



## Projektaufgabe: Intelligente Vergabe von Gutscheinen im Online-Handel

Quelle: Data Mining CUP 2010

### Ausgangssituation

Viele Kunden nehmen in einem Online Shop nur einmal eine Bestellung vor. Dafür, dass sie später nicht erneut bestellen, gibt es vielfältige Gründe. Durch geeignete Kundenbindungsmaßnahmen versuchen Onlinehändler dem entgegen zu wirken. Hier kann z.B. der Einsatz von Gutscheinen, die in einem gewissen zeitlichen Anschluss an eine Bestellung versandt werden, ein probates Mittel sein, Kunden zu einem Folgekauf zu animieren. Allerdings ist das Versenden von Gutscheinen an alle Kunden aus wirtschaftlicher Sicht keine gute Lösung, da einige Kunden auch einen zweiten Kauf ohne Zusendung eines solchen Gutscheins tätigen würden.

### Szenario

Neben klassischen Medien wie Büchern und CDs bietet ein Händler in seinem Internetshop eine Vielzahl an Hörbüchern, Downloads, eBooks und eBook Readern an. Aus der Gesamtheit der Kunden sollen nun jene Kunden identifiziert werden, die mit einem Gutschein zu einem Folgekauf animiert werden können. In diesem Szenario wird davon ausgegangen, dass ein Folgekauf innerhalb der ersten 90 Tage nach der Erstbestellung angestrebt wird.

Anhand der vorhandenen Merkmale der Erstbestellung eines Kunden wie z. B. Bestellmenge pro Warengruppe, Anrede und Sendungsgewicht ist eine Entscheidung zu treffen ob dem Kunden ein Gutschein in Höhe von 5 Euro zugesandt werden soll. Es sollen nur jene Kunden einen Gutschein erhalten, welche sich nicht von selbst für eine erneute Bestellung entschieden hätten.

Wird ein Gutschein an einen Erstbesteller gesendet, der jedoch sowieso wieder dort eingekauft hätte, entstehen dem Händler 5 Euro Verlust. Wird jedoch einem Kunden ein Gutschein zugesendet, der ohne diesen Gutschein nicht erneut gekauft hätte, generiert das beim Händler im Durchschnitt einen Gewinn von 1,50 Euro. Daraus ergibt sich folgende Kostenmatrix:

		Tatsächlich	
		Kein Wiederkäufer (0)	Wiederkäufer (1)
Vorhergesagt	Kein Gutschein (0)	0	0
	Gutschein (1)	1,5	-5

Das Ziel des Händlers ist nun natürlich, seinen Gewinn zu maximieren. Dafür ist ein Data Mining Modell auf 80% der vorliegenden Daten zu erstellen, das eine möglichst genaue Vorhersage ermöglicht, welchen Kunden ein Gutschein zugesendet werden soll und welchen nicht.

Die für die Erstellung eines Data Mining Modells zur Verfügung stehenden Daten befinden sich in der Datei training.csv. Sofern für ein Attribut keine Werte angegeben wurden, sind diese Werte nicht

bekannt, sie sind in diesem Fall als fehlende Werte zu behandeln. Die Beschreibung der einzelnen Attribute befindet sich in features.pdf

### **Projektbericht**

Schreiben Sie einen kurzen Bericht über die von Ihnen vorgenommenen Schritte zur Datenvorbereitung sowie welche Data Mining Methoden Sie angewandt haben.

Beurteilen Sie Ihr Ergebnis unter Anwendung der genannten Kostenmatrix auf die in der Datei test.csv zur Verfügung gestellten Testdaten. Welche finanzielle Verbesserung erreicht der Online-Händler, wenn er Ihren Klassifikator einsetzt?

Der Projektbericht ist spätestens am **27. Januar 2019** abzugeben.