

Information Mining - Wintersemester 2014

Dipl.-Inform. Vu Tran, LF 139

vtran@is.inf.uni-due.de

Übungsblatt 4

Aufgabe 8: Attributauswahl für Spam-Klassifikation

Bei der Klassifikation von Spam-E-Mails geht es darum, eingehende E-Mails entweder in die Klasse **Spam** oder in die Klasse **Kein_Spam** einzuordnen.

Welche Informationen einer E-Mail könnte man als Attribute zur Klassifikation von Spam-E-Mails verwenden? Nenne anhand von Beispielen mindestens fünf verschiedene Attribute oder Attributgruppen.

Aufgabe 9: Assoziationsregeln

Die nachfolgende Tabelle gibt Verkaufsdaten eines (sehr kleinen) Supermarkts an. Die erste Spalte (Nr.) dient der einfacheren Identifikation der Kunden und ist *kein* reguläres Attribut. Alle weiteren Spalten geben an, ob der Kunde das jeweils angegebene Produkt gekauft hat (✓) oder nicht (×).

Nr.	Bier	Kartoffelchips	Klopapier	Zahnbürste
1	✓	×	✓	×
2	✓	✓	✓	×
3	×	✓	✓	✓
4	×	✓	✓	✓
5	✓	✓	×	×
6	✓	✓	✓	✓
7	×	×	✓	✓
8	×	×	×	✓
9	×	×	✓	✓

- (a) Gib die Unterstützung und die Konfidenz der nachfolgenden Assoziationsregeln an:
- (i) Kartoffelchips → Bier
 - (ii) Klopapier und Bier → Zahnbürste
- (b) Der Geschäftsführer des Supermarkts schaut sich die erste Assoziationsregel an und sagt: „Es gibt also einen kausalen Zusammenhang zwischen Kunden, die Kartoffelchips kaufen, und Kunden, die Bier kaufen.“ Hat er recht? Begründe Deine Antwort.