

**Information Mining - Wintersemester 2014**

Dipl.-Inform. Vu Tran, LF 139

vtran@is.inf.uni-due.de

**Übungsblatt 6****Aufgabe 12: Lernen von Assoziationsregeln mit *RapidMiner***

Lerne mit *RapidMiner* aus fiktiven Kundendaten<sup>1</sup> eines Unternehmens Assoziationsregeln. Der Prozess soll dazu die folgenden Schritte durchführen:

1. Einlesen der Daten
2. Aufbereitung der Daten
  - a. Diskretisierung von numerischen Attributen
  - b. Umwandlung aller Attribute in binominale Attribute
3. Berechnung von häufigen Item-Sets
4. Berechnung von Assoziationsregeln

Hinweis: Für die letzten beiden Schritte werden die Operatoren **FP-Growth** und **Create Association Rules** benötigt.

**Aufgabe 13: Nützlichkeit von Attributen: Informationsgewinn**

Die nachfolgende Tabelle zeigt die Daten einiger Patienten, bei denen ein Verdacht auf Schweinegrippe bestand (markiert durch „✓“). Das Attribut **Mexiko** gibt an, ob der Patient in letzter Zeit in Mexiko war (✓) oder nicht (×). Das Attribut **Alter** gibt die Altersgruppe des Patienten an (**jung**, **mittelalt** oder **alt**). Außerdem hat jeder Patient eine **Lieblingsfarbe** (**rot** oder **blau**).

	<b>Mexiko-Aufenthalt</b>	<b>Alter</b>	<b>Lieblingsfarbe</b>	<b>Schweinegrippe</b>
1	✓	jung	blau	✓
2	×	jung	blau	×
3	×	mittelalt	rot	✓
4	✓	mittelalt	rot	✓
5	×	alt	blau	×
6	×	alt	rot	×
7	✓	jung	rot	×
8	✓	mittelalt	blau	✓
9	✓	jung	blau	✓
10	✓	alt	rot	✓

<sup>1</sup>[http://www.is.inf.uni-due.de/courses/im\\_ws14/uebung/data\\_a12.csv](http://www.is.inf.uni-due.de/courses/im_ws14/uebung/data_a12.csv)

Berechne jeweils für die Attribute **Mexiko**, **Alter** und **Lieblingsfarbe** den Informationsgewinn (definiert wie in der Vorlesung), wenn anhand dieser die Instanzen aufgeteilt werden.