



## **BOOTCAMP ATLÂNTICO AVANTI**

**ALUNO:** Vinícius Arantes Emrich Leão

**EMAIL:** viniciusemrich@outlook.com

**CURSO:** Machine Learning

### **Atividade 1 - Conceitos básicos de Machine Learning**

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

Machine Learning, é um campo da inteligência artificial que permite que máquinas aprendam a identificar padrões, fazer previsões e realizar tarefas de forma autônoma a partir da alimentação e treinamento inicial com grandes bases de dados. Ou seja, é um processo de desenvolvimento de algoritmos e técnicas que permitem as máquinas aprenderem algo com base nas experiências anteriores obtidas com os dados e irem melhorando seu desempenho nas tarefas sem serem explicitamente programados para tal.

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

O conjunto de treinamento, o conjunto de validação e o conjunto de teste são etapas fundamentais do processo de machine learning, usados para treinar, validar e testar modelos de aprendizado de máquina.

##### **a) Conjunto de Treinamento:**

É o conjunto de dados usado para treinar o modelo de machine learning. Consiste em exemplos de classificações e características dessas que o modelo usa para aprender os padrões nos dados.

##### **b) Conjunto de Validação:**

É usado para ajustar os parâmetros do modelo. Este conjunto é necessário para evitar que o modelo se ajuste muito bem aos dados de treinamento, mas não generaliza bem para dados novos, o que é chamado de overfitting.

### **c) Conjunto de Teste:**

É a fase de avaliar a performance do modelo avaliado usando o conjunto de teste. Este conjunto é totalmente separado do conjunto de treinamento e de validação e contém dados que o modelo nunca viu antes. Avaliar o modelo com o conjunto de teste fornece uma estimativa imparcial de sua performance em dados reais e ajuda a determinar o quão bem o modelo generaliza para novos dados.

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

Entendo que lidar com dados ausentes em um conjunto de dados de treinamento para machine learning é muito importante para garantir a precisão dos resultados. Nesse contexto, simplificarmente, eu seguiria alguns passos para fazer esse tratamento de dados:

Um passo primeiro seria fazer uma identificação de dados ausentes, buscando entender o porque disso, como falhas na coleta, erros de entrada. A partir disso, poderá serem excluídos os dados ausentes, preenchidos com um valor padrão ou pode-se usar estatística para estimar valores a partir dos dados já disponíveis.

Após tratar os dados ausentes, é importante validar o tratamento, observando se a abordagem adotada não introduziu nenhum viés nos dados ou distorceu os resultados do modelo.

E por fim é preciso manter um monitoramento contínuo durante a produção e uso do modelo, para detectar e lidar com novos dados ausentes que possam surgir.

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

Uma matriz de confusão é uma tabela que mostra o desempenho de um modelo de machine learning na tarefa de classificação. Esta tabela compara as previsões feitas pelo modelo com os valores reais dos dados. Ou seja, é uma ferramenta essencial para avaliar a performance de modelos de classificação em machine learning, por fornecer uma visão completa do

desempenho do modelo e avaliar se as predições estão coerentes, possibilitando a identificação de pontos de aprimoramento.

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

Entendo que algoritmos de machine learning têm potencial para serem aplicados em qualquer área de conhecimento, devido à sua capacidade de extrair informações de dados e automatizar processos de tomada de decisão. Mas existem alguns campos nos quais sua utilização se destacaria além dos já citados no enunciado, seriam esses por exemplo:

- a) **Finanças:** No setor financeiro, os algoritmos podem ser aplicados para detecção de fraudes, previsão de riscos, análise de crédito, otimização de investimentos, análise de mercado, entre outras aplicações voltadas para tomada de decisões financeiras mais precisas e eficientes.
- b) **Varejo:** No varejo, os algoritmos de machine learning podem ser utilizados para prever demanda, personalizar recomendações de produtos, otimizar preços, realizar análises de mercado, entre outras aplicações voltadas para melhorar a experiência do cliente e aumentar as vendas.
- c) **Segurança:** Na área da segurança, algoritmos de machine learning são utilizados para detectar anomalias e intrusões em sistemas de computadores e redes, analisar imagens e vídeos de vigilância para identificar comportamentos suspeitos, e gerenciar riscos por meio da análise de dados. Essas aplicações ajudam a fortalecer a proteção contra ameaças e a promover ambientes mais seguros.