

**Universidade Tecnológica Federal do Paraná**  
**Departamento de Informática**

**Mineração de Dados**  
**Thiago H Silva**

**Exercício 4**

**1- Encontre o(s) conjunto(s) com um suporte de 33% do conjunto de transações abaixo:**

T1 = tree, cup, paper  
T2 = book, tomato, pen  
T3 = pen, book, tree  
T4 = tomato, pen, cup  
T5 = tree, paper, book  
T6 = pen, book

**2- Dado o conjunto de transações:**

{1,3,5,7,8}
{3,4,5,7}
{2,3,5,6}
{2,4,5,8,9}
{2,3,5,11}
{1,2,3,7,9}

**Gere regras para os conjuntos relativos à linha e coluna, nessa ordem, para todas as células da matriz abaixo. Por exemplo,  $2 \rightarrow 3$ ,  $2 \rightarrow 5$  e  $2 \rightarrow 7$  para a primeira linha da matriz.**

	{2}	{3}	{5}	{7}
{2}	X			
{3}		X		
{5}			X	
{7}				X

**Informar todas as regras com confiança mínima de 73%.**

**3- Realize os seguintes passos:**

- Utilize alguma implementação do Apriori, por exemplo, a da biblioteca mlxtend (python) ou do pacote Arules (R).
- Utilize os dados sobre compras de supermercado “supermercado.csv”, disponível no moodle.
- Garanta que um conjunto de itens não contenha dois itens iguais.

- Execute o Apriori com as configurações: suporte mínimo = 0.001 e confiança = 0.25.
- Gere regras que contenham no consequente somente itens do tipo "bottled beer".
- Avaliação: o que você encontrou de interessante? Faça uma breve análise das regras e resultados encontrados.

**4 (opcional – pode ser um pontapé para a APS)** - Considere o *dataset* de *reviews* disponibilizado do Yelp, e em seguida:

- a) Preprocesse o dataset agrupando os tipos de locais visitados por usuário único que fizeram pelo menos dois *reviews*. Faça uma análise de padrões frequentes encontrados nos tipos de locais visitados. Você deve encontrar valores de parâmetros que forneçam resultados interessantes nesse *dataset*. Você encontrou algum padrão inesperado ou curioso?
- b) Faça uma análise de padrões frequentes considerando agora tipos de locais visitados e se um local é popular ou não. Note que aqui você não precisa agrupar informações por usuário único, você pode analisar somente as informações dos *businesses*. Você deve identificar parâmetros que forneçam resultados interessantes. Quais regras possuem como consequentes locais populares nesse contexto? Analise os resultados.