

Atividade 01

1. Explique, com suas palavras, o que é machine learning?

Pense no Machine Learning como um jeito de **ensinar computadores a aprenderem sozinhos**, sem que a gente precise programar cada coisinha que eles devem fazer. É como se a gente desse muitos exemplos para o computador, e ele, a partir desses exemplos, aprende a identificar padrões e a tomar decisões.

Imagine que você quer ensinar um computador a reconhecer a diferença entre um gato e um cachorro em fotos. Em vez de escrever um código que diga: "se tiver orelha pontuda e bigodes, é gato", você mostra ao computador **milhares de fotos de gatos e milhares de fotos de cachorros**.

O computador, então, vai "estudar" essas fotos, procurando por características que são mais comuns em gatos (como orelhas, formato do rosto, etc.) e outras que são mais comuns em cachorros. Depois de "aprender" com esses exemplos, se você mostrar uma foto nova, ele consegue dizer se é um gato ou um cachorro, mesmo que nunca tenha visto aquela foto específica antes.

2. Explique o conceito de conjunto de treinamento, conjunto de validação e conjunto de teste em machine learning.

Para o computador aprender de verdade e não só "decorar", a gente divide os dados em três grupos:

Conjunto de Treinamento: É a **maior parte dos dados**, que o computador usa para **aprender**. É como as aulas e os livros.

Conjunto de Validação: É uma parte menor, usada para **ajustar o modelo** durante o aprendizado e evitar que ele "decore" demais (o que chamamos de *overfitting*). É como os exercícios ou simulados.

Conjunto de Teste: É uma parte dos dados que o computador **nunca viu**. Usamos ela **apenas uma vez, no final**, para saber o quão bem o modelo realmente aprendeu e se comportará com dados novos. É como a prova final.

3. Explique como você lidaria com dados ausentes em um conjunto de dados de treinamento.

É comum que alguns dados venham com "buracos" (dados ausentes). Precisamos lidar com eles, pois os modelos de Machine Learning não sabem o que fazer com informações faltando. As formas mais comuns são:

- **Remover:** Se faltam poucos dados, podemos jogar fora as linhas ou colunas que têm esses buracos. Mas cuidado para não perder informações importantes!
- **Preencher (Imputar):** É a forma mais comum. A gente "adivinha" e preenche os buracos com algum valor, como:

- A **média** ou **mediana** (para números).
- A **moda** (o valor que mais aparece, para categorias).
- Um **valor constante** (tipo zero ou "desconhecido").

A escolha da melhor forma de lidar com os dados ausentes depende do tipo de dado e de quantos dados estão faltando. O importante é sempre verificar e tratar esses "buracos" antes de treinar seu modelo.

4. O que é uma matriz de confusão e como ela é usada para avaliar o desempenho de um modelo preditivo?

A **Matriz de Confusão** é uma tabela que nos ajuda a ver em detalhes onde o nosso modelo de Machine Learning acerta e erra, especialmente quando ele precisa classificar algo em "sim" ou "não". Ela mostra:

- **Verdadeiro Positivo (VP):** O modelo previu "sim" e a realidade era "sim" (acerto).
- **Verdadeiro Negativo (VN):** O modelo previu "não" e a realidade era "não" (acerto).
- **Falso Positivo (FP):** O modelo previu "sim", mas a realidade era "não" (um erro, tipo um alarme falso).
- **Falso Negativo (FN):** O modelo previu "não", mas a realidade era "sim" (um erro mais perigoso, tipo não detectar uma doença).

Com base nessa matriz, podemos calcular métricas importantes como:

- **Acurácia:** Porcentagem total de acertos.
- **Precisão:** Dos que o modelo disse "sim", quantos eram "sim" de verdade (importante quando um FP é ruim).
- **Recall:** Dos que eram "sim" de verdade, quantos o modelo conseguiu pegar (importante quando um FN é ruim).

Essas métricas nos ajudam a entender se o modelo é bom para o nosso objetivo, dependendo de qual tipo de erro é mais crítico para a gente.

5. Em quais áreas (tais como construção civil, agricultura, saúde, manufatura, entre outras) você acha mais interessante aplicar algoritmos de machine learning?

Machine Learning é útil em tudo, mas se destaca muito em áreas como:

- **Saúde:** Para diagnósticos mais rápidos, descoberta de medicamentos e medicina personalizada.
- **Agricultura:** Para otimizar plantações (água, adubo), prever colheitas e doenças.
- **Manufatura:** Para prever falhas de máquinas, controlar a qualidade e otimizar a produção.

Essas áreas têm dados complexos e problemas onde o ML pode trazer grandes avanços e benefícios.