

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

Machine Learning é uma área da Inteligência Artificial através da qual os computadores são capazes de processar dados e obter aprendizados através de modelos e algoritmos.

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

O conjunto de treinamento são os dados selecionados para que o modelo seja treinado

O conjunto de validação é um conjunto de dados usado para avaliar o desempenho do modelo durante do treinamento

O conjunto de teste são dados selecionados para testar se o modelo foi corretamente treinado e validado

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

É necessário identificar os dados ausentes e fazer a escolha de retirar completamente as tuplas ou ainda, preencher os dados ausentes com dados neutros. Essa decisão dependerá do contexto.

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

É uma tabela específica que tem como intuito permitir a visualização da performance do algoritmo de classificação. A matriz compara os valores previstos com os valores reais para entender a porcentagem de acurácia e previsão do modelo.

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

Todas as áreas que geram dados são passíveis de melhorias através do machine learning. Eu me interessei muito pela aplicação de machine learning no contexto de negócios em geral (vendas, melhorias de processo produtivo, marketing, etc) e aplicações de machine learning na área de saúde (detecção de doenças através de imagens, predição de doenças através de modelos genéticos, etc).