

UNIVERSIDADE FEDERAL DE SANTA MARIA

CENTRO DE TECNOLOGIA

Trabalho 1 de Mineração de Dados

Aluno: Diogo Rocha Marques

Pré-processamento

Devido à impossibilidade de utilizar a biblioteca `jsonlite` no RStudio, optei por converter o arquivo JSON para CSV. Realizei o download do arquivo e o salvei no formato CSV, o que resolveu imediatamente os erros relacionados aos caracteres Unicode presentes no arquivo JSON, além de corrigir o problema das vírgulas no arquivo.

Linguagem Escolhida

A linguagem R foi escolhida pela sua facilidade em manipular arquivos CSV. Também foram realizados testes em Python utilizando a biblioteca `pandas`. Anexo ambos os scripts: o funcional em R e o teste em Python.

Funcionamento

Resultados das Regras de Associação

1. **Exibição de Regras:** Após a execução do código, o primeiro resultado exibido é uma lista de todas as regras de associação encontradas. Se não houver regras, a mensagem "Nenhuma regra encontrada." será apresentada.

Interpretação:

- Cada regra possui um antecedente (produtos que implicam a compra de outros) e um consequente (produtos que são adquiridos como resultado).
 - É importante observar as métricas associadas a cada regra, como suporte e confiança, que indicam a relevância da regra.
2. **Verificação da Presença de "Doce":** O código verifica a presença do produto "Doce" nas transações.

Interpretação:

- Se a mensagem indicar que "Doce" está presente, isso sugere que há compras que incluem este item, o que é relevante para entender sua popularidade nas transações.
- Caso contrário, deve-se investigar se há um problema nos dados ou se realmente não foram realizadas compras de "Doce".

3. **Regras que Implicam a Compra de "Doce":** O código filtra e exibe regras que têm "Doce" como consequente.

Interpretação:

- Se regras forem encontradas, é fundamental analisar as antecedências, revelando quais produtos são frequentemente comprados junto com "Doce". Por exemplo, se a regra "Pão -> Doce" aparecer, isso indica que clientes que compram pão também tendem a comprar doces.
 - Essas informações podem ser valiosas para estratégias de marketing, promoções cruzadas ou disposição de produtos nas prateleiras.
4. **Métricas de Regras:** As regras podem ser acompanhadas de métricas como suporte, confiança e lift.
- **Suporte:** Proporção de transações que contêm tanto a antecedente quanto a consequente, ajudando a entender a popularidade de uma regra.
 - **Confiança:** Indica a probabilidade de que a consequente ocorra quando a antecedente está presente, sendo uma medida da força da regra.
 - **Lift:** Relação entre a confiança da regra e a probabilidade de observar a consequente independentemente. Um lift maior que 1 indica uma associação positiva entre antecedente e consequente.

Conclusões

- **Insights de Negócio:** As regras de associação ajudam a entender o comportamento do consumidor e como diferentes produtos estão inter-relacionados nas compras.
- **Decisões Estratégicas:** A análise das regras pode orientar decisões de marketing, como a criação de promoções para produtos frequentemente adquiridos juntos, melhorando a experiência do cliente e aumentando as vendas.
- **Oportunidades de Marketing:** Se "Doce" se mostra um item frequentemente comprado em conjunto com outros produtos, campanhas promocionais direcionadas podem ser eficazes.

Análise da Saída do Apriori

1. **Parâmetros da Especificação:**

- **Confidence (Confiança):** Mínima de 0.5, significando que, para uma regra ser válida, pelo menos 50% das vezes que o antecedente (lhs) é comprado, o consequente (rhs) também deve ser adquirido.
- **Support (Suporte):** Mínimo de 0.05, ou seja, uma regra só é considerada se for encontrada em pelo menos 5% das transações.
- **Maxlen (Comprimento Máximo):** 10, definindo que a regra pode conter até 10 itens no lado esquerdo ou direito.
- **Count:** O suporte absoluto mínimo é 3, indicando que a regra deve aparecer em pelo menos 3 transações.

2. **Transações:**

- O conjunto de dados contém 23 itens e 79 transações.

- O algoritmo considerou um total de 21 itens após a filtragem.

3. **Regras Geradas:** As regras de associação encontradas foram as seguintes:

lhs	rhs	support	confidence	coverage	lift	count
{Presunto Perdigo}	{Pão Francês}	0.0633	0.5556	0.1139	2.1944	5
{Queijo Minas}	{Pastel Queijo}	0.0633	0.5000	0.1266	2.4688	5
{Refri - Coca-Cola}	{Queijo Mussarela}	0.0759	0.5000	0.1519	1.6458	6
{Pão Francês}	{Queijo Mussarela}	0.1266	0.5000	0.2532	1.6458	10

Interpretação das Regras:

- **Regra 1:** {Presunto Perdigo} => {Pão Francês}
 - **Support:** 6.33% das transações incluem tanto presunto quanto pão francês.
 - **Confidence:** Há 55.56% de chance de que um cliente que compra presunto também adquira pão francês.
 - **Lift:** A relação entre presunto e pão francês é 2.19 vezes maior do que o esperado ao acaso, indicando uma forte associação.
 - **Regra 2:** {Queijo Minas} => {Pastel Queijo}
 - **Support:** 6.33% das transações incluem queijo minas e pastel de queijo.
 - **Confidence:** Há 50% de chance de que clientes que compram queijo minas também comprem pastel de queijo.
 - **Lift:** A associação é 2.47 vezes maior do que o esperado.
 - **Regra 3:** {Refri - Coca-Cola} => {Queijo Mussarela}
 - **Support:** 7.59% das transações incluem refrigerante Coca-Cola e queijo mussarela.
 - **Confidence:** Há 50% de chance de que um cliente que compra Coca-Cola também compre queijo mussarela.
 - **Lift:** A relação é 1.65 vezes maior do que o esperado, mas é uma associação mais fraca em comparação com as anteriores.
 - **Regra 4:** {Pão Francês} => {Queijo Mussarela}
 - **Support:** 12.66% das transações incluem pão francês e queijo mussarela.
 - **Confidence:** Há 50% de chance de que um cliente que compra pão francês também compre queijo mussarela.
 - **Lift:** A associação é 1.65 vezes maior do que o esperado.
4. **Ausência de "Doce":**
- A saída indicou que "Doce" não está presente nas transações. Isso significa que, nas transações analisadas, não houve registros de compras que incluíssem o item "Doce".
 - Portanto, não é possível gerar regras que impliquem a compra de "Doce", uma vez que não existem transações onde "Doce" seja um item presente.

Conclusão para o Relatório

- **Regras de Associação:** As regras de associação identificadas demonstram relações significativas entre produtos frequentemente comprados juntos. A análise das métricas de

suporte, confiança e lift sugere que certos produtos estão relacionados de maneira que pode ser explorada para estratégias de venda e marketing.

- **Relevância da Análise:** Compreender essas associações pode auxiliar na otimização de estoques, promoções e na disposição dos produtos em lojas físicas ou online.
- **Falta de Dados sobre "Doce":** A ausência do item "Doce" nas transações implica que a estratégia de marketing envolvendo esse produto pode necessitar de revisão, possivelmente aumentando a visibilidade ou promovendo combinações que incluam "Doce".