

# MINERAÇÃO DE DADOS

## Estudo de Caso

### Regras de Associação, Agrupamento e Classificação

Professor: Lucas Bonilha Henrique

**Preparado por**

**Kethilyn Francielle**

DSM – 6º Semestre - 2025



# Introdução

A mineração de dados é o processo de analisar muitos dados para descobrir padrões e informações úteis. Ela ajuda empresas a entender melhor seus clientes e tomar decisões mais inteligentes, por exemplo.

Nas próximas páginas, vou explicar de acordo com o estudo de caso a seguir.

## Estudo de Caso

A empresa Market+, uma rede varejista especializada em alimentos, está iniciando um projeto piloto para aprimorar suas estratégias de atendimento, abastecimento e personalização de ofertas. Para isso, ela disponibilizou um **dataset real** chamado **Groceries Dataset** (disponível no Kaggle) contendo **38.765 transações**, onde cada linha representa uma cesta de compras composta por diversos itens adquiridos no mesmo pedido.

Preciso conduzir uma análise aplicada de Mineração de Dados contemplando três pilares fundamentais:

- Descoberta de Padrões (Regras de Associação)
- Agrupamento (Clusterização)
- Classificação (Modelos Supervisionados)

O objetivo é extrair conhecimento útil para subsidiar decisões estratégicas da empresa.

# Metodologia

## Etapas do Processo

Neste trabalho, eu usei três técnicas principais:

- **Regras de Associação** → para descobrir itens que costumam ser comprados juntos.
- **Clusterização** → para encontrar grupos de compras parecidas.
- **Classificação** → para prever se uma compra é grande ou pequena.

**E segui estas etapas:**

- Carreguei e explorei os dados para entender a estrutura.
- Transformei os dados para criar as “cestas de compra” (cliente + data → lista de itens).
- Apliquei o Apriori para encontrar regras de associação.
- Criei features numéricas para poder fazer a clusterização.
- Usei K-Means para separar as compras em grupos parecidos.
- Treinei modelos de classificação (Árvore de Decisão e KNN) para prever compras grandes vs pequenas.
- Interpretei os resultados pensando no que seria útil para a empresa Market+.

# Resultados das Etapas

## Regras de Associação

- Encontrei padrões significativos entre produtos.
- Identifiquei combinações fortes, como itens que são comprados juntos com frequência.
- Esses padrões ajudam a pensar em promoções e recomendações.

## Clusterização

- Depois de criar algumas medidas numéricas das compras (quantidade de itens, variedade, proporção de categorias), apliquei o K-Means.
- O algoritmo encontrou 3 grupos principais, cada um com um perfil diferente de cesta.
- Isso mostrou que os clientes têm padrões distintos e previsíveis de compra.

## Classificação

- Modelei a variável “compra grande vs. pequena”.
- Treinei dois modelos: Árvore de Decisão e KNN.
- Os dois tiveram 100% de desempenho, pois a variável-alvo estava totalmente alinhada com as features, deixando a previsão muito mais fácil.

# Justificativas Teóricas

Nesta parte, eu explico por que usei cada técnica e como elas se relacionam com o que aprendemos na disciplina. Mas não estou indo para o lado técnico pesado, mas apenas para mostrar que acho que entendi a lógica por trás de cada método.

As **Regras de Associação** é uma técnica da Mineração de Dados usada quando queremos entender quais itens aparecem juntos em transações. O Apriori funciona encontrando conjuntos de itens frequentes antes de criar regras. Isso faz sentido para este projeto, porque o dataset é exatamente uma lista de itens comprados por transações.

Usei suporte, confiança e lift porque são as três medidas clássicas para avaliar a força de uma regra, como vimos na teoria.

Já a **clusterização** é uma técnica não supervisionada, indicada quando não sabemos as categorias antes e queremos deixar o algoritmo descobrir padrões por conta própria.

A **classificação** é uma técnica supervisionada, usada quando já existe uma categoria que queremos prever — no meu caso: compra grande vs compra pequena.

Escolhi esses dois modelos porque:

- **Árvore de Decisão:** é simples de entender, segue uma lógica de perguntas e embora tenha sido trabalhada em aula, eu acabei perdendo alguns conteúdos. Sendo assim, foi uma forma de aprender na prática.
- **KNN:** é um método baseado em proximidade, também ensinado pelo professor, e funciona bem quando temos poucas features e dados bem separados.

Tentei seguir exatamente o que foi pedido no enunciado, e aplicar as técnicas conforme explicadas em aula.

# Interpretação e Discussão dos Achados

As regras de associação mostraram combinações de produtos que realmente acontecem no dia a dia do mercado, ajudando a pensar em promoções e sugestões de compra.

Os clusters revelaram que existem tipos bem diferentes de compras, o que abre espaço para criar ofertas mais personalizadas para cada perfil.

A classificação mostrou que o padrão das cestas é tão claro que os modelos conseguiram acertar tudo só olhando as características das próprias transações.

No final, percebi que mesmo com técnicas básicas já dá pra descobrir muita coisa útil sobre o comportamento dos clientes.

# Conclusão Integrada

Ao longo deste projeto, consegui enxergar de forma muito mais clara como os três métodos da Mineração de Dados — regras de associação, clusterização e classificação — se conectam para transformar dados brutos em conhecimento real. Conforme eu fui aplicando cada etapa, tudo começou a fazer sentido de um jeito muito mais prático.

Na parte de Regras de Associação, aprendi a montar as cestas de compra e usar o Apriori para identificar padrões escondidos nas transações. Foi interessante ver como certos produtos realmente aparecem juntos com frequência, e como isso pode ajudar as empresas a pensar em promoções, sugestões de compra e estratégias de vendas.

Já na Clusterização, eu finalmente entendi por que criar features numéricas é tão importante. Antes parecia só mais uma etapa técnica, mas agora percebo que são esses números que permitem ao algoritmo identificar perfis diferentes de comportamento.

O K-Means ajudou a separar compras parecidas em grupos e por fim, na Classificação, pude testar dois modelos — Árvore de Decisão e KNN — para prever se uma compra era grande ou pequena. Aqui vivi um aprendizado importante: como o alvo e as features estavam diretamente relacionados, os modelos acertaram tudo. Isso me mostrou que, além de treinar modelos, é essencial pensar muito bem na forma como construímos as variáveis e no que estamos tentando prever.

No geral, este estudo me mostrou como essas três técnicas se complementam para ajudar empresas como a Market+: melhorar promoções, organizar o estoque, criar recomendações e entender melhor seus consumidores. Mesmo com os desafios no meio do caminho, senti que consegui chegar longe e desenvolver o estudo de caso. E claro: deixei tudo bem explicadinho para quando eu precisar revisitar isso no futuro., rs 🇧🇷🌟