

Profissão: Cientista de Dados



GLOSSÁRIO



Metodologia



Dica: para encontrar rapidamente a palavra que procura aperte o comando CTRL+F e digite o termo que deseja achar.

- **Implante o modelo**
- **Faça o setup do ambiente**
- **Faça seu primeiro projeto em Python**



Implante o modelo



Implante o modelo

● **Avaliação dos resultados**

Processo de quantificar o impacto financeiro do modelo no negócio, comparando-o com o retorno sobre o investimento do projeto. Difere da avaliação de desempenho, que se concentra na eficácia do modelo em si.

● **Gestão de risco de modelos**

Prática de gerenciar vários modelos em execução ao mesmo tempo, tomando decisões sobre quando desligar um modelo, quando substituí-lo e quando ativar a equipe de desenvolvimento de novos modelos. É especialmente importante em ambientes com um grande volume de modelos em execução simultaneamente.



Implante o modelo

● **Implantação**

Etapa sensível que envolve a decisão sobre a estrutura a ser usada para produzir o produto, como ele será disponibilizado para os clientes finais e como será utilizado. Também envolve o planejamento de como a ferramenta será monitorada após a implantação.

● **Monitoramento**

Processo de criação de indicadores para acompanhar o desempenho do modelo ao longo do tempo e verificar se ele continua identificando os mesmos padrões. É crucial para garantir que o modelo esteja funcionando corretamente e para determinar quando é necessário revisá-lo.



Faça o setup do ambiente



Faça o setup do ambiente

• Markdown

Linguagem de marcação leve que você pode usar para adicionar formatação a texto em documentos de texto simples.

• Commit

Comando usado no Git para salvar as alterações feitas em um arquivo. Cada 'commit' é acompanhado por uma mensagem que descreve as alterações feitas.



Faça seu primeiro projeto em Python



Faça seu primeiro projeto em Python

● **Análise Descritiva Bivariada**

É uma análise estatística que envolve a consideração de duas variáveis ao mesmo tempo. É útil para entender a relação entre duas variáveis.

● **Análise de Distribuição Univariada**

É uma análise estatística que envolve a consideração de uma única variável. É útil para entender a distribuição e a natureza dos dados.



Faça seu primeiro projeto em Python

● **Acurácia**

É uma métrica de avaliação de modelos que mede a proporção de previsões corretas feitas pelo modelo em relação ao total de previsões.

● **Algoritmo Random Forest**

É um método de aprendizado de máquina que opera construindo múltiplas árvores de decisão durante o treinamento e produzindo a classe que é a moda das classes (classificação) ou a média das previsões (regressão) das árvores individuais.



Faça seu primeiro projeto em Python

● Conjuntos de Treinamento e Teste

São subconjuntos de dados usados para treinar um modelo de aprendizado de máquina (conjunto de treinamento) e para avaliar sua performance (conjunto de teste).

● Matriz de Confusão

É uma tabela que permite a visualização do desempenho de um algoritmo de aprendizado de máquina. Ela mostra as previsões corretas (verdadeiros positivos e verdadeiros negativos) e as previsões incorretas (falsos positivos e falsos negativos).



Faça seu primeiro projeto em Python

● **Overfitting**

É um problema comum em aprendizado de máquina, onde um modelo é treinado de forma a se ajustar muito bem aos dados de treinamento, mas não generaliza bem para novos dados.

● **Variáveis Dummy ou Indicadoras**

São variáveis binárias criadas para representar uma variável com duas ou mais categorias.



Bons estudos!

