

OFICINA 2

usarei como exemplo um supermercado

1- *Entendimento do negócio*

No contexto de um supermercado, o foco pode ser aumentar as vendas, melhorar a gestão de estoque ou personalizar recomendações para os clientes. Identificar padrões de compra dos clientes para otimizar promoções e sugerir produtos relevantes, aumentando o ticket médio de compras.

2 - *Entendimento dos dados*

Em um supermercado, os dados podem incluir:

- **Histórico de compras** (produtos adquiridos, data, hora, cliente)
- **Dados demográficos** dos clientes (idade, localização, comportamento de compra)
- **Estoque de produtos** (itens mais vendidos, sazonalidade)
- **Promoções anteriores** e seus impactos

Esses dados vão nos guiar a entender padrões de consumo e a definir as estratégias mais eficazes.

3 - *Preparação dos dados*

Nesta etapa, transformamos os dados brutos em um formato adequado para análise. As principais atividades incluem:

Limpeza de Dados: Remover registros duplicados, lidar com valores ausentes (ex.: preenchendo com a média) e corrigir erros de formatação.

Criação de novas Features: Criar novas variáveis úteis, como frequência de compras, valor médio gasto por cliente e categorias de produtos mais compradas.

Normalização e Categorização: Ajustar os dados para garantir que estejam prontos para os algoritmos de aprendizado de máquina, como converter datas em períodos do dia (manhã/tarde/noite) e transformar variáveis categóricas em valores numéricos.

4 - **Modelagem**

Nesta fase, aplicamos técnicas de Mineração de Dados para extrair padrões e insights. Para o supermercado, podemos usar por exemplo:

Algoritmos de Clusterização (K-Means, DBSCAN) para agrupar clientes com perfis de compra semelhantes para oferecer promoções personalizadas.

5- **Avaliação**

Após a modelagem, avaliamos a eficácia dos modelos para garantir que os insights gerados estão alinhados aos objetivos de negócio.

Algumas métricas de avaliação são:

Para previsões de vendas: RMSE (Root Mean Squared Error) e MAE (Mean Absolute Error)

Para agrupação de clientes: Silhouette Score (avaliar a qualidade dos clusters)

Se os modelos não estiverem performando bem, ajustamos hiperparâmetros e refinamos os dados para melhorar os resultados.