

Tarefa EBAC: Questionário

Machine Learning 2: Tipos de aprendizados

Lucas Amaral Taylor

24 de maio de 2025

1. Diferencie com suas palavras aprendizado supervisionado de aprendizado não supervisionado.

Enquanto isso, o *aprendizado supervisionado* é aquele que busca por padrões com um treinamento prévio. Temos que *aprendizado não supervisionado* é aquele que a partir do *dataset* a máquina busca por padrões em dados não rotulados.

2. Pesquise e exemplifique pelo menos mais 2 exemplos de problemas de negócios em que podemos usar aprendizado supervisionado.

A partir da referência *Supervised Machine Learning: A Vital Modern Business Tool* (s.d.), temos que o aprendizado supervisionado pode ser utilizado para:

(a) **Segurança e identificação de fraudes.** O exemplo fornecido pela referência e, também dado em aula, é a identificação de e-mails maliciosos (*spam*). Atualmente, empresas provedoras de e-mail treinam algoritmos para identificação por meio de estrutura do texto, palavras-chave e presença de *links* maliciosos através de uma *dataset* de e-mails classificados ou não como *spam* atribuem um *score* (NBC News, 2017). A identificação destes e-mails é fundamental para evitar o vazamento de dados sensíveis da empresa .

(b) **Recomendação de produtos com base no comportamento do cliente em um site de compras.** O algoritmo iria por buscar por recomendações personalizadas com grande potencial de compra com base no histórico de visualização e compra de produtos.

3. Pesquise e exemplifique pelo menos mais 1 exemplo de problemas de negócios

em que podemos usar aprendizado NÃO supervisionado.

Utilizando como referência *Unsupervised Machine Learning in Business* (s.d.), temos que podemos utilizar o aprendizado não supervisionado para a identificação de anomalias. O processo de *clustering* possibilita a detecção de *outliers* e diferenças nos dados. Por exemplo, uma empresa financeira pode usar o aprendizado de máquina não supervisionado para alertá-la sobre cobranças estranhas no cartão da empresa, o que indica transações fraudulentas.

- 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.**

Temos que os “rótulos” (*labels*) são retirados na etapa de teste, pois nesta etapa aplicamos uma prova no algoritmo. Na etapa seguinte, chamada avaliação, “revelamos” estes resultados e temos a acurácia do modelo e, a partir dela, podemos inferir melhorias e a taxa de erro do modelo.

Em suma, retiramos os rótulos para testar o quanto bem o modelo está funcionando.

- 5. Nas atividades, você tem trabalhado com uma base de dados onde nosso objetivo será de predizer *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.**

O tipo de aprendizado ideal para prevermos o *Churn* da base que temos é o *não supervisionado*, por dois principais motivos:

- (a) O *aprendizado não supervisionado* busca identificar padrões com base no *dataset* que temos, como o objetivo é prever o *Churn*, i.e., identificar o possível *Churn* é necessário realizar o processo de *clustering* para encontrar padrões.
- (b) O *aprendizado supervisionado* precede uma resposta para a etapa de avaliação. Como queremos identificar um potencial *Churn*, i.e., *predizer*, não temos a resposta necessária para a aplicação do método.

Referências

- NBC News (2017). *How Email Spam Filters Work Based On Algorithms — Mach — NBC News*. https://www.youtube.com/watch?v=_i0xylrN4Io. Acesso em 24 de maio de 2025.
- Supervised Machine Learning: A Vital Modern Business Tool* (s.d.). <https://mailchimp.com/resources/supervised-machine-learning/>. Acesso em 24 de maio de 2025.
- Unsupervised Machine Learning in Business* (s.d.). <https://mailchimp.com/resources/unsupervised-machine-learning/>. Acesso em 24 de maio de 2025.