

Gabriel Walker Schulze – 02627211200

Profissão: Cientista de Dados v2

Responda as questões a seguir:

1. Diferencie com suas palavras aprendizado supervisionado de aprendizado não supervisionado.
2. Pesquise e exemplifique pelo menos mais 2 exemplos de problemas de negócios em que podemos usar aprendizado supervisionado.
3. Pesquise e exemplifique pelo menos mais 1 exemplo de problemas de negócios em que podemos usar aprendizado NÃO supervisionado.
4. Por que é essencial que na etapa de teste do aprendizado supervisionado a base de teste não contenha os "rótulos" de dados a serem previstos? Justifique com suas palavras.
5. Nas atividades, você tem trabalhado com uma base de dados onde nosso objetivo será de prever Churn dos nossos clientes. Vocês já realizaram a limpeza e tratamento dos dados e já estão realizando a análise exploratória. Qual tipo de aprendizado você acredita ser o ideal para prevermos o Churn da base que temos? Justifique com suas palavras.

Respostas às questões:

1. Diferença entre aprendizado supervisionado e não supervisionado:
No aprendizado supervisionado, utilizamos dados já rotulados (com respostas conhecidas) para treinar o modelo, como classificar e-mails em "spam" ou "não spam". Já no aprendizado não supervisionado, não temos rótulos pré-definidos, e o algoritmo busca padrões ou agrupamentos nos dados, como segmentar clientes por comportamento semelhante sem uma categoria prévia.
2. Exemplos de problemas de negócios com aprendizado supervisionado:
 - a. Previsão de inadimplência: Bancos podem usar dados históricos de clientes (como renda, score de crédito e pagamentos passados) para prever se um novo cliente terá alta probabilidade de não pagar um empréstimo.
 - b. Diagnóstico médico automatizado: Hospitais podem treinar modelos com imagens de exames (como raios-X) já diagnosticados para identificar doenças automaticamente em novos pacientes.
3. Exemplo de problema com aprendizado não supervisionado:

- a. Recomendação de produtos: E-commerces podem analisar padrões de compra (sem rótulos prévios) para agrupar produtos frequentemente comprados juntos e sugerir itens relacionados, como "clientes que compraram café também levaram açúcar".
4. Importância de não ter rótulos na base de teste (aprendizado supervisionado):
A base de teste deve simular situações reais, onde o modelo não conhece a resposta correta. Se os rótulos estivessem disponíveis durante o teste, o modelo poderia "enganar" o sistema, memorizando os dados em vez de aprender padrões generalizáveis. A avaliação justa exige que o modelo faça previsões cegas, comparadas depois com os rótulos verdadeiros para medir sua acurácia.
5. Tipo de aprendizado ideal para prever Churn:
Como o objetivo é prever se um cliente deixará ou não a empresa (Churn), e a base já contém dados históricos com rótulos (clientes que saíram ou não), o ideal é usar aprendizado supervisionado. Técnicas como regressão logística ou árvores de decisão podem identificar padrões nos dados (como tempo de contrato, frequência de uso) que antecedem o Churn, permitindo ações preventivas.