

# Mineração de Dados

## Trabalho 8

Arthur do Prado Labaki - 11821BCC017

23-05, 2023

GBC212

## Estudo do Conceito

Regras de Associação são um conceito importante na mineração de dados, que envolve a descoberta de relações interessantes e frequentes entre itens em conjuntos de dados. Essas regras descrevem padrões de associação entre diferentes itens em um conjunto de dados, permitindo que os usuários identifiquem correlações ou dependências entre eles.

As regras de associação são geralmente expressas na forma "se X, então Y", onde X e Y são conjuntos de itens. Essas regras são calculadas com base na frequência de ocorrência conjunta de diferentes itens em um conjunto de dados. Dois termos importantes associados às regras de associação são "suporte" e "confiança".

**Suporte:** Mede a frequência com que um conjunto de itens ocorre em um conjunto de dados. É calculado como a proporção de transações no conjunto de dados em que o conjunto de itens aparece juntos.

**Confiança:** Indica a probabilidade de que um item Y seja adquirido quando o item X é adquirido. É calculada como a proporção de transações que contêm o conjunto X e Y em relação ao número total de transações que contêm o conjunto X.

Além do suporte e da confiança, outros critérios podem ser utilizados para avaliar a qualidade das regras de associação, como o valor lift, que mede a importância da relação entre os itens em uma regra.

A descoberta de regras de associação envolve a análise de grandes conjuntos de dados em busca de relações interessantes e úteis. Essas regras podem ser aplicadas para tomada de decisões, planejamento estratégico, segmentação de mercado, recomendação de produtos, entre outros propósitos.

## Resolução do Exercício 1)

Dado o banco de dados abaixo:

Customer ID	Transaction ID	Items Bought
1	0001	$\{a, d, e\}$
1	0024	$\{a, b, c, e\}$
2	0012	$\{a, b, d, e\}$
2	0031	$\{a, c, d, e\}$
3	0015	$\{b, c, e\}$
3	0022	$\{b, d, e\}$
4	0029	$\{c, d\}$
4	0040	$\{a, b, c\}$
5	0033	$\{a, d, e\}$
5	0038	$\{a, b, e\}$

*Figura 1: Base de dados*

### Item A)

Para calcular o suporte, temos que realizar a seguinte formula:

$$\text{Suporte}(X) = (\text{Número de transações em que X ocorre}) / (\text{Número total de transações})$$

Com a nossa base de dados, temos que:

- $\text{Suporte}((e)) = 8/10 = 0.8$  ou 80%
- $\text{Suporte}((b, d)) = 2/10 = 0.2$  ou 20%
- $\text{Suporte}((b, d, e)) = 2/10 = 0.2$  ou 20%

### Item B)

Para calcular o suporte, temos que realizar a seguinte formula:

$$\text{Confiança}(A \rightarrow B) = \text{Suporte}(A \cup B) / \text{Suporte}(A)$$

Utilizando os resultados do exercício anterior, temos:

- $\text{Confiança}((b, d) \rightarrow (e)) = 0.2 / 0.2 = 1$  ou 100%

- $\text{Confi\~{a}n\~{c}a}((e) \rightarrow (b, d)) = 0.2 / 0.8 = 0.25$  ou 25%

A confian\~{c}a n\~{a}o \u00e9 sim\u00e9trica porque, se invertermos o antecedente e o conseq\u00fcente (b,d e e), a confian\~{c}a resultante pode ser diferente.

### Item C)

Recalculando o item A, considerando cada Customer ID. temos:

- $\text{Suporte}((e)) = 4/5 = 0.8$  ou 80%
- $\text{Suporte}((b, d)) = 2/5 = 0.4$  ou 40%
- $\text{Suporte}((b, d, e)) = 2/5 = 0.4$  ou 40%

### Item D)

Calculando a confian\~{c}a para esses novos valores, temos:

- $\text{Confi\~{a}n\~{c}a}((b, d) \rightarrow (e)) = 0.4 / 0.4 = 1$  ou 100%
- $\text{Confi\~{a}n\~{c}a}((e) \rightarrow (b, d)) = 0.4 / 0.8 = 0.5$  ou 50%

### Item E)

Supondo s1 e c1 regras de associa\~{c}o (suporte e confian\~{c}a respectivamente), para cada Transaction ID e s2 e c2 as mesmas regras, mas agora para cada Customer ID, podemos afirmar que embora possa haver alguma rela\~{c}o entre c1 e c2 ou s1 e s2, n\~{a}o h\~{a} uma rela\~{c}o direta ou fixa entre essas m\u00e9tricas. A rela\~{c}o depender\~{a} das caracter\u00edsticas espec\u00edficas dos dados transacionais e dos padr\u00f5es de compra dos clientes analisados.

## Item F)

Executando o algoritmo a priori com minsup de 50% e minconf 60%, temos que:

Contar a frequência de cada item na base de dados e eliminar os que não satisfazem o minsup:

a - 7 - 70%

b - 6 - 60%

c - 5 - 50%

d - 7 - 70%

e - 9 - 90%

Resultado: (a,b,c,d,e)

Gerar os conjuntos de itens candidatos de tamanho 2 e eliminar os que não satisfazem o minsup:

(a,b) - 4 - 40%

(a,c) - 3 - 30%

(a,d) - 6 - 60%

(a,e) - 7 - 70%

(b,c) - 3 - 30%

(b,d) - 3 - 30%

(b,e) - 6 - 60%

(c,d) - 4 - 40%

(c,e) - 4 - 40%

(d,e) - 7 - 70%

Resultado: ((a,d), (a,e), (b,e), (d,e))

Gerar os conjuntos de itens candidatos de tamanho 3 e eliminar os que não satisfazem o minsup:

(a,b,e) - 4 - 40%

(a,d,e) - 6 - 60%

(b,d,e) - 3 - 30%

Resultado: (a,d,e)

Não é possível gerar conjuntos de itens candidatos de tamanho maior que o maior conjunto de itens frequente encontrado, então agora é necessário gerar as regras de associação a partir dos conjuntos de itens frequentes, em que é preciso calcular a confiança de cada regra possível usando a fórmula:

$$\text{Confiança}(X \rightarrow Y) = \text{Suporte}(X \cup Y) / \text{Suporte}(X)$$

As regras que satisfazem o minconf de 60% são:

$a \rightarrow d$  (Confiança = 85.71%)

$a \rightarrow e$  (Confiança = 100%)

$b \rightarrow e$  (Confiança = 100%)

$d \rightarrow a$  (Confiança = 85.71%)

$d \rightarrow e$  (Confiança = 100%)

$e \rightarrow a$  (Confiança = 77.78%)

$e \rightarrow d$  (Confiança = 77.78%)

$a \rightarrow d,e$  (Confiança = 85.71%)

$d \rightarrow a,e$  (Confiança = 85.71%)

$e \rightarrow a,d$  (Confiança = 66.67%)