

## **CURSO: <BÁSICO EM MACHINE LEARNING>**

- **Atividade 01 - Giovanna Caetano**
- Tipo: Diagnóstica.
- Tema: Conceitos básicos de machine learning.
- Conteúdo: Machine learning.
- Participantes: Individual.
- Avaliação do aluno.
  - Objetivo: Avaliar o conhecimento do aluno sobre conceitos básicos de machine learning.
  - Nota: Sem valor numérico (apenas um guia para o processo educacional).
  - Critérios avaliados: Respostas com coerência, coesão e com exemplos.
- Informações complementares: A atividade é composta por 5 questões dissertativas. Os resultados são utilizados para divisão de grupos da atividade 04;
- **AO CONCLUIR A ATIVIDADE: ENVIAR APENAS O LINK DO REPOSITÓRIO GITHUB (ESPECIFICAR A BRANCH) PÚBLICO.**

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

Machine learning é uma tecnologia que permite que computadores aprendam a fazer coisas sem serem programados diretamente para isso. Por exemplo, em vez de dizer exatamente como reconhecer um gato, mostramos ao computador muitas fotos de gatos e de outros animais. O computador então aprende por conta própria a identificar os gatos, melhorando com o tempo à medida que vê mais exemplos. Isso é útil para várias tarefas, como

recomendar filmes, diagnosticar doenças e até dirigir carros sozinhos.

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

Em machine learning, dividimos os dados em três conjuntos principais: treinamento, validação e teste. Dividimos os dados dessa forma para garantir que o modelo aprenda bem e seja capaz de generalizar para novos dados, mantendo uma boa performance.

**Conjunto de Treinamento:** Este é o maior grupo de dados e é usado para ensinar o modelo. O modelo aprende padrões e características a partir desses dados.

**Conjunto de Validação:** Esse grupo de dados é usado para ajustar os hiperparâmetros do modelo e evitar o overfitting, que é quando o modelo fica bom demais nos dados de treinamento, mas ruim em novos dados. Ele ajuda a verificar como o modelo está se saindo e fazer ajustes necessários.

**Conjunto de Teste:** Este conjunto de dados é usado apenas no final do processo para avaliar o desempenho real do modelo em novos dados que ele nunca viu antes. Ele dá uma ideia de como o modelo vai se comportar no mundo real.

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, aqui estão algumas estratégias comuns, como remover Linhas ou Colunas com dados ausentes, e se a perda de informações não for significativa, podemos simplesmente removê-las. Substituir os valores ausentes por outros valores. Adicionar uma nova coluna que indica se o valor estava ausente ou presente, mantendo as informações sobre a ausência dos dados. Escolher a melhor estratégia depende do contexto dos dados e do impacto dos valores ausentes no 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 avalia o desempenho de modelos de classificação comparando previsões com resultados reais em quatro categorias: Verdadeiros Positivos (VP), Falsos Positivos (FP), Verdadeiros Negativos (VN) e Falsos Negativos (FN). Ela ajuda a calcular métricas importantes como acurácia e precisão, que indicam a eficácia do modelo em diferentes aspectos.

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

Saúde: Diagnóstico de doenças, personalização de tratamentos, análise de imagens médicas e previsão de surtos de doenças.