

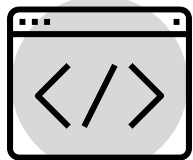
Profissão: Cientista de Dados



BOAS PRÁTICAS



Metodologia



- **Conheça o Crisp-DM**
- **Entenda dados e negócios**
- **Prepare os dados**
- **Implante o modelo**
- **Faça o setup do ambiente**
- **Faça seu primeiro projeto em Python**



Conheça o Crisp-DM

- **Entenda o negócio:** antes de começar a coletar e analisar dados, é crucial entender o negócio e o problema que você está tentando resolver. Isso ajudará a orientar sua análise e garantir que você esteja coletando os dados corretos.
- **Entenda os dados:** depois de coletar os dados, é importante entender o que eles representam, como são coletados e quais são suas limitações. Isso ajudará a garantir que você esteja usando os dados corretamente e a evitar erros de interpretação.



Conheça o Crisp-DM

- **Prepare os dados:** os dados raramente vêm em um formato que é imediatamente útil. Você precisará limpar, transformar e, possivelmente, combinar várias fontes de dados para prepará-los para análise.
- **Modele os dados:** depois que os dados estiverem preparados, você poderá começar a modelar. Isso pode envolver uma variedade de técnicas, dependendo do problema que você está tentando resolver.



Conheça o Crisp-DM

- **Avalie o modelo:** depois de criar um modelo, é importante avaliá-lo para garantir que ele esteja funcionando como esperado. Isso pode envolver a comparação do desempenho do modelo com um conjunto de dados de teste ou a verificação de sua precisão.
- **Implemente o modelo:** depois de avaliar e ajustar o modelo, é hora de implementá-lo. Isso pode envolver a integração do modelo em um sistema existente ou a criação de um novo sistema para usar o modelo.



Conheça o Crisp-DM

- **Repita conforme necessário:** a análise de dados é um processo iterativo. Você pode precisar repetir várias etapas do processo à medida que coleta novos dados, ajusta seu modelo ou refina sua compreensão do problema.



Entenda dados e negócios

- **Defina claramente o objetivo do projeto:** o objetivo pode ser o objetivo do negócio ou da modelagem. Este deve ser bem definido e alinhado com o restante do projeto.
- **Identifique as fontes de dados:** as fontes de dados podem vir de bancos de dados, planilhas excel, ou até mesmo da web. É importante definir como será feita a coleta dos dados e a frequência de atualização dos mesmos.



Entenda dados e negócios

- **Garanta a qualidade dos dados:** a qualidade dos dados é fundamental para a análise. Dados fora do padrão ou com erros podem dificultar a análise. Portanto, é importante verificar a qualidade dos dados antes de iniciar a análise.
- **Prepare um dicionário de dados:** a entrega final desta etapa é um banco de dados montado ou um dicionário de dados, que é uma planilha com o nome de todas as variáveis e seus significados. Isso ajuda a garantir que todos na equipe entendam os dados que estão sendo usados.



Prepare os dados

- **Selecione os dados com base no objetivo do projeto:** escolha variáveis que são relevantes para a análise que você está realizando. Não inclua dados que não contribuem para o objetivo do projeto.
- **Construa novas características:** combine variáveis para produzir novas características que possam ser úteis para a análise. Isso pode envolver a criação de variáveis derivadas ou a aplicação de transformações aos dados.



Prepare os dados

- **Integre seus dados:** junte diferentes bases de dados em uma única base para análise. Isso pode envolver a combinação de dados de diferentes fontes ou a junção de tabelas relacionadas.
- **Formate seus dados:** converta seus dados para o formato desejado para análise. Isso pode envolver a conversão de tipos de dados, a normalização de valores ou a codificação de variáveis categóricas.



Prepare os dados

- Escolha a técnica de modelagem apropriada:** a técnica de modelagem que você escolhe deve ser apropriada para o objetivo do projeto. Por exemplo, se você está tentando prever uma variável contínua, você pode usar a regressão; se você está tentando classificar observações em categorias, você pode usar a classificação.
- Construa e avalie seu modelo:** a construção do modelo envolve várias tentativas e modificações até que um modelo satisfatório seja alcançado. A avaliação do modelo é uma etapa quantitativa, onde a eficácia do modelo é avaliada com base em métricas específicas.



Prepare os dados

- **Avalie os resultados do modelo e implemente o modelo:** após a construção e avaliação do modelo, os resultados devem ser avaliados para determinar se o modelo atende aos objetivos do projeto. Se o modelo for satisfatório, ele pode ser implementado.



Implante o modelo

- Avalie os resultados do modelo em termos de impacto financeiro no negócio, não apenas em termos de métricas de desempenho do modelo. Isso ajudará a entender o retorno sobre o investimento do projeto.
- Planeje a implantação desde o início do projeto. Decida a estrutura a ser usada para produzir o produto, como ele será disponibilizado para os clientes finais e como será utilizado.



Implante o modelo

- Monitore o desempenho do modelo após a implantação. Crie indicadores para acompanhar o desempenho do modelo ao longo do tempo e verifique se ele continua identificando os mesmos padrões.
- Crie um relatório do projeto durante o desenvolvimento do projeto. Este relatório é crucial para manter o conhecimento adquirido durante o desenvolvimento do projeto dentro da empresa.



Implante o modelo

- Gerencie o risco de modelos. Isso envolve o gerenciamento de vários modelos em execução ao mesmo tempo e a tomada de decisões sobre quando desligar um modelo, quando substituí-lo e quando ativar a equipe de desenvolvimento de novos modelos. Isso é especialmente importante em ambientes com um grande volume de modelos em execução simultaneamente.



Faça o setup do ambiente

- Utilize o GitHub para compartilhar, discutir e desenvolver projetos de maneira colaborativa. Além disso, use-o para construir um portfólio de seus projetos de ciência de dados.
- Utilize o Jupyter Notebook com frequência para seus projetos de ciência de dados. Crie uma nova pasta no Jupyter Notebook para seus projetos e mantenha seus arquivos organizados.



Faça o setup do ambiente

- Ao instalar o Anaconda, escolha a opção "instalar apenas para mim" para evitar problemas de permissões.
- Organize seu código e projeto utilizando células de texto em Markdown no Jupyter Notebook. Elas podem ser usadas para adicionar comentários, explicações e formatações ao seu projeto.



Faça o setup do ambiente

- **Mantenha seu perfil no Github atualizado:** isso pode ajudar a construir sua presença online e tornar seu trabalho mais visível para outros profissionais da área.
- Organize seus projetos em repositórios: Ao criar um novo repositório para cada projeto, você mantém seu trabalho organizado e fácil de navegar.



Faça o setup do ambiente

- Faça upload de arquivos relevantes: Certifique-se de que todos os arquivos necessários para entender e replicar seu projeto estão presentes no repositório.
- Faça 'commit' de suas alterações: Sempre que fizer alterações ou adicionar novos arquivos, faça um 'commit' para salvar essas alterações. Inclua uma mensagem de 'commit' clara para documentar o que foi alterado ou adicionado.



Faça o setup do ambiente

- Verifique se seus arquivos são renderizados corretamente: Após fazer o upload e o 'commit' de um arquivo, verifique se ele é renderizado corretamente no Github.



Faça seu primeiro projeto em Python

- Carregue seus dados corretamente. No caso do pandas, você pode usar a função **pd.read_csv** para carregar dados de um arquivo CSV.
- Lembre-se de que a variável resposta (ou variável dependente) é crucial para o seu modelo preditivo. Certifique-se de entender bem o que ela representa e como ela é medida.



Faça seu primeiro projeto em Python

- Na etapa de preparação dos dados, é importante transformar todas as variáveis em formato numérico para que os algoritmos possam trabalhar com elas. Uma das maneiras de fazer isso é através da criação de variáveis 'dummy' ou indicadoras.
- Na etapa de modelagem, é crucial selecionar a técnica de modelagem adequada para o seu conjunto de dados e problema. Além disso, é importante desenhar um teste adequado e avaliar o modelo corretamente.



Faça seu primeiro projeto em Python

- Evite o overfitting dividindo os dados em conjuntos de treinamento e teste. Isso ajuda a garantir que o modelo possa generalizar bem para dados não vistos.
- Na etapa de implementação, pense em como a regra gerada pelo modelo pode ser usada para tomar decisões automáticas. Isso pode envolver a integração do modelo em um sistema de tomada de decisões existente.



Bons estudos!

