

Data Science do ZERO

Validação Cruzada

Validação Cruzada

- Também conhecida como Cross Validation.
- Técnica usada para validar a performance de um modelo de **Machine Learning**.
- Esta técnica consiste em separar os dados em partes distintas (folds).
- As partes são usadas para treino e teste do algoritmo. Os dados usados para treino sempre serão **diferentes** dos dados usados para teste.



Validação Cruzada

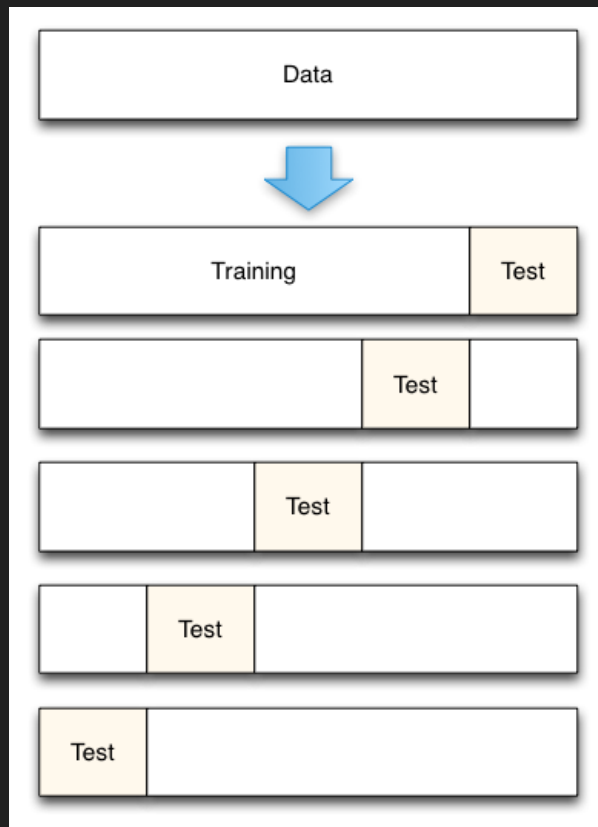
- O Problema de separar dados de treino e teste é ter dados muito iguais em ambos os conjuntos.
- Conjuntos muito semelhantes irão **mascarar** o desempenho real do modelo.
- A Validação Cruzada resolve o problema de **variância** nos dados.
- Ao treinar e testar o modelo com todo o dado disponível, o modelo se torna mais confiável.

Peso	CabeloCurto	Altura	LetraFinalNome	Classe
58	Não	1,62	A	Mulher
61	Não	1,74	A	Mulher
100	Sim	1,85	O	Homen
90	Não	1,90	O	Homen

Peso	CabeloCurto	Altura	LetraFinalNome	Classe
58	Não	1,62	A	?
61	Não	1,70	A	?
101	Sim	1,87	O	?
90	Sim	1,90	O	?

Validação Cruzada

- Validação Cruzada separa dados em partes distintas(folds).
- A quantidade de partes é definida pelo usuário.
- Uma parte é destinada para **teste** enquanto todas as outras são usadas para **treino**.
- Com essa configuração assegura-se que o algoritmo é treinado e testado com todo o dado disponível.
- Resolve o problema de variância nos dados.



Hands on!