

Desafio nº1: Explique de forma resumida todas as etapas que compõe a metodologia

A metodologia CRISP-DM é um framework iterativo e incremental que fornece uma estrutura para o desenvolvimento de projetos de ciência de dados. É uma metodologia flexível que pode ser adaptada a diferentes tipos de projetos e contextos.

As seis etapas da metodologia CRISP-DM são as seguintes:

1. Entendimento do negócio

Nesta etapa, o objetivo é entender o problema que o projeto de ciência de dados pretende resolver. Isso envolve compreender os objetivos do negócio, as necessidades dos usuários e os dados disponíveis.

Por exemplo, uma empresa de e-commerce pode querer desenvolver um modelo de previsão de abandono de carrinho. Para isso, a empresa precisa entender as razões pelas quais os clientes abandonam seus carrinhos, quais dados estão disponíveis para analisar e quais são os objetivos da empresa ao desenvolver o modelo.

2. Entendimento dos dados

Nesta etapa, o objetivo é entender os dados que serão usados no projeto. Isso envolve coletar os dados, limpar e preparar os dados, e explorar os dados para identificar padrões e tendências.

No exemplo do e-commerce, a empresa precisa coletar dados sobre os clientes, como histórico de compras, comportamento de navegação e informações demográficas. A empresa também precisa limpar os dados para remover erros e inconsistências, e preparar os dados para análise. Por exemplo, a empresa pode agrupar os dados por cliente ou por produto.

3. Preparação dos dados

Nesta etapa, o objetivo é preparar os dados para que possam ser usados para construir o modelo. Isso envolve transformar os dados para atender aos requisitos do modelo, e reduzir o tamanho dos dados para melhorar o desempenho do modelo.

No exemplo do e-commerce, a empresa pode precisar transformar os dados de texto em dados numéricos, ou reduzir o tamanho dos dados removendo características irrelevantes.

4. Modelagem

Nesta etapa, o objetivo é construir o modelo de ciência de dados. Isso envolve escolher o tipo de modelo adequado para o problema, e ajustar os parâmetros do modelo para obter o melhor desempenho.

No exemplo do e-commerce, a empresa pode escolher um modelo de regressão logística para prever a probabilidade de um cliente abandonar seu carrinho. A empresa pode ajustar os parâmetros do modelo para considerar fatores como o histórico de compras do cliente, o comportamento de navegação do cliente e as informações demográficas do cliente.

5. Avaliação

Nesta etapa, o objetivo é avaliar o desempenho do modelo. Isso envolve usar dados de teste para estimar a precisão do modelo.

No exemplo do e-commerce, a empresa pode usar dados de compras anteriores para avaliar o desempenho do modelo. A empresa pode calcular a precisão do modelo, ou a probabilidade de o modelo prever corretamente se um cliente abandonará seu carrinho.

6. Implantação

Nesta etapa, o objetivo é implantar o modelo em produção. Isso envolve colocar o modelo em uso para gerar previsões ou insights.

No exemplo do e-commerce, a empresa pode implantar o modelo em seu site para alertar os clientes sobre o risco de abandono de carrinho. A empresa também pode usar o modelo para melhorar suas estratégias de marketing e vendas.

A metodologia CRISP-DM é uma ferramenta valiosa para o desenvolvimento de projetos de ciência de dados. Ela ajuda os cientistas de dados a garantir que os projetos sejam bem-sucedidos, fornecendo uma estrutura para o processo de desenvolvimento.