# Questionário CRISP DM (S2D3)

Professor: Lauro Teixeira

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

Olás tudo bom pessoal ?

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.

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. O que é o CRISP-DM? Por favor, explique em suas próprias palavras o que você entende por CRISP-DM e qual é o seu objetivo em projetos de mineração de dados.

CRISP-DM é um modelo de processo amplamente aceito para projetos de ciência de dados, projetado para guiar profissionais através das complexidades de um projeto, desde a compreensão do problema de negócios até a implantação do modelo. Seu objetivo é fornecer uma estrutura metodológica para planejar, executar e avaliar de dados de forma eficaz.

1. Escreva um breve parágrafo sobre os passos do CRISP-DM.

Incluem Entendimento do Negócio, Entendimento dos Dados, Preparação dos Dados, Modelagem, Avaliação e Implantação. Esses passos são iterativos e flexíveis, permitindo ajustes conforme o projeto evolui.

1. Por que a avaliação de modelos é uma etapa essencial no CRISP-DM?

É essencial para garantir que o modelo atenda às necessidades do negócio, avaliando sua precisão, eficácia e aplicabilidade antes de ser implantado em um ambiente de produção.

1. Um ponto interessante sobre o CRISP-DM é como todos os seus passos basicamente geram em torno dos dados, até mesmo as avaliações e feedbacks. Por que um modelo centrado em seus dados pode ajudar a projetar uma solução mais eficiênte ?

Focar nos dados ajuda a garantir que a solução seja fundamentada na realidade observável, aumentando a relevância e eficácia da solução ao abordar diretamente os problemas identificados através da análise de dados.

1. Você sente falta de alguma etapa no CRISP-DM ? Justifique.

Algumas pessoas podem sentir falta de uma ênfase explícita na ética e na privacidade dos dados, o que é crucial na era atual da ciência de dados.

1. Como a dificuldade de manipular, obter, limpar, analisar dados pode dificultar a vida do cientista de dados durante o design de alguma soluca de ciência de dados ?

Dificuldades na obtenção, limpeza e análise de dados podem atrasar o progresso, aumentar os custos e até mesmo comprometer a validade dos resultados, tornando a tarefa do cientista de dados mais complexa.

1. Nesta questão, gostaria que vocês fizessem um breve pesquisa sobre: "Tempo gasto em etapas de um projeto de machine learning", e conte aqui um pouco do que aprenderam. Além disso, gostaria que vocês relacionassem as etapas do CRISP-DM com o tempo gasto em cada etapa de um projeto. Por fim escreva uma conclusão com sua opnião.

Pesquisas mostram que cientistas de dados gastam a maior parte do tempo na preparação dos dados, seguida pela modelagem e ajuste dos modelos. Relacionando ao CRISP-DM, as etapas de Entendimento e Preparação dos Dados são as mais demoradas, mas essenciais para a qualidade do projeto.

1. No Questionário sobre ML Canvas, o último exercício era a construção do ML Canvas de um projeto de "Criar um projeto de Machine Learning para prever preço das ações da bolsa de valores", descreva um pouco dos passo a passo deste mesmo projeto utilizando o Crisp-DM e escreva sobre a relação entre os projetos.

O processo começaria com o Entendimento do Negócio, definindo objetivos claros; seguido pelo Entendimento dos Dados, coletando dados históricos de preços de ações; na Preparação dos Dados, limpeza e organização dos dados seriam essenciais; a Modelagem envolveria selecionar e treinar modelos preditivos; a Avaliação testaria a precisão do modelo; e a Implantação colocaria o modelo em uso prático.

1. Sobre ML Canvas e CRISP-DM: Você acredita que é possível usar elementos de ambos os modelos em um único projeto de análise de dados? Se sim, como eles poderiam se complementar?

Sim, é possível e benéfico combinar ambos. Enquanto o CRISP-DM oferece uma estrutura de processo detalhada, o ML Canvas pode ajudar a esclarecer e comunicar a visão e os componentes chave do projeto, complementando o CRISP-DM com um resumo visual e estratégico do projeto.

1. Sobre ML Canvas e CRISP-DM: Existem algum passo que se preocupa com "ética" em algum de seus processos ? Pesquise um pouco sobre "Ética em Machine Learning" e proponha como este tópico agregaria durante a modelagem de um projeto de ciência de dados.

A inclusão de considerações éticas é vital para garantir que os projetos de ML sejam justos, transparentes e não discriminatórios. No CRISP-DM, poderia haver uma etapa adicional dedicada à avaliação ética, onde questões de viés, privacidade e consentimento dos dados são analisadas e endereçadas antes da modelagem e implantação.