



Aluno: Thiago Guilherme Bezerra de Aguiar

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

R: Aprendizado de máquina é um conceito muito amplo, podendo ir desde ensinar o computador a reconhecer padrões, fazer previsões, até fazer ele aprender sem supervisão, como realizar certas tarefas. Fazemos isso por meio de datasets que são usados como “alimento” para os programas, é deles que o modelo vai aprender a fazer previsões ou classificações de novos dados. M.L. ainda pode ser aplicado para as mais diversas coisas, como treinar um algoritmo para jogar algum jogo de forma automatizada.

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

R: Conjunto de treino é usado para treinar o modelo, é de onde ele vai aprender as informações que precisa para fazer o que foi designado a ele. Conjunto de validação é usado para comparar diferentes modelos e hiperparâmetros. Por fim, conjunto de teste é uma parte do dataset que é composta por dados não vistos durante o treino e é usada para validar o modelo, é dessa parte que vamos tirar as métricas de avaliação do modelo.

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

R: Depende de qual tipo de dados estão faltando e de onde estão faltando, por exemplo, se for peso, podemos fazer um preenchimento usando média. Já se for um outlier, podemos simplesmente deletar toda a linha com dados faltantes. Existem formas diversas de se tratar dados faltantes que vão ser melhor ou pior dependendo de cada caso e dataset.

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

R: Matriz de confusão é usada principalmente em classificações binárias, ela serve para ter uma noção do desempenho do classificador, sendo composta por Verdadeiros Positivos (VP), Falsos Positivos (FP), Verdadeiros Negativos

(VN) e Falsos Negativos (FN). Onde VP são os que foram corretamente classificados na classe alvo e VN os corretamente classificados na outra classe. Em problemas não binários os Negativos se referem a todas as classes que não sejam a alvo.

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

R: Acredito que em todas as áreas o Machine Learning pode ajudar de alguma forma, principalmente com os conhecimentos que temos sobre o assunto constantemente evoluindo. A área que mais estou vendo sobre, e que mais vejo artigos relacionados é a área da saúde, por ser ampla e ter espaço para uso de M.L. em diversas etapas. Seja no uso de NLP, análise de dados para predição de Sepsis, ou análise de imagens para identificação de diversas doenças.