

MACHINE LEARNING
Atividade 01
Paloma Pinheiro de Lima

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

Machine Learning é um subconjunto de Inteligência Artificial que permite que os computadores aprendam a partir de exemplos e dados, por experiência.

Pode ser dividido em:

Aprendizado Supervisionado (classificação): tipo mais comum de Machine Learning, envolvendo a alimentação de um algoritmo com dados rotulados. Seu objetivo é o de treinar o algoritmo para que ele possa então prever os rótulos corretos para novos dados.

Aprendizado não Supervisionado (agrupamento de dados): é usado quando não há rótulos disponíveis, o algoritmo tenta encontrar padrões nos dados ou nas estruturas sem a orientação de rótulos.

Aprendizado por reforço (baseado em tentativa e erro): é utilizado para ensinar um agente a tomar decisões em um ambiente, recebendo recompensas ou penalidades de acordo com suas ações. Com o tempo, o agente aprende a tomar decisões que sejam capazes de maximizar a recompensa.

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

O Conjunto de Treinamento refere-se aos exemplos usados para treinar o modelo no processo de aprendizado. Com este conjunto de dados, o modelo aprende a relação entre as variáveis de entrada e as variáveis de saída.

O Conjunto de Validação é usado para comparar diferentes modelos e hiperparâmetros. É usado para avaliar o desempenho do modelo durante o treinamento, identificando se ele está se ajustando bem aos dados ou não.

O Conjunto de Teste é utilizado como comprovação de que aquele modelo realmente está aprendendo algo com o conjunto de dados fornecido, calculando seu desempenho final após o treinamento e o ajuste de parâmetros.

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

Algumas técnicas que podem ser utilizadas para lidar com dados ausentes em um conjunto de dados são:

- Remoção de Dados Ausentes: remover linhas ou colunas com valores ausentes, caso sua remoção não afete de forma significativa o tamanho do conjunto de dados.
- Imputação de Dados: substituir os valores ausentes pela média, mediana ou moda; substituir os valores ausentes com um valor fixo; usar técnicas de machine learning para prever os valores ausentes com base nas outras variáveis.
- Algoritmos: alguns algoritmos de machine learning podem lidar com dados ausentes internamente, sem a necessidade explícita de imputação.
- Sinalização de valores ausentes: criar uma variável que indique a presença de um valor ausente em uma coluna específica.
- Imputação Múltipla: realizar várias imputações diferentes e combinar seus resultados para considerar a incerteza da imputação.

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

A Matriz de Confusão é uma ferramenta usada para avaliar o desempenho de um modelo preditivo, em especial em problemas de classificação. Ela permite visualizar a performance do modelo ao comparar previsões feitas com valores reais. Ela é uma tabela que descreve a performance de um modelo de classificação ao mostrar os acertos e erros de cada classe.

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

Os algoritmos de Machine Learning podem ser aplicados em diversas áreas.

Na saúde, por exemplo, os algoritmos podem ajudar a diagnosticar doenças a partir de imagens médicas, analisar dados genéticos para recomendar tratamentos específicos, prever surtos e auxiliar na gestão hospitalar com a otimização de recursos, como leitos e suprimentos.

Na agricultura, pode ser útil na detecção de pragas e doenças utilizando imagens, pode prever safras através da análise de dados climáticos e de solo, e pode também otimizar o uso de água, fertilizantes e pesticidas com base em dados.

Na construção civil, pode ser utilizada na otimização do uso de materiais e equipamentos, na análise de dados de incidentes para melhoria de segurança no ambiente de trabalho, assim como podem ser utilizados sensores e machine learning no monitoramento de integridade das estruturas, prevendo falhas.

Na manufatura, previsões de falhas em máquinas e equipamentos podem ser feitos através de machine learning, assim como o controle de qualidade que pode ser realizado com o uso da visão computacional. Processos também podem ser otimizados e o desenvolvimento de sistemas inteligentes para automação torna-se possível.

Na logística e no transporte, a previsão de demanda por serviços, a determinação de rotas mais eficientes e a condução autônoma de veículos podem ser beneficiadas pelo uso de machine learning.