

Übungsblatt 4

Apriori 15 Punkte

Data Mining

Wintersemester 2016/17

Abgabe: 16.12.2016 9:45 Uhr

Aufgabe 1 Maße für Assoziationsregeln 5 Punkte

Ein Online-Buchgeschäft möchte eine Datenbank mit 10.000 Kunden (= Transaktionen) analysieren, die jeweils eines oder mehrere von 500 verschiedenen Büchern gekauft haben. Zur Entdeckung von Assoziationsregeln wird der Algorithmus Apriori mit einem Minimum Support von 3% und einer minimalen Konfidenz von 75% verwendet

Aufgabe 1.1 Support und Konfidenz, 4 Punkte

Es wird festgestellt, dass die beiden häufigsten Verkäufe „Der Prozeß“ von Franz Kafka (Buch1) und „Lebensansichten des Katers Murr“ von E.T.A. Hoffmann (Buch2) sind. Buch1 wurde von 6.000 Kunden und Buch2 von 8.000 Kunden gekauft. 4.000 Kunden kauften beide Bücher. Welche der beiden Assoziationsregeln findet sich im Output des Assoziationsregel-Lerners?

- a) Buch1 \rightarrow Buch2
- b) Buch2 \rightarrow Buch1
- c) beide
- d) keine von beiden

Geben Sie Support und Konfidenz für beide Regeln an.

Aufgabe 1.2 Lift, 1 Punkt

Wenn man annimmt, dass alle Kunden, die beide Bücher gekauft haben, zuerst Buch1 und später Buch2 gekauft haben: Wie interpretieren Sie den Einfluss des Kaufs von Buch1 auf den Kauf von Buch2? Verwenden Sie das Maß Lift, um den Einfluss zu erklären.

Aufgabe 2 Lernen von Assoziationsregeln mit R 10 Punkte

Aufgabe 2.1 Supermarket Transaction Baskets, 10 Punkte

Die Datei `supermarket_transaction_baskets.csv` enthält Daten von Kassensbons eines Supermarkts. Erstellen Sie in R ein Model mit Aprior, das aus diesen Daten Assoziationsregeln lernt. Beantworten Sie die Fragen, die Sie in dem R-File finden.

Erläutern Sie Konfidenz, Support und Lift anhand einer selbst gewählten Regel aus dieser Aufgabe.

Aufgabe 3 Projekt

Bis Semesterende sollen Sie in Gruppen von vier bis fünf Studierenden ein reales Data Mining Projekt von dem Datenverständnis, Datenaufbereitung, Modell Lernen und Bewertung durchführen. Die Projektergebnisse werden von den Gruppen am 20.01.2017 nachmittags vorgestellt.

Sie finden die Beschreibung und die Daten für das Projekt in dem Git-Ordner Project.

Churn Management

Der deutsche Markt der Energieversorger ist seit 1999 dereguliert. Das heißt, dass seit diesem Zeitpunkt die Stromkunden in Deutschland frei ihren Anbieter wechseln können – vorher war dies nicht möglich.

Das ist eine neue Situation für Anbieter und Kunden. Der Kunde wird oft einen Wechsel vollziehen, um durch günstigere Preise an den Stromkosten zu sparen.

Die Anbieter ihrerseits stehen vor der neuen Herausforderung, ihre Kunden mit guten und interessanten Angeboten binden zu müssen. Sie versuchen dabei, ihre Kunden individuell und bedürfnisgerecht anzusprechen. Vor diesem Hintergrund wird Data Mining u. a. zur Berechnung von Kundenprofilen eingesetzt.

Ein Energieversorger möchte mit zielgerichteten Kundenbindungsmaßnahmen seine Kunden halten. Dazu möchte das Unternehmen potenziellen Kündigern einen günstigeren Tarif (Tarif Rabatt) anbieten. Da dieser Tarif für den Energieversorger einen geringeren Gewinn bedeutet, will er die voraussichtlich treuen Kunden von diesem Angebot nicht informieren, um nicht zu vielen von diesen Kunden den Tarif Rabatt gewähren zu müssen.

Ziel des Einsatzes von Data Mining ist es hierbei, die potenziellen Kündiger von den treuen Kunden zu unterscheiden. Potenziellen Kündigern wird der Tarif Rabatt angeboten, um sie damit zu binden. Insgesamt ist es für den Energieversorger deutlich günstiger, einen potenziellen Kündiger weiterhin im Tarif Rabatt zu versorgen als ihn ganz zu verlieren.