

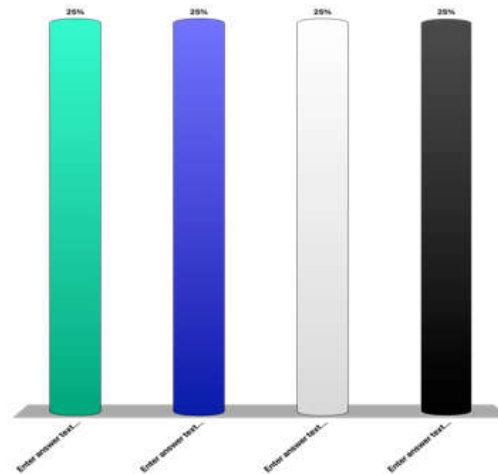
An insurance company wants to find typical, but unknown, causes for specific injuries from earlier cases. This likely is based on ...

- A. local rule discovery
- B. predictive modelling
- C. descriptive modelling
- D. exploratory data analysis



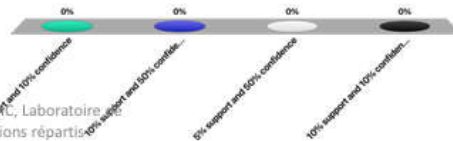
Let's assume that the transactions are stored in a relation $T(x, A1, \dots, A5)$, where x is the customer and each attribute $A1, \dots, A5$ can have 3 different values. How many different items exist after reduction to a single dimension?

- A. 5
- B. 243
- C. 125
- D. 15



10 itemsets out of 100 contain item A, of which 5 also contain B. The rule $A \rightarrow B$ has:

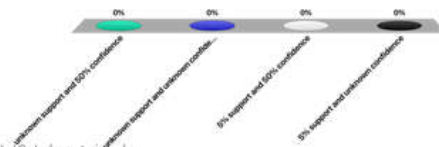
- A. 5% support and 10% confidence
- B. 10% support and 50% confidence
- C. 5% support and 50% confidence**
- D. 10% support and 10% confidence



© 2018, Karl Aberer, EPFL, Laboratoire
systèmes d'informations réparties

10 itemsets out of 100 contain item A, of which 5 also contain B. The rule $B \rightarrow A$ has:

- A. unknown support and 50% confidence
- B. unknown support and unknown confidence
- C. 5% support and 50% confidence
- D. 5% support and unknown confidence**



© 2018, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

Given the frequent 2-itemsets $\{1,2\}$, $\{1,4\}$, $\{2,3\}$ and $\{3,4\}$, how many 3-itemsets are generated and how many are pruned?

A. 2, 2

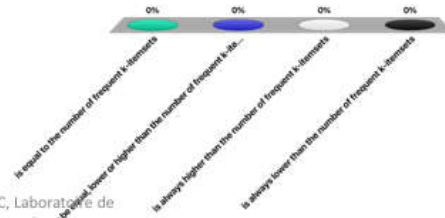
B. 1, 0

C. 1, 1

D. 2, 1

After the join step, the number of $k+1$ -itemsets ...

- A. is equal to the number of frequent k -itemsets
- B. can be equal, lower or higher than the number of frequent k -itemsets
- C. is always higher than the number of frequent k -itemsets
- D. is always lower than the number of frequent k -itemsets



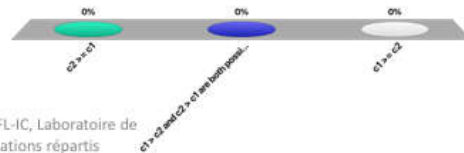
© 2018, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations répartis

If rule $\{A,B\} \rightarrow \{C\}$ has confidence c_1 and rule $\{A\} \rightarrow \{C\}$ has confidence c_2 , then ...

A. $c_2 \geq c_1$

B. $c_1 > c_2$ and $c_2 > c_1$
are both possible

C. $c_1 \geq c_2$



© 2018, Karl Aberer, EPFL-IC, Laboratoire de
systèmes d'informations répartis

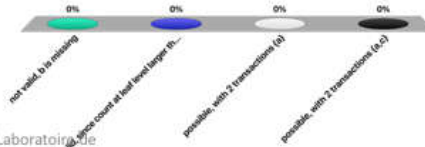
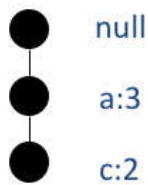
A false negative in sampling can only occur
for itemset with support smaller than ...

- A. the threshold s
- B. $p \cdot s$
- C. $p \cdot m$
- D. None of the above



The FP tree below is ...

- A. not valid, b is missing
- B. not valid, since count at leaf level larger than 1
- C. possible, with 2 transactions {a}
- D. possible, with 2 transactions {a,c}



© 2018, Karl Aberer, EPFL-IC, Laboratoire de systèmes d'informations réparties