



**INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO  
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**ANDREAS MUNTE FOERSTER – [andreas.foerster@usp.br](mailto:andreas.foerster@usp.br) – #7143997**

**EXPERIMENTAÇÃO COM ALGORÍTMO APRIORI**

**São Carlos  
Outubro de 2018**



## **Enunciado do Problema**

Em anexo, está disponibilizada uma base de dados do supermercado Dotto, que contém todas as transações realizadas em um determinado período.

Cada linha do arquivo refere-se a uma transação (uma compra) e contém todos os produtos que foram adquiridos nessa transação. Faça os processamentos que julgar adequados.

Escolha alguma implementação já pronta do algoritmo Apriori (no Weka, R, Python, ou outra), gere as regras de associação e analise as regras obtidas, considerando medidas de interesse (suporte, confiança, lift, etc). Em sua análise, apresente regras que julgar interessantes para o domínio em questão e diga se são regras óbvias ou se alguma novidade pode ser obtida.

Escreva um breve relatório, relatando as decisões de projeto (processamentos realizados) e as análises realizadas.

## **Metodologia**

Neste trabalho, foi opado pela linguagem de programação Python devido a sua fama de excelente ferramenta para executar algoritmos de análise de dados. A implementação do algoritmo Apriori foi provida pela biblioteca `efficient_apriori`.

Como os dados utilizados para este exercício emulam registros de vendas de um supermercado, muitos dos produtos são similares em sua natureza e poderiam ser agrupados. Para permitir uma análise mais ampla, resultados do algoritmo forma colegados com e sem o agrupamento de itens semelhantes. Vale resaltar que o agrupamento de itens semelhantes requer uma análise cautelosa dos dados existentes. O agrupamento realizado, de forma alguma é ideal e foi feito apenas para exemplificar a alteração dos resultados do algoritmo.

O código utilizado para este exercício pode ser obtido pelo repositório Git em [https://github.com/Maglethong-USP/SCC0230\\_association-exercise](https://github.com/Maglethong-USP/SCC0230_association-exercise).

Parâmetros do algoritmo utilizados foram:

- Confiança Mínima: 50%
- Suporte Mínimo: 2% (devido ao volume de dados)

Apenas os dados com maior Lift foram coletados para este relatório.

## Resultados

### Dados não agrupados:

Item 1	Item 2	Confiança	Suporte	Lift
Creme de Leite Nestle	Leite Moça	0.61	0.05	5.48
Coca Cola + Creme de Leite Nstle	Leite Moça	0.66	0.03	5.98
Fermento Royal	Leite Moça	0.71	0.03	6.37
Biscoito Nestle + Leite Moça	Creme de Leite Nestle	0.54	0.02	6.38
Coca Cola + Leite Moça	Creme de Leite Nsetle	0.55	0.03	6.52
Biscoito Nestle + Creme de Leite Nestle	Leite Moça	0.79	0.24	7.12

### Dados agrupados:

Item 1	Item 2	Confiança	Suporte	Lift
Bombril + Papel Higiênico	Toalha de Papel	0.55	0.02	8.34
Fermento Royal	Leite Moça + Papel Higiênico	0.59	0.02	8.58
Biscoito Nestle + Leite Moça + Papel Higiênico	Toalha de Papel	0.58	0.02	8.78
Detergente Limpol + Papel Higiênico	Toalha de Papel	0.58	0.03	8.78
Leite Moça + Toalha de Papel	Biscoito Nestle + Papel Higiênico	0.6	0.02	9.75
Biscoito Nestle + Creme de Leite Nestle	Leite Moça + Papel Higiênico	0.67	0.02	9.79
Biscoito Nestle + Toalha de Papel	Leite Moça + Papel Higiênico	0.71	0.02	10.35

## **Discussão**

Observando os resultados é notável como alguns produtos comumente utilizados em receitas brasileiras de doces aparecem agrupados com Lift alto.

Adicionalmente, nota-se como um simples agrupamento de alguns itens apenas ignorando marcas, traz resultados com lift muito mais alto. Tal alteração revela, por exemplo, tendências naturais de compras semanais ou mensais, englobando material de limpeza, papel higiênico e Papel toalha. A hipótese inicial de produtos utilizados em receitas de doces estarem relacionados é fortalecida neste resultado com a presença de mais combinações favoráveis com lift superior.

Para possibilitar uma análise ainda melhor, mais agrupamentos de itens com marcas distintas poderiam ser feitos. Tal agrupamento é delicado e trabalhoso, mas pode ser extremamente benéfico para a análise final. Para este exemplo, os poucos agrupamentos que foram realizados já evidenciam tendências pouco visíveis no primeiro experimento, provando como esta etapa pode ser de grande valor para a análise final.