

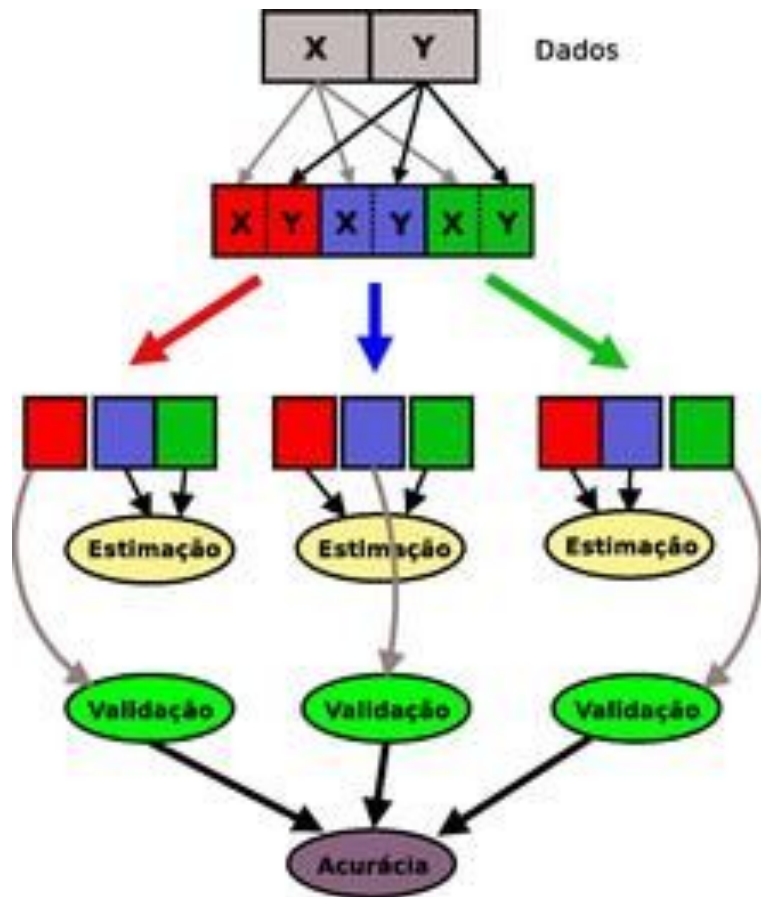
Cientista de Dados na Prática

```
346 .widget-area-sidebar {
347   font-size: 13px;
348 }
349
350
351
352 /* =Menu
353 -----
354
355 #access {
356   display: inline-block;
357   height: 69px;
358   float: right;
359   margin: 11px 28px 0px 0px;
360   max-width: 800px;
361 }
362
363 #access {
364   font-size: 13px;
365   list-style-type: none;
366   margin: 0 0 0 -0.8125em;
367   padding-left: 0;
368   z-index: 99999;
369   text-align: right;
370 }
371
372 #access {
373   display: inline-block;
374   text-align: left;
```



Seja Bem-Vindo!

Dicas e Truques



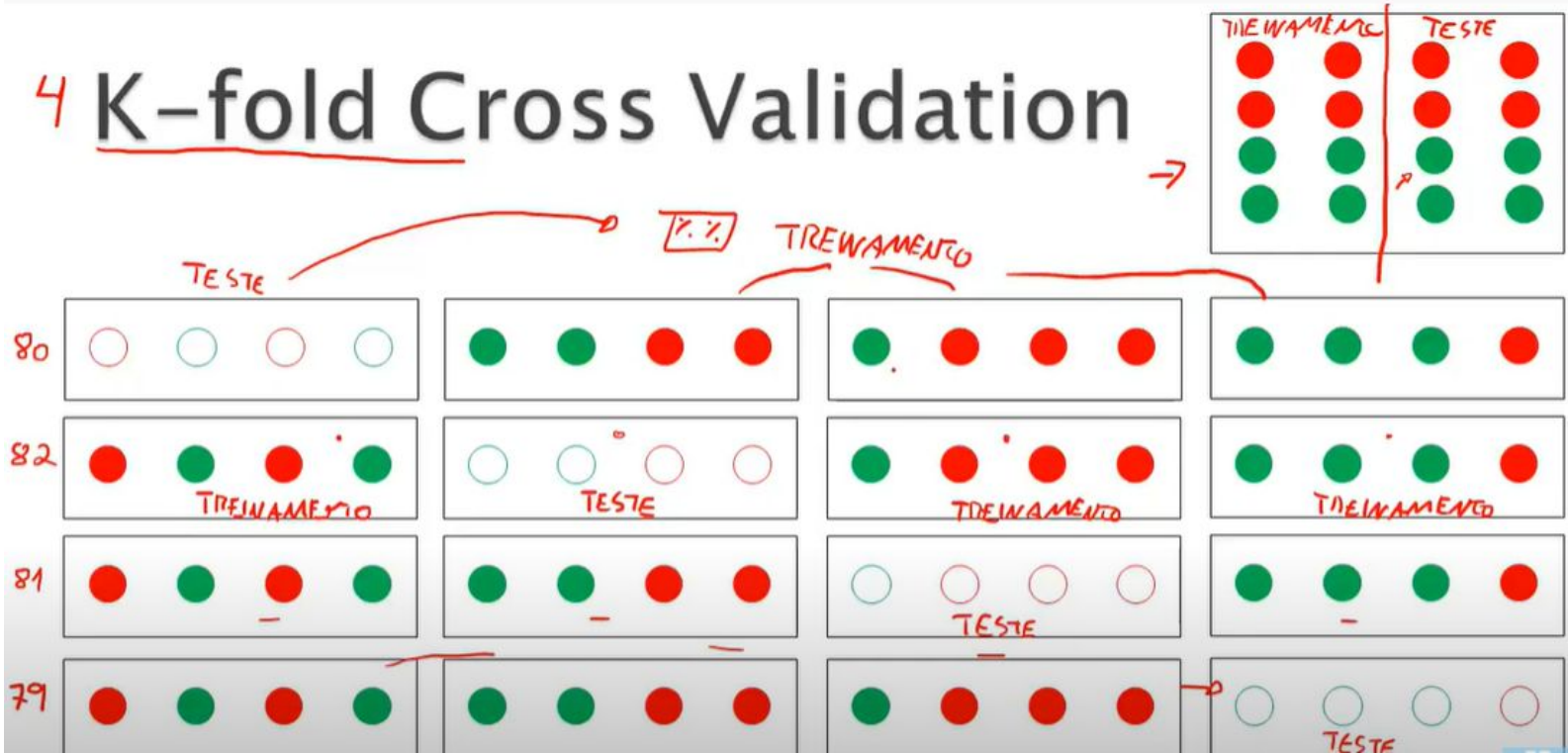
Validação Cruzada





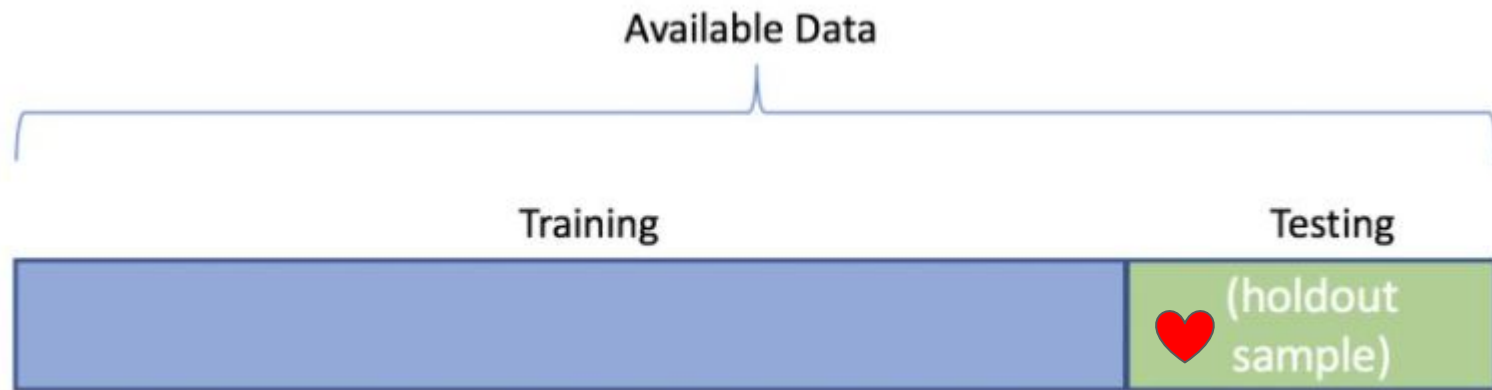
Por que usar CV?

4 K-fold Cross Validation





Imagina o seguinte:



registro importante para o conjunto de dados



Se o registro importante♥ ficou somente na amostra de teste, seu resultado não será tão bom quanto poderia ser se vc tivesse utilizado ele também na fase de treinamento da sua Máquina Preditiva.



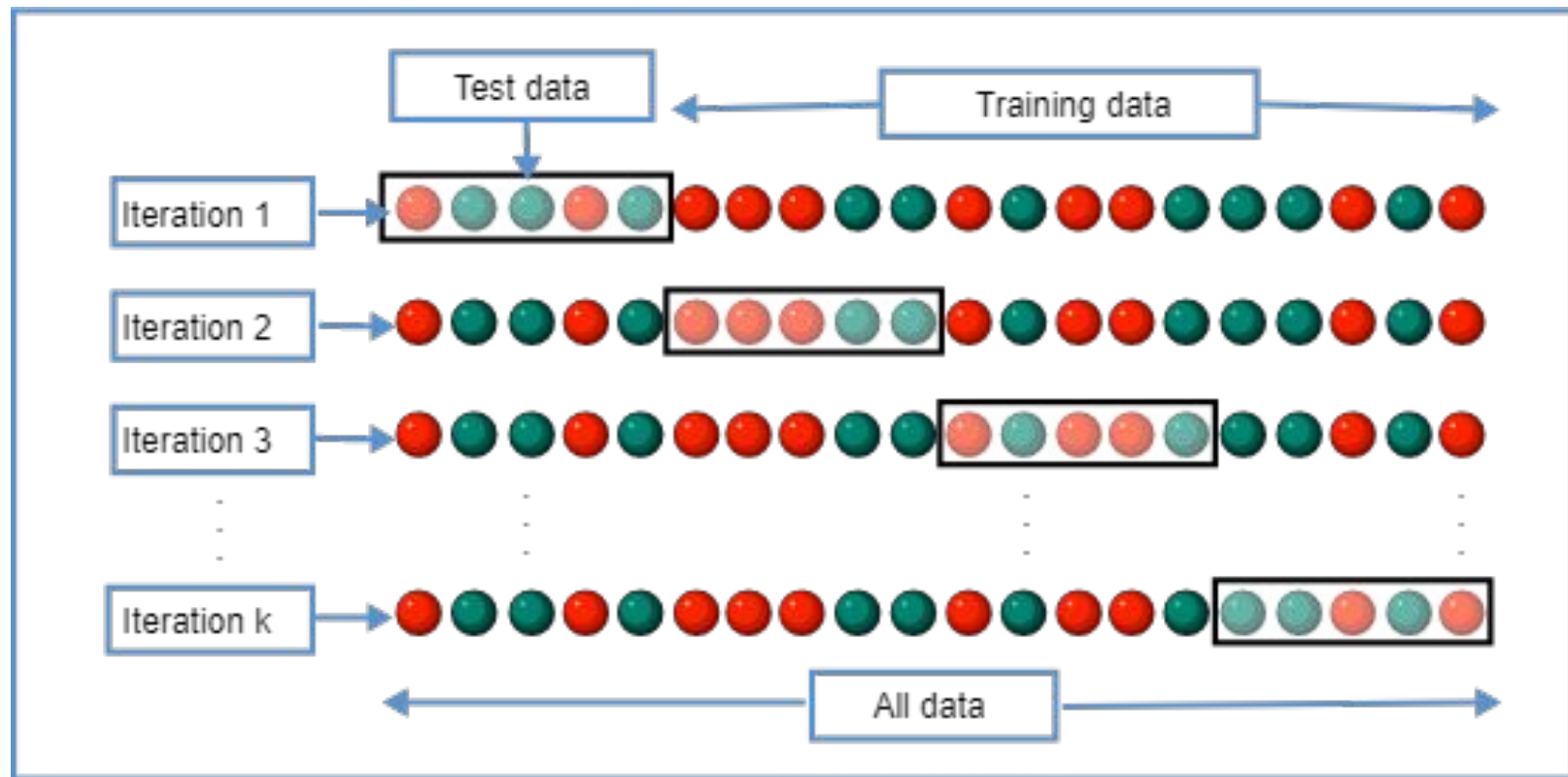
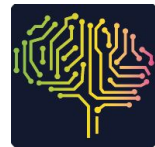
Imagina quantos registros importantes♥ podem ter na base de teste.....



Ou seja, sua
Máquina poderia ter
uma assertividade
maior



...é aí que entra a
Validação Cruzada





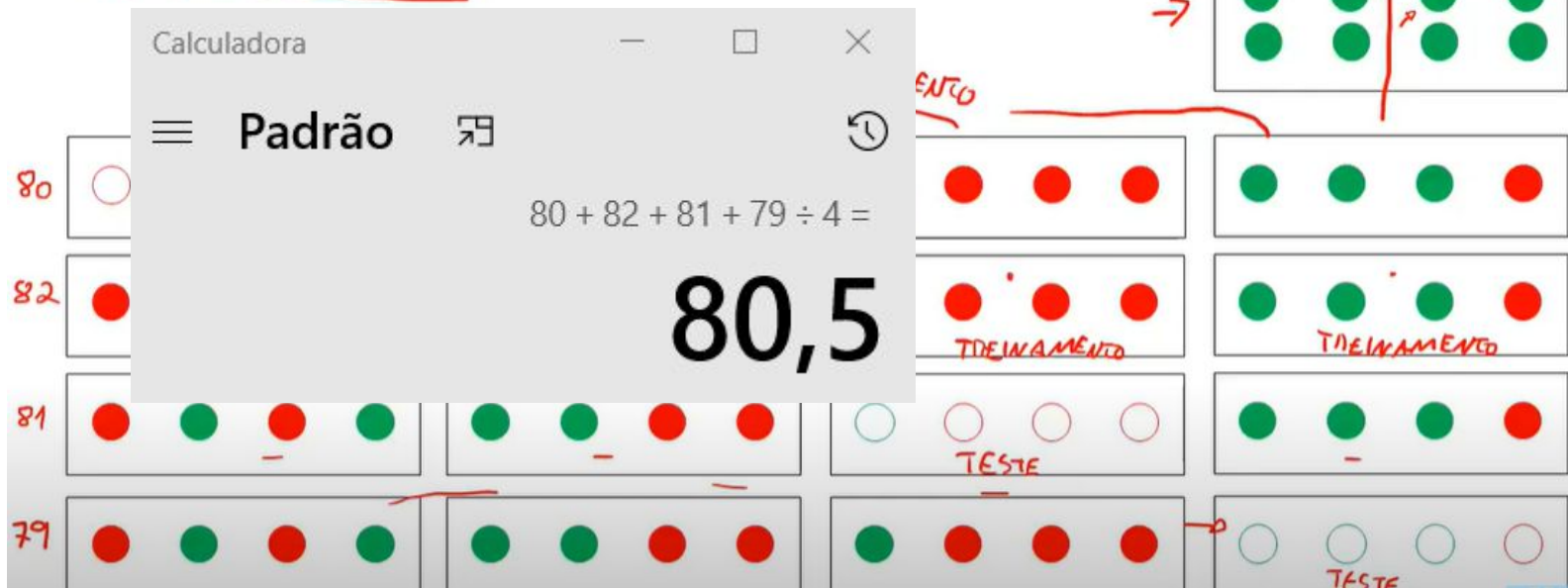
k é a quantidade
de separações (folds)
da base de dados
em treino e teste



Assim, todos os registros
são utilizados na base de
treino e na base de teste



4 K-fold Cross Validation





Simbóra para o





de 74% para 78%

