

**Übung zur Vorlesung  
 „Einsatz und Realisierung von Datenbanksystemen“  
 im Sommersemester 2007**

Richard Kuntschke (richard.kuntschke@in.tum.de)

**Blatt 8**

**Aufgabe 1**

Entwerfen Sie einen Algorithmus, um den Klassifikationsbaum, wie er in Abbildung 1 exemplarisch gezeigt ist, automatisch zu ermitteln. Dazu gehe man top-down vor: Man wählt ein Vorhersage-Attribut und einen Wert dafür, der die Eingabemenge in zwei „sinnvolle“ Partitionen zerlegt. Man partitioniere die Eingabemenge demgemäß und gehe rekursiv den Baum hinab bis man Partitionen erhält, deren Elemente das gleiche abhängige Attribut haben.

Aufgabennummer im Buch: 17.4

**Aufgabe 2**

Beschreiben Sie in eigenen Worten die Auswertungsstrategie für die Joinberechnung des Bitmap-Star-Joins, wie sie in Abbildung 2 dargestellt ist.

**Aufgabe 3**

Beweisen Sie Folgendes: Wenn man für einen *frequent itemset*  $F = L \cup R = L^+ \cup R^-$  zwei Assoziationsregeln  $L \Rightarrow R$  und  $L^+ \Rightarrow R^-$  mit  $L \subseteq L^+$  und  $R^- \subseteq R$  betrachtet, so gilt immer:

$$confidence(L^+ \Rightarrow R^-) \geq confidence(L \Rightarrow R)$$

Aufgabennummer im Buch: 17.5

Schadenshöhe			
wiealt	Geschlecht	Autotyp	Schäden
45	w	Van	gering
18	w	Coupé	gering
22	w	Van	gering
38	w	Coupé	gering
19	m	Coupé	hoch
24	m	Van	hoch
40	m	Coupé	hoch
40	m	Van	gering
⋮	⋮	⋮	⋮

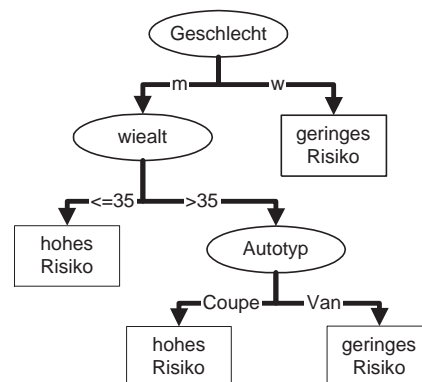


Abbildung 1: Klassifikationsschema für Haftpflicht-Risikoabschätzung

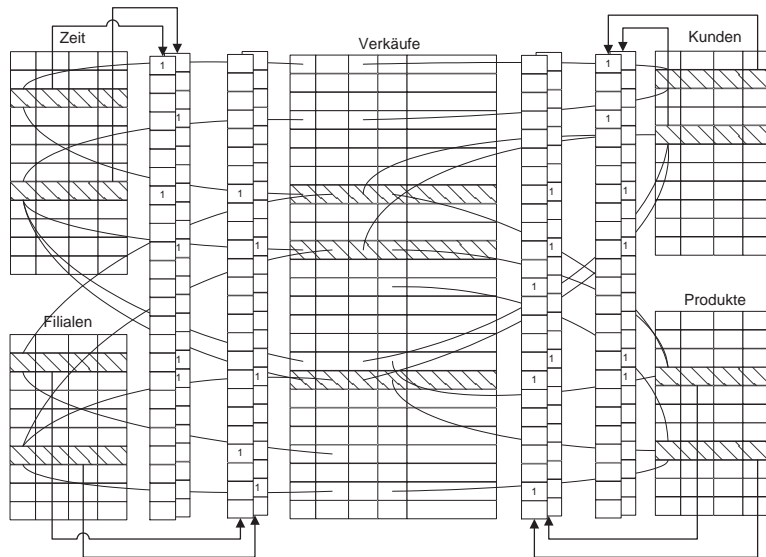


Abbildung 2: Auswertungsstrategie für die Joinberechnung des Bitmap-Star-Joins

#### Aufgabe 4

Zeigen Sie die vierte Phase des à priori-Algorithmus für unser Beispiel in Abbildung 3. Damit eine Menge von Produkten ein *frequent itemset* ist, muss sie in mindestens  $3/5$  aller Verkäufe enthalten sein, d.h.  $minsupp = 3/5$ .

Aufgabennummer im Buch: 17.6

VerkaufsTransaktionen	
TransID	Produkt
111	Drucker
111	Papier
111	PC
111	Toner
222	PC
222	Scanner
333	Drucker
333	Papier
333	Toner
444	Drucker
444	PC
555	Drucker
555	Papier
555	PC
555	Scanner
555	Toner

Zwischenergebnisse	
FI-Kandidat	Anzahl
{Drucker}	4
{Papier}	3
{PC}	4
{Scanner}	2
{Toner}	3
{Drucker, Papier}	3
{Drucker, PC}	3
{Drucker, Scanner}	
{Drucker, Toner}	3
{Papier, PC}	2
{Papier, Scanner}	
{Papier, Toner}	3
{PC, Scanner}	
{PC, Toner}	2
{Scanner, Toner}	
{Drucker, Papier, PC}	
{Drucker, Papier, Toner}	3
{Drucker, PC, Toner}	
{Papier, PC, Toner}	

Abbildung 3: Datenbank mit Verkaufstransaktionen (links) und Zwischenergebnisse des à priori-Algorithmus (rechts)