

Aufgabe 3

1

Schritt k=1

Itemset	Support
Windeln	3/8
Bier	5/8
Chips	6/8
TV-Zeitschrift	4/8
Zahnpasta	2/8

Schritt k=2

Itemset	Support	
Windeln, Bier	3/8	max.
Windeln, Chips	1/8	fällt raus
Windeln, TV-Zeitschrift	0/8	fällt raus
Windeln, Zahnpasta	1/8	fällt raus
Bier, Chips	3/8	
Bier, TV-Zeitschrift	2/8	
Bier, Zahnpasta	1/8	fällt raus
Chips, TV-Zeitschrift	3/8	
Chips, Zahnpasta	1/8	fällt raus
TV-Zeitschrift, Zahnpasta	0/8	fällt raus

Schritt k=3

Itemset	Support
Bier, Chips, TV-Zeitschrift	2/8

Für Schritt k=4 lässt sich kein weiteres Itemset konstruieren, da es kein Itemset gäbe, welches keine Subsets enthalten würden, welche nicht den Mindestsupport erfüllen würden

Somit ergeben sich folgende Itemsets mit Mindestsupport von 25%: {{Windeln}, {Bier}, {Chips}, {TV-Zeitschrift}, {Zahnpasta}, {Windeln, Bier}, {Bier, Chips}, {Bier, TV-Zeitschrift}, {Chips,

2

Man könnte die Ergebnisse so interpretieren, dass die in den Itemsets zusammenstehenden Items sind Dinge, welche häufig zusammen gekauft werden.

Bei einem größeren Datensatz könnten mehr Zusammenhänge auftreten und vielleicht auch größere Zusammenhänge. Ebenso wäre es dort wahrscheinlich sinnvoll keine untere Schranke zu wählen, welche ein wenig kleiner als 25% ist

3

Der Lift gibt an, wie stark X zu Y korreliert. Würde Y keine Abhängigkeit von X haben, so würde in 50% der Fälle Y auftreten, wenn X auftritt. In diesem Fall wäre der Lift dann 1. Würde Y häufiger als 50% auftreten, so würde X Y implizieren und der Lift wäre größer 1. Würde Y seltener als 50% auftreten, so würde X $\neg Y$ implizieren und der Lift wäre kleiner 1.

transaktionsID	milch	brot	butter	bier	windeln	eier	frucht
1	1	1	0	0	0	0	1
2	0	0	1	0	0	1	1
3	0	0	0	1	1	0	0
4	1	1	1	0	0	1	1
5	0	1	0	0	0	0	0

$$lift(milch, brot \Rightarrow butter) = \frac{0.2}{0.4 \cdot 0.4} = 1.25$$

{milch,brot} und {butter} sind also leicht positiv korreliert