



Aluna: Carol Lima Fonseca

Bootcamp: Machine Learning

Email: carolfonseca@alu.ufc.br

Turma: 20

### **Atividade 01**

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

Digamos que você precisa tomar decisões de maneira automatizada e eficiente, tornando uma coleção de dados em bons insights, assim é o Machine Learning, uma área da Inteligência Artificial que é treinada para aquela determinada situação (em diferentes áreas) e sua previsibilidade é melhorada a cada aprendizado. Podemos pensar em estações meteorológicas, possuem um extenso dataset, no entanto, o ML que será capaz de deduzir de forma eficiente próximos eventos de climáticos com base nos dados.

2. Explique o conceito de conjunto de treinamento, conjunto de validação e conjunto de teste em machine learning.
  - Treinamento: Treinamento do modelo, tempo de ajustar parâmetros e reconhecimento de padrões das variáveis.
  - Validação: Uma avaliação durante o treinamento com dados diferentes dos usados no treinamento para verificar se o modelo “quebra” com novas variáveis.
  - Teste: Fora do período de treinamento, é feita uma avaliação final do modelo para saber se ele realmente funciona como o esperado.

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

Primeiramente, podemos identificar esses dados nulos a partir de funções da biblioteca Pandas. Em seguida analisamos o “peso/proporção” e tipo dos dados para saber se seria viável remoção de maneira que não afete o treinamento do ML ou então preenchimento daquele dado nulo de acordo com alguma estratégia, como a mediana dos valores.

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

É uma ferramenta que nos auxilia a verificar o desempenho do modelo usando dados reais representado visualmente por uma tabela, como abaixo na figura. Dividido em:

- Verdadeiro positivo: Teste de covid positivo para um caso positivo
- Verdadeiro negativo: Teste de covid negativo para um caso negativo
- Falso positivo: Teste de covid positivo para um caso negativo
- Falso negativo: teste de covid negativo para um caso positivo

## CLASSIFICAÇÃO DO MODELO

REAL	VP 70	FN 10
	FP 30	VN 50

acertos

erros

VP - Verdadeiros Positivos

VN - Verdadeiros Negativos

FP - Falsos Positivos

FN - Falsos Negativos

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

Tenho interesse em soluções na área ambiental, que de certa forma atinge agricultores e políticas públicas buscadas por prefeituras. Estou buscando conhecimento para trabalhar em modelos de previsão da temperatura do Ceará, com foco em áreas de risco em desertificação.

### Referências

ALURA, Conheça a formação Machine Learning na Prática: fundamentos e aplicações.

Disponível em: [Fundamentos do Machine Learning](#)

ABNER, Conjuntos de treino, validação e teste em Machine Learning. Disponível em:

[Conjuntos em Machine Learning](#)

Jacob Murel Ph.D., O que é uma matriz de confusão?. Disponível em:

<https://www.ibm.com/br-pt/think/topics/confusion-matrix>