

Bootcamp Atlântico - Machine Learning

Atividade 1

Cauan Gomes Dos Santos Barbosa

1. O que é Machine Learning?

Machine Learning (aprendizado de máquina) é uma forma de ensinar computadores a **aprenderem com dados** e tomarem decisões sem que alguém precise programar tudo passo a passo. É como dar exemplos para uma máquina e ela vai entendendo padrões para fazer previsões ou reconhecer coisas. Por exemplo, se você mostrar várias fotos de gatos e cachorros para um programa, ele aprende a identificar o que é um gato e o que é um cachorro.

2. O que são conjunto de treinamento, conjunto de validação e conjunto de teste?

- **Conjunto de treinamento (training set):** É como o material de estudo do modelo. Usamos ele para ensinar o modelo, mostrando vários exemplos e as respostas certas.
 - **Conjunto de validação (validation set):** Esse é usado durante o treinamento para conferir se o modelo está aprendendo bem ou se está decorando os dados. Ajuda a ajustar o modelo.
 - **Conjunto de teste (test set):** Depois que o modelo está pronto, usamos o conjunto de teste para ver como ele se comporta com dados novos, que ele nunca viu antes.
-

3. Como lidar com dados ausentes no conjunto de dados?

Quando um dado está faltando, temos algumas opções:

1. **Remover a linha ou coluna:** Se poucos dados estão ausentes, às vezes é mais fácil ignorar esses pontos.
 2. **Preencher com a média ou mediana:** Por exemplo, se um valor numérico está faltando, dá para colocar a média dos outros.
 3. **Interpolação:** Para dados que seguem uma ordem, podemos estimar o valor que falta com base nos vizinhos.
 4. **Modelos de imputação:** Em casos mais complexos, usamos um modelo para prever o valor ausente com base em outros dados.
-

4. O que é uma matriz de confusão?

A matriz de confusão é uma tabela que mostra como um modelo fez as previsões, comparando o que ele previu com as respostas certas. Ela ajuda a ver se o modelo cometeu erros e que tipo de erro foi.

Exemplo:

- **Verdadeiro Positivo (VP):** Previu que era positivo e estava certo.
- **Verdadeiro Negativo (VN):** Previu que era negativo e estava certo.
- **Falso Positivo (FP):** Previu que era positivo, mas estava errado.
- **Falso Negativo (FN):** Previu que era negativo, mas estava errado.

A matriz ajuda a calcular métricas como **acurácia, precisão e recall**, que indicam a qualidade do modelo.

5. Em quais áreas é mais interessante aplicar Machine Learning?

- **Saúde:** Para prever doenças, analisar exames, e até sugerir tratamentos.
- **Agricultura:** Para prever colheitas ou monitorar a saúde das plantas.
- **Construção Civil:** Para prever o custo de projetos ou detectar falhas estruturais.
- **Manufatura:** Para prever quando uma máquina vai quebrar e evitar problemas na produção.
- **Finanças:** Para identificar fraudes ou ajudar a planejar investimentos.

Eu acho a **saúde** e a **manufatura** muito interessantes! Na saúde, machine learning pode salvar vidas com diagnósticos mais rápidos e precisos. Já na manufatura, pode ajudar a evitar prejuízos e melhorar a eficiência das fábricas.