

A. Mineração de Dados em Larga Escala – 02.04.2025

Name:

Student Number:

Suppose there are 100 items, numbered 1 to 100, and also 100 baskets, also numbered 1 to 100. Item i is in basket b if and only if i divides b with no remainder. Thus, item 1 is in all the baskets, item 2 is in all fifty of the even-numbered baskets, and so on.

Basket 12 consists of items $\{1,2,3,4,6,12\}$, since these are all the integers that divide 12.

Answer the following questions:

- (a) If the support threshold is 5, which items are frequent?
- (b) If the support threshold is 5, which pairs of items are frequent?
- (c) What is the sum of the sizes of all the baskets?

What is the confidence of the following association rules?

- (a) $\{5,7\} \rightarrow 2$.
- (b) $\{2,3,4\} \rightarrow 5$.
- (c) What is the lift of the previous association rules?

Notes:

$$\text{Confidence}(A \rightarrow B) = \text{Support}(A \cup B) / \text{Support}(A)$$

$$\text{Lift}(A \rightarrow B) = \text{Confidence}(A \rightarrow B) / (\text{Support}(B) / N)$$

$$= [\text{Support}(A \cup B) \times N] / [\text{Support}(A) \times \text{Support}(B)]$$

Where:

Support(X) = number of baskets containing itemset X

N = total number of baskets