# Questionário 12 PASSOS DE CASSIE (S1D3)

Professor: Lauro Teixeira

---------------------------------------------------------------------------------------------------------------------------

Olás tudo bom pessoal ?

Este documento contém perguntas para serem respondidas sobre o que estudamos e vimos sobre os 12 passos de Cassie Kozyrkov. Ao contrário das atividades dos módulos anteriores que eram muito mais práticas, e tudo se resolvia em código, aqui vamos pensar mais no aspecto teórico dos projetos.

As perguntas a seguir são bem gerais e servem mais para formalizar o conteúdo visto e fixar o conhecimento. Sintam se livres para também transformar esse questionario como um documento de referência sua de estudos sobre o tema. Peço que respondam da forma mais completa o possível, esse questionario também será avaliado como desempenho dos alunos.

Em algumas perguntas sobre os 12 passos de cassie, estão bem gerais sobre toda a área de Ciência de dados, então caso precise, pode responder tendo em mente o dia a dia do seu trabalho. Só lembre-se de deixar claro que a resposta dada está sobre a perspectiva do seu contexto.

Por fim, usem a internet ou ChatGPT como um meio de referencia ou para tirar dúvidas pontuáis, e não para resolver as questões. As questão estão bem simples para poderem resolver sem problemas.

---------------------------------------------------------------------------------------------------------------------------

## Questões:

1. Quais são os "12 Passos de Cassie" na Ciência de Dados? Cite todos e escreva com suas próprias palavras um breve res​​​​umo sobre o significado de cada passo.

1 - Pergunte o problema certo. Define a importância de entender claramente o problema de negócio antes de mergulhar na análise de dados.

2 - Consiga os dados certos. Enfatiza a necessidade de coletar dados relevantes e de qualidade para resolver o problema.

3 - Explore os dados. Incentiva uma análise exploratória para familiarizar-se com os dados e encontrar insights preliminares.

4 - Prepare os dados. Destaca a importância de limpar e formatar os dados adequadamente para análise.

5 - Construa modelos. Orienta sobre a escolha e construção de modelos estatísticos ou de aprendizado de máquina para fazer previsões ou inferências.

6 - Escolha a medida de sucesso. Define a importância de estabelecer métricas claras para avaliar o sucesso do modelo.

7 - Divida os dados. Recomenda dividir os dados em conjuntos de treinamento, validação e teste para avaliar o modelo de forma justa.

8 - Treine o modelo. Guia o processo de ajuste do modelo aos dados de treinamento.

9 - Avalie o modelo. Enfatiza a necessidade de usar o conjunto de validação para ajustar hiperparâmetros e o conjunto de teste para testar a performance do modelo.

10 - Lance o modelo. Discute a etapa de colocar o modelo em produção, tornando-o operacional e acessível.

11 - Monitore o modelo. Destaca a importância de monitorar o desempenho do modelo em produção para garantir sua eficácia contínua.

12 - Atualize o modelo. Reconhece a necessidade de atualizar o modelo regularmente com novos dados ou para ajustar às mudanças nas condições ou no comportamento do usuário.

1. Quais são os benefícios de seguir a metodologia de Cassie Kozyrkov em projetos de IA e Ciência de Dados? Seja no seu trabalho ou em qualquer novo projeto da área.

A metodologia oferece uma abordagem sistemática e estruturada para projetos de IA e Ciência de Dados, garantindo que o problema certo está sendo abordado com os dados certos e de maneira ética. Ela ajuda a minimizar riscos e a maximizar a eficácia das soluções propostas, garantindo que os modelos sejam relevantes, precisos e seguros.

1. Quais passos ou aspectos dos "12 Passos de Cassie" você acha mais desafiadores ou complexos de implementar em projetos reais?

Os passos mais desafiadores podem incluir a obtenção dos dados certos devido a limitações de acesso ou qualidade dos dados, a avaliação e o lançamento do modelo em um ambiente de produção devido a desafios técnicos e operacionais, e o monitoramento contínuo e atualização do modelo, que exige uma infraestrutura robusta e um comprometimento contínuo.

1. Ná prática, você realmente acha que é possível seguir todos os 12 passos ou mesmo se são obrigatórios ?

Embora seja ideal seguir todos os 12 passos para garantir a eficácia e a eficiência do projeto, na prática, a aplicabilidade de cada passo pode variar dependendo do contexto específico do projeto. Alguns passos podem ser mais críticos do que outros, e a flexibilidade para adaptar a metodologia às necessidades específicas do projeto é importante.

1. Proponha uma alteração ou um novo passo (específique onde ele entraria) para tornar a framework mais interessante para ser utilizada na prática. Justifique.

Proposta de Novo Passo: Passo 4.5 - "Avaliação Preliminar de Viés e Justiça"

Este novo passo seria inserido entre o Passo 4 (Prepare os dados) e o Passo 5 (Construa modelos), para enfatizar a importância de avaliar e mitigar potenciais vieses nos dados antes de construir modelos. A justificativa para este passo é que os modelos de Machine Learning (ML) e Inteligência Artificial (IA) podem perpetuar ou até amplificar vieses presentes nos dados de treinamento. Avaliar os dados com um olhar crítico para viés e justiça pode ajudar a prevenir a produção de modelos que produzam resultados discriminatórios ou injustos, promovendo soluções mais éticas e equitativas.

1. Explique a importancia do suposto passo 0 da Cassie. E descreva, você acha que toda solução de ciência de dados é sobre Machine Learning/ IA ?

O Passo 0 é fundamental porque estabelece a base para uma abordagem pragmática e orientada a problemas na ciência de dados. Ele questiona a necessidade real de aplicar IA ou ML, abrindo espaço para considerar soluções mais simples que podem ser mais apropriadas para o problema em questão. Isso não só economiza recursos, mas também alinha as expectativas e esforços com as necessidades reais do projeto.

Nem toda solução de ciência de dados é sobre Machine Learning/IA. Muitos problemas podem ser resolvidos com análises estatísticas básicas, visualizações de dados, ou até mesmo com processos de negócio melhor estruturados. O foco deve estar em resolver o problema da maneira mais eficaz e eficiente, não necessariamente na utilização da tecnologia mais avançada disponível.

1. Sobre o Passo 2, o quanto pode ser relevantes os dados abertos que estão na internet para um projeto ?

Dados abertos disponíveis na internet podem ser extremamente relevantes para um projeto, especialmente em estágios iniciais de pesquisa e desenvolvimento. Eles podem oferecer insights valiosos, complementar conjuntos de dados internos com novas dimensões ou perspectivas, e ajudar na validação de modelos. Por exemplo, dados abertos sobre padrões climáticos, tendências econômicas ou demográficas podem enriquecer análises em diversos campos, desde a saúde pública até a gestão de cadeias de suprimentos.

1. Sobre o Passo 8, diga quais problemas geralmente podem acontecer se seus modelos não estiverem validados e seguros para serem utilizados em produção.

Modelos não validados e inseguros podem causar uma série de problemas quando colocados em produção, incluindo:

Decisões Incorretas: Resultados imprecisos podem levar a decisões de negócios erradas.

Viés e Injustiça: Modelos podem perpetuar ou amplificar vieses, levando a resultados discriminatórios.

Vulnerabilidades de Segurança: Falhas de segurança podem ser exploradas, comprometendo dados sensíveis.

Rejeição por Parte dos Usuários: Falta de confiança nos resultados do modelo pode levar à rejeição do sistema pelos usuários finais.

1. Sobre o Passo 10, qual a importância de preparar um modelo para colocar no ar ? Em sua opnião isto é pode ser mais difícil do que treinar o próprio modelo ?

Preparar um modelo para produção é crucial para garantir sua escalabilidade, segurança e integração com sistemas existentes. Este passo pode ser mais desafiador do que treinar o modelo, pois envolve questões como otimização de performance, gerenciamento de infraestrutura, monitoramento em tempo real e garantia de que o modelo possa ser atualizado sem interrupções. A complexidade aumenta com a necessidade de garantir que o modelo funcione bem em condições de mundo real, onde a variabilidade e a imprevisibilidade dos dados podem ser maiores do que nos conjuntos de dados de treinamento.

1. Sobre o passo 12, qual a importância de monitorar a performance de um modelo em produção ? Quais problemas é possível detectar ?

O monitoramento contínuo é vital para assegurar que o modelo mantenha seu desempenho ao longo do tempo e sob condições variáveis. Problemas detectáveis incluem:

Degradação do Modelo: Performance pode diminuir à medida que padrões nos dados mudam.

Erros Inesperados: Novos tipos de dados ou padrões não vistos durante o treinamento podem causar erros.

Viés ou Injustiça Emergentes: Novos dados podem revelar vieses não detectados anteriormente.

A monitorização permite intervenções rápidas, como o reajuste do modelo ou a coleta de novos dados para treinamento, garantindo que o modelo continue relevante e eficaz.