

Exercícios sobre Regras de Associação

Pós Graduação em Big Data

Prof. Fabrício Jailson Barth

1. Dado a lista de cinco transações que incluem os itens A, B, C e D:

A,B,C

A,C

B,C

A,D

A,C,D

Quais são os conjuntos de itens (*itemsets*) que satisfazem o suporte mínimo de 0,5? (Dica: um conjunto de itens pode incluir mais que um item)

2. Um estabelecimento de varejo tem uma base de dados que armazena 10.000 transações realizadas no último verão. Após analisar os dados, um time de cientistas de dados identificou os seguintes dados:

- $\{bateria\}$: foi identificado em 6.000 transações.
- $\{protetorsolar\}$: foi identificado em 5.000 transações.
- $\{sandalia\}$: foi identificado em 4.000 transações.
- $\{tigela\}$: foi identificado em 2.000 transações.
- $\{bateria, protetorsolar\}$: foi identificado em 1.500 transações.
- $\{bateria, sandalia\}$: foi identificado em .000 transações.
- $\{bateria, tigela\}$: foi identificado em 250 transações.
- $\{bateria, protetorsolar, sandalia\}$: foi identificado em 600 transações.

Responda as seguintes questões:

- Quais são os valores de suporte para os *itemsets* acima?
- Assumindo que o valor de suporte mínimo é 0,05, quais itens são considerados frequentes?
- Quais são os valores de confiança de $\{bateria\} \rightarrow \{protetorsolar\}$ e $\{bateria, protetorsolar\} \rightarrow \{sandalia\}$? Qual das duas regras é mais interessante?
- Liste todas as regras candidatas que podem ser formados a partir dos dados acima. Quais regras são consideradas interessantes levando-se em consideração uma confiança mínima de 0,25? Quais regras podem ser consideradas mais úteis (isto é, com menor probabilidade de ter sido gerada por um acaso)?