

Aufgabe 1

1

Die Confidence gibt an, welcher Anteil der Tupel, welche die Precondition erfüllen auch die Conclusion erfüllen. Also im Endeffekt wie sicher man sich sein kann, dass die Klassifikationsregel auch gilt.

Der Support gibt an, welchen Anteil der Tupel die die Precondition und die Conclusion erfüllen an allen Tupeln der Tabelle haben. Also im Endeffekt wie stark die Klassifikationsregel von den Daten aus der Datenbank gestützt wird.

2

1. $confidence = 4/4 = 1$

$support = 4/8 = 0.5$

2. $confidence = 2/2 = 1$ $support = 2/8 = 0.25$

3. $confidence = 1/1 = 1$

$support = 1/8$

4. $confidence = 1/1 = 1$ $support = 1/8$

3

1

Der A-Priori-Algorithmus findet in einem Datenset alle häufigen Itemsets, welche häufiger als ein gewisses, gegebenes Threshold auftauchen.

2

Alle Teilmengen eines häufigen Itemsets müssen häufig sein (Apriori-Eigenschaft).
Wenn ein Itemset nicht häufig ist, sind auch alle seine Obermengen nicht häufig.

4

Bei der einen Iteration hat man ein Itemset I_k , zu welchen ein weiteres Element, was noch nicht in der Menge I_k vorhanden ist hinzugefügt wird. Danach wird für die neu entstandenen Itemsets die Häufigkeit geprüft und damit bestimmt, welche weiter betrachtet werden.

Bei der anderen Iteration werden die Itemsets bestimmt, indem zwei Mengen I_k und J_k aus den Itemsets des Schrittes k betrachtet werden, welche nur ein gemeinsames Element enthalten. Von diesen wird dann die Differenz gebildet. Diese ist dann auch ein Itemset, für welche auch die Häufigkeit bestimmt wird. Ist diese häufig, so lassen sich die jeweiligen Element aus dem Schnitt nun zu den Mengen I_k und J_k hinzufügen, welche dann wieder häufige Itemsets ergeben