

Atividade 01 (ATIV-01)

CURSO: BÁSICO EM MACHINE LEARNING

Tema: Conceitos básicos de machine learning

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

Machine Learning é um subcampo da Inteligência Artificial que se concentra no desenvolvimento de algoritmos e modelos estatísticos que permitem a sistemas computacionais aprender padrões, fazer previsões ou tomar decisões diretamente a partir de dados, sem serem explicitamente programados para cada tarefa. O objetivo é generalizar a partir de exemplos.

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

O Conjunto de Treinamento é utilizado para ajustar os parâmetros do modelo. O Conjunto de Validação é empregado para otimizar hiperparâmetros e selecionar o modelo com melhor desempenho generalizado. O Conjunto de Teste fornece uma avaliação final e imparcial da performance do modelo em dados completamente não vistos.

Imagine que você está estudando para uma prova. O conjunto de **treinamento** são os exercícios que você usa para aprender a matéria. O de **validação** são simulados para ajustar seu estudo. O de **teste** é a prova final, que mostra o quão bem você realmente aprendeu.

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

Estratégias comuns incluem Imputação (substituição por medidas estatísticas como média/mediana, ou métodos mais sofisticados como KNN ou regressão), Deleção (remoção de instâncias ou variáveis com dados faltantes, se apropriado) ou utilização de modelos intrinsecamente robustos a dados ausentes. A escolha depende do padrão e extensão da ausência e da sugestão da literatura para trabalhos aplicados na área do conjunto de dados.

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

É uma tabela que resume o desempenho de um algoritmo de classificação, comparando as classes preditas com as classes reais. Ela detalha os resultados em Verdadeiros Positivos (TP), Verdadeiros Negativos (TN), Falsos Positivos (FP) e Falsos Negativos (FN), sendo fundamental para calcular métricas como Acurácia, Precisão, Recall e F1-score.

Em termos mais simples, é uma tabela que mostra quantos acertos e erros um modelo fez para cada tipo de resultado possível. Ela ajuda a entender onde o modelo acertou (previsões corretas) e onde ele se confundiu (previsões incorretas entre categorias).

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

No momento minhas aplicações de Machine Learning estão mais voltadas para Bioinformática e Biometria por causa dos cursos que tenho feito. Adicionalmente, tenho feito aplicações também relacionadas a Finanças Quantitativas (modelo de previsão e robôs).