

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

É um subconjunto de estudo da Inteligência Artificial, focado em explorar e construir algoritmos capazes de aprender/identificar padrões de dados, dados esses que “alimentam” o escopo de treinamento do algoritmo, possibilitando até mesmo fazerem previsões em cima de dados mais recentes. Antes do treinamento se tem um algoritmo e após o treinamento se tem um modelo de ML.

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 são usados especificamente para treinar o modelo, com exemplos ou amostras. O modelo usa esse conjunto para entender os padrões e relações entre os dados.
- **Conjunto de validação:** É o conjunto de dados que são utilizados para avaliar o aprendizado do modelo e refinar os seus parâmetros. A partir desse conjunto conseguimos dizer o quão bem o modelo está aprendendo e adaptando, dando margem para otimizações antes de ser colocado em teste.
- **Conjunto de teste:** Esse conjunto diz respeito aos dados que nunca foram entregues para o modelo, para avaliação imparcial final do modelo. Normalmente com dados que espelham exemplos do mundo real.

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

Há algumas estratégias para se lidar com dados ausentes, então dependeria da porcentagem de dados que está faltando. Pode-se limpar os dados que estão ausentes, mas isso só é interessante se forem menos da metade de dados disponíveis. A partir disso pode-se ou preenchê-los com dados de tendência central, como a média ou a moda, ou preencher com dados mais avançados como com os valores médios dos vizinhos mais próximos.

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

É uma tabela de visualização do desempenho de um algoritmo de classificação, sendo formada de duas colunas de condições reais (positiva ou negativa) e duas linhas de condições previstas (positiva ou negativa) para comparar se o modelo previu corretamente o que deveria ser positivo e o que deveria ser negativo. Ela avalia melhor o modelo preditivo, pois quando se tem um conjunto de dados muito desequilibrado, a precisão pode produzir resultados enganosos. Ex: um modelo classificador recebe 95 carros e 5 tratores e classifica todos como carros, então sua precisão aparente é de 95%, mas poderia dizer que ele acertou 100% dos carros e 0% dos tratores, o que pode indicar que o modelo está prevendo nada, apenas chutando sempre a mesma coisa, sem ter realmente tomado uma decisão informada.

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

Todas que tenham matérias(s) que pode(m) ser quantificada(s). Tendo em vista que ML pode ser aplicada para diversas finalidades, tanto comparativas quanto preditivas, certos modelos podem se adequar para cada situação que se deseja prever ou comparar dados da área para se tomar decisões embasadas no que se tem documentado quantificado. Por exemplo, a área comercial pode se beneficiar ao comparar compras e vendas para saber o período de estocagem ou de mais venda de certo material/produto.