

# Tarefa EBAC - MÓDULO 29

Lucas Amaral Taylor

16 de fevereiro de 2026

Para esta atividade, utilizarei como tema uma loja virtual, *e-commerce*. A motivação da escolha do tema vem de que é o tema mais abordado do curso, até o momento, além de ser um exemplo muito ilustrativo para as sete etapas apresentadas acima.

1. **Coleta de Dados.** A coleta de dados em um *e-commerce* ocorre em dois momentos: no cadastro do cliente e na sua atividade no site.

Para a realização de compras *online*, é necessário que o cliente faça um cadastro no *site* responsável pelo *e-commerce*. Nesse cadastro, podemos dividir as informações em dois tipos: informações de perfil do cliente e informações de pagamento. No que diz respeito às primeiras, temos, por exemplo, gênero, nome, endereço e CPF. Já nas informações de pagamento, incluem-se o método de pagamento, o endereço de cobrança e os dados associados ao método selecionado.

Por fim, o comportamento do cliente no site também faz parte da coleta de dados. Em *sites* de *e-commerce* mais robustos, são levadas em consideração não apenas as compras realizadas pelo cliente (incluindo as informações dos produtos comprados, como categoria, quantidade e marca, por exemplo), mas também os produtos visualizados durante a visita ao site, o tempo de permanência na página, o posicionamento do *mouse*, as pesquisas realizadas, entre outros.

2. **Preparação dos Dados.** Após a coleta, realizamos a etapa de preparação dos dados. Inicialmente, efetuamos o tratamento de dados sensíveis, como o CPF, tomando como base legal a Lei Geral de Proteção de Dados Pessoais (LGPD).

Além disso, realizamos o tratamento de dados ausentes e nulos (remoção ou imputação), a exclusão de dados irrelevantes, o tratamento de inconsistências (como erros de digitação e categorias duplicadas), a identificação e o tratamento de *outliers*, a conversão de tipos

de dados e, por fim, a aplicação de técnicas de *One-hot encoding* e *Label encoding*.

3. **Escolha do Modelo.** Durante o nosso curso, modelos de *e-commerce* buscavam traçar o perfil do cliente a fim de entender o comportamento de compra para promover recomendações que podem ser convertidas em compra. Para reduzir o escopo de opções possíveis para o nosso objetivo, vamos restringir o funcionamento do modelo para o seguinte: a promoção de  $X\%$  de desconto no *site*, converteu em aumento de compras? Se sim, qual perfil foi alcançado na promoção? Qual foi seu impacto?

Este problema é um problema de classificação. Estamos interessados em saber qual público adotou, ou seja, podemos classificar o cliente em adotou (1) ou não adotou (0). Isto posto, dado a natureza do problema, podemos utilizar os modelos de classificação aprendidos no curso, como: Naive Bayes, Árvore de decisão ou regressão logística.

4. **Treinamento do Modelo.** Para o treinamento do modelo, como de costume, separamos os dados em duas bases: base de treino e base de teste. Em relação ao treinamento, estamos interessados na variável de adoção da promoção: adotou (1) ou não adotou (0). Essa variável foi utilizada como variável *target* e, portanto, foi omitida do conjunto de variáveis explicativas durante o treinamento do modelo. Além disso, foi realizado o balanceamento das classes, a fim de evitar viés no aprendizado decorrente de possíveis desproporções entre clientes que adotaram e que não adotaram a promoção. Todos os modelos listados acima são supervisionados; assim, para realizar o treinamento, analisamos se, a partir dos dados, o modelo conseguiu classificar corretamente se o cliente adotou ou não a promoção.
5. **Avaliação do Modelo.** Para a avaliação do modelo, verificamos as métricas obtidas, principalmente, na base de teste do modelo<sup>1</sup>. As métricas avaliadas são: precisão, *recall* e *f1-score*, cada uma definida matematicamente abaixo:

- $\text{Precisão} = \frac{\text{Número total de previsões corretas}}{\text{Número total de previsões}}$

- $\text{Recall} = \frac{\text{Verdadeiros positivos}}{\text{Verdadeiros positivos} + \text{Falsos negativos}}$

- $F_1 = \frac{2 \times \text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}}$

6. **Ajuste e Otimização.** Após o treinamento inicial do modelo, realizamos ajustes e etapas de otimização com o objetivo de melhorar seu desempenho. Esses processos baseiam-se na definição e configuração dos hiperparâmetros, como o parâmetro de regularização e o

---

<sup>1</sup>Geralmente, também verificamos as métricas também na base de treino. Contudo, entre as duas, a mais relevante para os nossos objetivos é a base de treino

tipo de penalidade adotado. Além disso, utilizamos *solvers* para tornar mais eficientes alguns processos internos do treinamento, como o *solver newton-cg*, por exemplo. Por fim, destaca-se que uma ferramenta fundamental para a otimização do modelo e para a definição dos hiperparâmetros é o *Grid Search*, que testa diferentes combinações desses parâmetros a fim de identificar aquela que melhor se adequa ao modelo.

7. **Implantação e Monitoramento.** Após o treinamento, o modelo é implementado no cotidiano da empresa, ou seja, deixa a fase de testes e validação e passa a operar em ambiente real. Nessa etapa, recomenda-se a adoção de boas práticas, como a limpeza e a documentação do código, a fim de facilitar futuras manutenções. Além disso, realiza-se a integração do modelo com plataformas de nuvem, como *AWS*, *Google Cloud* ou *Azure*.

Modelos relacionados a vendas não são estáticos, uma vez que o desempenho comercial, especialmente no contexto de vendas *online*, é influenciado por fatores como condições econômicas, entrada e saída de produtos, decisões estratégicas das marcas e mudanças no comportamento do consumidor. Considerando o problema proposto, ainda que uma promoção apresente maior adesão por parte de um público  $X$  em determinado momento, esse padrão pode se alterar ao longo do tempo, como em diferentes períodos do ano. Dessa forma, reconhecendo o caráter dinâmico dos dados, torna-se necessário monitorar continuamente tanto as informações de entrada quanto a acurácia do modelo. Sempre que necessário, devem ser realizadas manutenções ou atualizações para garantir sua adequação e desempenho ao longo do tempo.