



Como garantir que ocorreu aprendizado?

☰ Ciclo	Ciclo 04: As garantias de aprendizado
# Aula	26
🕒 Created	@January 30, 2023 4:59 PM
☑ Done	☑
☑ Ready	☑

Objetivo da Aula:

- ☐ A garantia do aprendizado
- ☐ Joaquim: O DS novato
- ☐ Resumo
- ☐ Próxima Aula

Conteúdo:

▼ 1. A garantia do aprendizado

A única forma real de garantir o aprendizado do algoritmo é comparar as classificações com os rótulos reais, no caso da classificação.

É exatamente essa comparação entre as previsões e os valores reais que ocorre o treinamento dos algoritmos.

▼ 1.1 Exemplo de Aprendizagem

Cor	Combustível	Preço	Número de Portas	Tipo de Veículo	Classificação	Acertos ou Erros
Azul	Gasolina	R\$ 30.000	4	Carro	Carro	1
Vermelho	Gasolina	R\$ 22.500	2	Carro	Moto	0
Verde	Álcool	R\$ 40.000	4	Carro	Carro	1
Preto	Diesel	R\$ 35.000	4	Carro	Carro	1
Prata	Gasolina	R\$ 10.000	0	Moto	Moto	1
Branco	Gasolina	R\$ 12.000	0	Moto	Moto	1
Amarelo	Álcool	R\$ 12.500	0	Moto	Moto	1
Azul	Gasolina	R\$ 11.000	0	Moto	Moto	1
Cinza	Diesel	R\$ 300.000	2	Ônibus	Ônibus	1
Verde	Diesel	R\$ 350.000	2	Ônibus	Moto	0
Azul	Diesel	R\$ 290.000	2	Ônibus	Carro	1
Preto	Diesel	R\$ 320.000	2	Ônibus	Carro	0

Amarelo	Diesel	R\$ 280.000	2	Ônibus	Moto	0
Vermelho	Diesel	R\$ 270.000	2	Ônibus	Ônibus	1

Porém, nós queremos que os algoritmos sejam capazes de fazer classificações corretas para dados que nunca foram vistos. Para isso, é necessário realizar testes para validar o aprendizado. Nesse teste, o algoritmo recebe um novo conjunto de dados, que não foram usados durante o treinamento e precisa ser capaz de classificar corretamente dados nunca vistos.

Essa capacidade é chamada de generalização.

▼ 2. Joaquim: O DS novato

“Joaquim era um jovem Cientista de Dados recém-empregado na empresa de seus sonhos. Ele estava empolgado com a oportunidade de trabalhar com dados e mostrar o que tinha aprendido durante a pós-graduação na FAPONE.

No seu primeiro mês de trabalho, Joaquim recebeu um problema de negócio, coletou e tratou um grande conjunto de dados e treinou um algoritmo de Machine Learning para prever se a cliente pertenceria a classe A ou B no ranking dos clientes, priorizando o atendimento do time de vendas. Se o cliente fosse classificado como A, ele seria priorizado nas ligações.

Com o algoritmo treinado, Joaquim se lembrou que era necessário fazer uma validação do aprendizado com dados que não foram usados no treinamento do algoritmo. Como ele havia usado todos os dados disponíveis para o treinamento, sua decisão foi publicar o algoritmo em Produção e usar os dados nunca vistos para medir a capacidade de generalização do algoritmo.

Quando ele começou a verificar os resultados do algoritmo em produção, ficou horrorizado. O algoritmo estava dando classificações tão ruins que estava causando prejuízos à empresa. Cada previsão errada a empresa deixava de faturar R\$ 4897 reais. Quanto a conta chegou a R\$ 480.000,00, o CFO (Chief Finance Office), entrou em contato com o chefe de venda para questionar o motivo das vendas estarem diminuindo tanto.

O chefe das vendas o (CSO) investigou o processo do seu time e verificou que eles estavam dando preferência nas ligações para pessoas com baixo potencial de conversão em venda, devido as recomendações do algoritmo feito pelo time de Cientistas de Dados.

O chefe de dados da empresa, questionou seu time e pediu que o algoritmo fosse reavaliado. Joaquim logo percebeu que tinha feito algo errado durante o treinamento. Ele então começou a verificar todos os passos que havia feito durante o treinamento para entender o que estava acontecendo.

Depois de horas de análise, ele descobriu que tinha esquecido de avaliar o desempenho do algoritmo antes de colocá-lo em produção. Ele percebeu que precisava ter usado métricas para verificar o desempenho do algoritmo antes de colocá-lo em produção.

Joaquim retirou o algoritmo imediatamente do ambiente de Produção e explicou ao time o ocorrido. Depois dessa experiência, ele passou a usar métricas para verificar o desempenho do algoritmo antes de colocá-lo em produção”

Como avaliar a capacidade de generalização do algoritmo antes de colocá-lo em Produção?

▼ 3. Resumo

1. A validação do aprendizado garante o aprendizado real do algoritmo.
2. O mais importante para um algoritmo de Machine Learning é a sua capacidade de generalização.
3. Não faça como o Joaquim, faça testes de validação antes de publicar um algoritmo em Produção.

▼ 4. Próxima aula

A estratégia do Treino-Teste