

Mineração de Padrões Frequentes em Dados de E-Commerce

Alunos: Gilmar Junio, Fabrício Chaves

1. Contexto

Análise de transações de vendas de um e-commerce do Reino Unido para identificar padrões de compra frequentes.

2. Relevância

- Compreender o comportamento de compra dos clientes.
- Melhorar estratégias de *cross-selling* e *up-selling*.
- Otimizar estoque e promoções com base em produtos frequentemente comprados juntos.

3. Objetivos

Principal:

Identificar itens frequentemente comprados juntos para recomendações de produtos e estratégias de marketing.

Específicos:

- Encontrar *itemsets* frequentes (conjuntos de produtos que aparecem juntos em transações).
- Extrair regras de associação (ex.: "Se compra X, então compra Y").
- Analisar diferenças nos padrões por país ou tipo de cliente.

4. Critérios de Sucesso

- Identificar regras de associação.
- Gerar visualizações claras (ex.: matriz de calor, grafos de associação).
- Propor ações práticas (ex.: promoções combinadas).

5. Recursos

- **Dataset:** 500 mil transações de e-commerce (CSV).
- **Ferramentas:** Python e bibliotecas relevantes.
- **Algoritmos:** Apriori, FP-Growth.

6. Requisitos

- Pré-processamento para filtrar transações canceladas (`TransactionNo` com "C").
- Agrupar itens por transação (formato: `{item1, item2, ...}`).

7. Suposições

- Padrões de compra são consistentes ao longo do período analisado.

8. Riscos e Contingências

- **Risco:** Dados esparsos (muitos itens únicos).
 - **Solução:** Ajustar `min_support` ou agrupar produtos similares.

9. Objetivos da Mineração

- Aplicar Apriori/FP-Growth para encontrar *itemsets* frequentes.
- Gerar regras com suporte, confiança e *lift*.
- Filtrar regras não triviais (`lift > 1`).

10. Tarefas

Pré-processamento:

- Limpar transações canceladas.
- Converter dados para formato transacional.

Mineração:

- Aplicar Apriori com `min_support=0.01`.
- Gerar regras com `min_confidence=0.5`.

Avaliação:

- Analisar regras por métricas (suporte, confiança, *lift*).

Visualização:

- Gráfico de redes ou matriz de calor para regras.