

KESAVA MENEZES

Atividade Diagnóstica - Machine Learning

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

Machine Learning (ML) é uma área da inteligência artificial que desenvolve sistemas capazes de aprender automaticamente com a experiência e identificar padrões em dados. **Existem duas abordagens: aprendizado supervisionado**, que usa dados classificados, e **aprendizado não supervisionado**, que aprende por conta própria. O **ML** é aplicado em áreas como detecção de problemas e recomendações personalizadas, ajudando empresas a reter clientes e se tornando comum em setores como saúde e finanças, além de aparecer em sugestões de produtos em plataformas de streaming e e-commerce.

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

Em Machine Learning, os **conjuntos de treinamento, validação e teste** são essenciais para **construir e avaliar modelos**. O **conjunto de treinamento** é usado para ensinar o modelo a reconhecer padrões entre as variáveis independentes e a variável dependente. Depois de treinar, o modelo é ajustado usando o **conjunto de validação**, que ajuda a comparar diferentes abordagens e verificar se o modelo está se saindo bem sem tocar nos dados que serão usados para teste final. O **conjunto de teste**, por sua vez, é reservado exclusivamente para avaliar o desempenho do modelo em dados que nunca foram vistos antes, garantindo que ele funcione bem em situações do mundo real. Essa divisão é importante para evitar que o modelo se ajuste demais aos dados de treinamento, o que poderia prejudicar sua capacidade de fazer previsões precisas no futuro.

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

Para lidar com **dados ausentes** em um conjunto de dados de treinamento, uma boa opção é a **imputação**, que consiste em substituir os valores faltantes por estimativas. Podendo usar a **média**, a **mediana** ou a **moda** dos dados que já existem, dependendo do tipo de informação. Outra alternativa é usar métodos mais avançados, como o **K-Nearest Neighbors**, que estima os valores com base em dados semelhantes. Também é possível treinar um modelo que

adivinha os valores ausentes usando outras características. Fazer a imputação corretamente é importante para que o modelo aprenda bem e faça previsões precisas no futuro.

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 ferramenta que avalia o desempenho de um modelo de classificação, comparando as previsões do modelo com os resultados reais. Ela organiza os dados em quatro categorias: verdadeiros positivos, falsos positivos, falsos negativos e verdadeiros negativos. Com essas informações, podemos calcular métricas como precisão, revocação, F1-Score e acurácia, que ajudam a entender como o modelo está se saindo e a identificar erros.

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

A **aplicação de Machine Learning na economia** oferece oportunidades para melhorar a tomada de decisões e prever tendências. Os **algoritmos** podem prever variáveis macroeconômicas, como PIB e inflação, auxiliar na avaliação de risco de crédito e detecção de fraudes no setor financeiro, além de otimizar a segmentação de mercado e personalização de ofertas na economia do consumidor. Eles também permitem analisar sentimentos de notícias e redes sociais para entender impactos no mercado, otimizar gastos públicos e modelar interações em sistemas econômicos. Com a capacidade de identificar padrões em grandes volumes de dados financeiros, o Machine Learning transforma a análise de dados, promovendo eficiência e eficácia em diversos setores da economia.