dados (diferentes conjuntos de treino e teste), em cada rotação um modelo é treinado e avaliado.

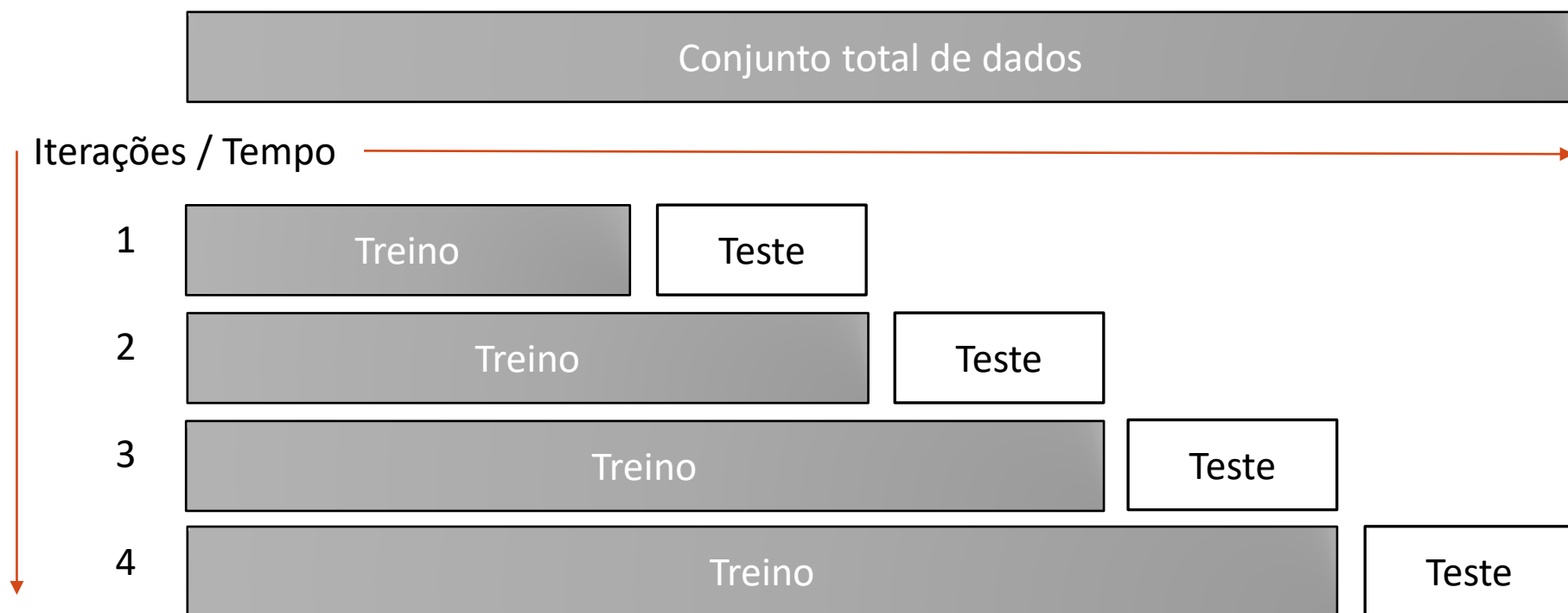
O mais utilizado é o ***10-fold cross-validation***.

	Conjunto total de dados			
Modelo 1	1	2	3	4
Modelo 2	1	2	3	4
Modelo 3	1	2	3	4
Modelo 4	1	2	3	4

Avaliação: validação de modelos

Growing window (ou incremental retraining): considera que os dados estão ordenados no tempo (últimos registos são mais recentes).

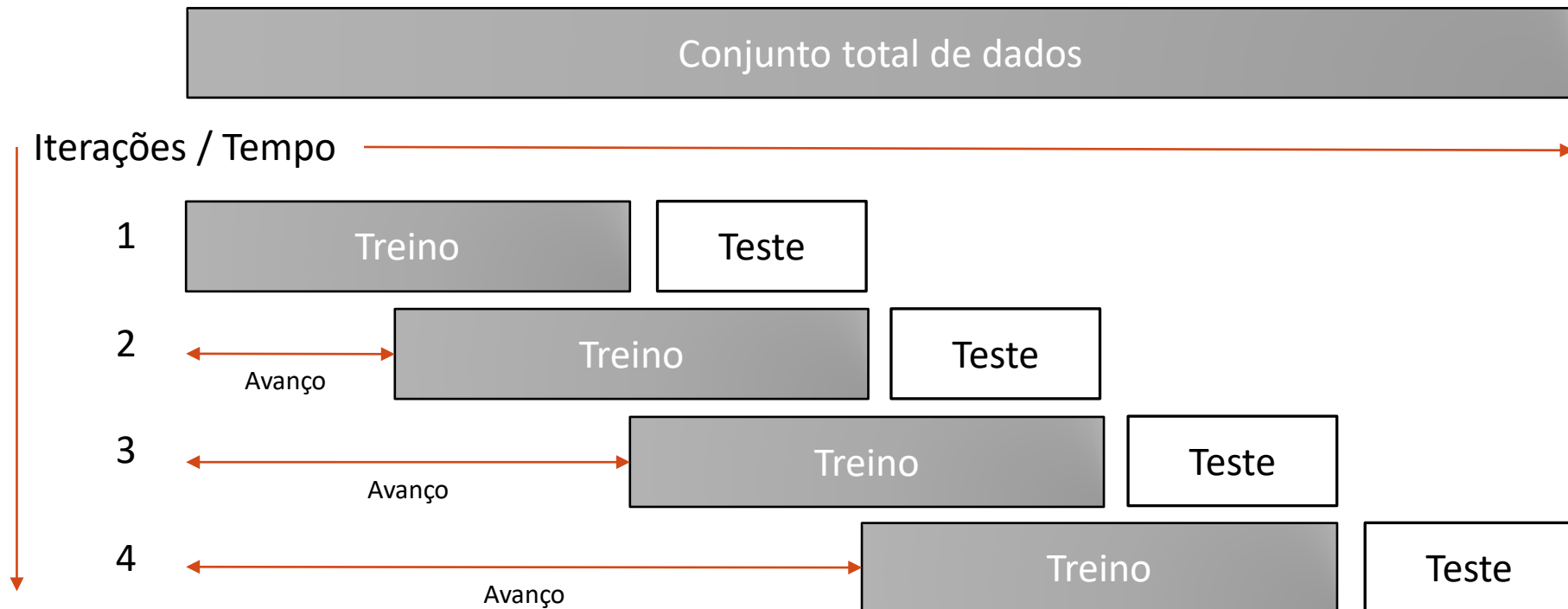
Consoante avançamos no tempo, o conjunto de dados **aumenta**.



Avaliação: validação de modelos

Rolling window: considera que os dados estão ordenados no tempo (últimos registos são mais recentes).

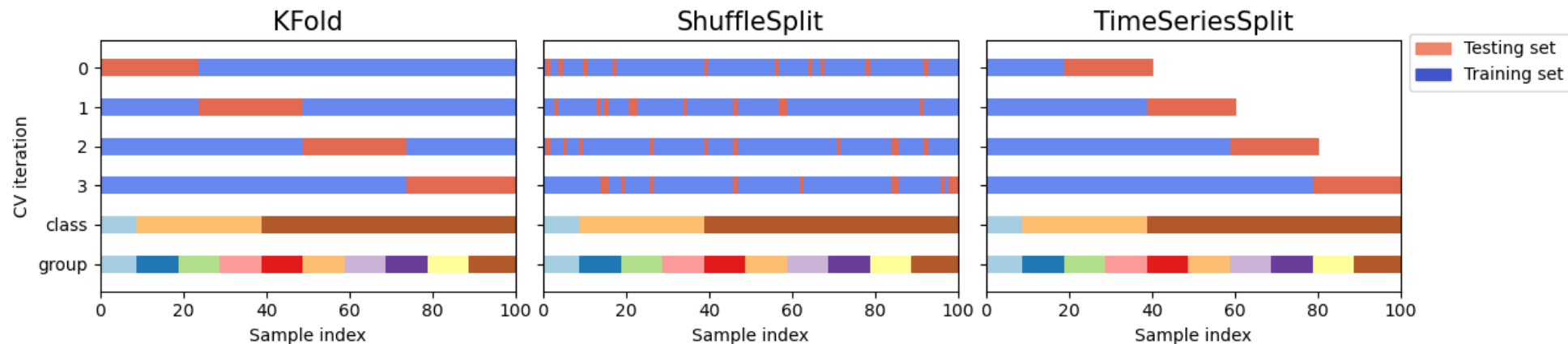
O conjunto de dados tem um tamanho **fixo**.



Avaliação: validação de modelos

Ferramentas:

- Python: https://scikit-learn.org/stable/modules/cross_validation.html
- R: <http://www.sthda.com/english/articles/38-regression-model-validation/157-cross-validation-essentials-in-r/>
- Rapidminer: https://docs.rapidminer.com/latest/studio/operators/validation/cross_validation.html





Aprendizagem Automática em Sistemas Empresariais

PEDRO PEREIRA

AULA 7