



CURSO: <BÁSICO EM MACHINE LEARNING>

- **Atividade 01 (ATIV-01)**

Gisele Melio de Oliveira

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

Machine learning é um ramo da inteligência artificial que permite que sistemas aprendam e tomem decisões a partir de dados, dessa forma, conseguem realizar tarefas de modo automático, sempre melhorando e aprendendo conforme for sendo apresentado a novas bases de dados. Além disso, os algoritmos utilizados no aprendizado da máquina conseguem realizar previsões e identificar padrões mediante às informações que são acrescentadas.

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

Os conjuntos de dataset de treino, validação e teste permitem a medição da performance de algoritmos de classificação e previsão.

Treino: esse conjunto é a base para o aprendizado, nessa fase, é importante se ter uma base de dados com tamanho significativo para que o modelo seja treinado e aprenda de forma competente.

Validação: esse conjunto utiliza uma parte dos dados que não são operados na fase de treinamento do modelo, com isso, pode-se realizar uma avaliação não-viesada, ou seja, sem influência de preconceitos ou fatores externos que podem prejudicar os resultados. Nesta etapa, ao utilizar dados inexplorados, podemos obter uma visão mais realista sobre como o modelo reagirá em situações do mundo real.

Teste: esse conjunto é empregado na avaliação final do modelo, oferecendo uma medida objetiva de seu desempenho. Assim, é possível verificar se o modelo opera efetivamente nas condições do mundo real. Essa análise é crucial para determinar se o modelo apresenta um bom desempenho no conjunto de teste, sugerindo que ele generaliza bem. Além disso, os resultados podem influenciar decisões sobre a implementação do modelo em aplicações práticas, como na produção ou em serviços.



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

Para entender e lidar com os dados ausentes, é necessário verificar as variáveis do banco de dados e a quantidade de dados faltantes. Sendo uma quantidade pequena, optaria pela remoção; porém, sendo uma quantidade significativa, optaria pela inclusão da média ou mediana.

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 utilizada para medir o desempenho de um modelo de classificação, resultando em um entendimento acerca dos acertos e erros do modelo, a partir de previsões, tendo uma visão geral do mesmo. Ela é uma tabela que reconhecerá vários tipos de erros, como por exemplo, em um reconhecimento de imagens, é importante verificar quantas imagens de carros foram classificadas como motos e quantas imagens de motos foram classificadas como carros, pois essa métrica é fundamental para analisar a precisão do modelo.

Os tipos de erros são:

Verdadeiro positivo (TP): Quando não ocorre fraude e o modelo corretamente indica que não há fraude.

Falso positivo (FP): Quando não há fraude, mas o modelo erroneamente classifica como se houvesse.

Falso negativo (FN): Quando há fraude, mas o modelo incorretamente classifica como se não houvesse.

Verdadeiro negativo (TN): Quando há fraude e o modelo corretamente identifica a presença dela.

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

Acho mais interessante a área da saúde.

