



Bootcamp: Machine Learning

Aluna: Débora Nunes Ferreira

1. Explique, com suas palavras, o que é Machine Learning?

Machine Learning, ou aprendizado de máquina, é uma área da inteligência artificial que envolve ensinar computadores a aprender e fazer previsões ou tomar decisões com base em dados.

2. Explique o conceito de conjunto de treinamento, conjunto de validação e conjunto de teste em Machine Learning.

Em Machine Learning, utilizamos três conjuntos principais de dados para desenvolver e avaliar modelos: o conjunto de treinamento é usado para ensinar o modelo, permitindo que ele aprenda com os dados fornecidos. Durante o treinamento, o modelo ajusta seus parâmetros com base nos exemplos do conjunto de treinamento. Após o treinamento, usamos o conjunto de validação para ajustar os hiperparâmetros do modelo e avaliar seu desempenho. Isso ajuda a verificar se o modelo está generalizando bem, ou seja, se está aprendendo padrões úteis em vez de apenas memorizar os dados de treinamento. O conjunto de teste é reservado para avaliar o desempenho final do modelo. Contendo dados que o modelo não viu durante o treinamento, o conjunto de teste oferece uma avaliação imparcial da capacidade do modelo de fazer previsões precisas em novos dados.

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, primeiro analisaria a distribuição e impacto desses dados. Em seguida, escolheria uma estratégia de imputação adequada, como preenchimento com média, moda ou técnicas avançadas de imputação baseadas em Machine Learning, garantindo a integridade dos dados para o treinamento do modelo.

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 compara as previsões feitas por um modelo com os resultados reais. Ela é usada para calcular métricas de desempenho, como precisão e recall, fornecendo uma visão clara de como o modelo está acertando e errando suas previsões.



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

Todas essas áreas são interessantes para aplicar algoritmos de Machine Learning, mas a área que mais tenho interesse em estudar e aprofundar técnicas de Machine Learning é a agricultura, devido ao seu grande impacto em nosso país. Recentemente, fiz um projeto na faculdade no qual utilizamos o algoritmo Random Forest para prever o tipo de cultura com base nas características do solo. O modelo alcançou uma acurácia de 99,9%. Esse resultado destaca o potencial do Machine Learning para melhorar a eficiência agrícola.