

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

Machine learning, ou aprendizado de máquina, é uma área da inteligência artificial que foca em criar algoritmos que permitem que as máquinas aprendam a partir dos dados. Em vez de programar a máquina explicitamente para realizar uma tarefa, você fornece a ela muitos exemplos (dados) e o algoritmo "aprende" a identificar padrões e tomar decisões por conta própria. É tipo ensinar uma criança a andar de bicicleta: você não explica cada movimento, mas dá a bicicleta e deixa ela praticar até aprender.

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

Conjunto de treinamento: É o conjunto de dados que usamos para treinar nosso modelo. O modelo aprende com esses dados, ajustando seus parâmetros para capturar os padrões presentes.

Conjunto de validação: É um conjunto de dados separado do conjunto de treinamento. Usamos para ajustar o modelo, selecionar os melhores parâmetros e evitar overfitting (quando o modelo se adapta muito bem aos dados de treinamento, mas não generaliza bem para novos dados).

Conjunto de teste: Este é usado para avaliar o desempenho final do modelo. Depois de treinado e ajustado, testamos o modelo com esse conjunto de dados, que ele nunca viu antes, para ver como ele se comporta no "mundo real".

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

Remoção: Se a quantidade de dados ausentes for pequena, você pode simplesmente remover essas entradas.

Imputação: Você pode preencher os dados ausentes com valores como a média, mediana ou moda da coluna. Também pode usar modelos mais complexos, como K-Nearest Neighbors, para imputar os valores.

Modelos específicos: Alguns algoritmos de machine learning conseguem lidar com dados ausentes naturalmente, como certas árvores de decisão.

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 nos ajuda a visualizar o desempenho de um modelo de classificação. Ela mostra as previsões corretas e incorretas feitas pelo modelo, divididas em categorias:

- True Positives (TP): Previsões corretas para a classe positiva.

- True Negatives (TN): Previsões corretas para a classe negativa.
- False Positives (FP): Previsões incorretas onde o modelo previu positivo, mas era negativo.
- False Negatives (FN): Previsões incorretas onde o modelo previu negativo, mas era positivo.

Com essa matriz, podemos calcular métricas como acurácia, precisão, recall e F1-score, que nos ajudam a entender quão bem o modelo está performando

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

Saúde: Machine learning é super interessante na área da saúde porque pode ajudar a diagnosticar doenças mais rapidamente e com mais precisão, prever surtos de epidemias, personalizar tratamentos e até mesmo descobrir novos medicamentos. Por exemplo, algoritmos podem analisar imagens de raios-X para detectar câncer de mama em estágios iniciais, o que pode salvar muitas vidas.

Agricultura: Na agricultura, machine learning pode revolucionar a forma como cultivamos alimentos. Pode prever rendimentos de colheitas, detectar doenças nas plantas em estágios iniciais e otimizar o uso de água e fertilizantes. Por exemplo, drones equipados com câmeras e sensores podem monitorar grandes áreas de plantações e identificar problemas que o olho humano pode não detectar tão rapidamente.

Manufatura: Na manufatura, machine learning é usado para otimizar processos de produção, prever falhas em máquinas e melhorar a qualidade dos produtos. Algoritmos analisam dados de sensores para prever problemas e programar manutenção preventiva, economizando custos e reduzindo o tempo de inatividade. Além disso, ajudam no controle de qualidade, identificando defeitos durante a fabricação para garantir produtos de alta qualidade.