

## Automatische Kategorisierung von Antworten bei Kundenbefragung

Entwicklung eines Text Mining Algorithmus für die OBI Group

### Herausforderung

Die OBI Group ist Marktführer der deutschen Bau- und Heimwerkermarktbranche und gehört mit mehr als 650 Märkten europaweit zu den bekanntesten Baumärkten. Um das Angebot und den Kundenservice kontinuierlich zu verbessern, bietet OBI seinen Kunden an, ihr Feedback mittels eines Online-Fragebogens mit dem Unternehmen zu teilen. Vor allem die offenen Fragen liefern dabei wertvolle und oft ungeahnte Erkenntnisse. Um die Antworten auf diese Fragen auswerten zu können, müssen sie jedoch vor der Auswertung zunächst manuell kategorisiert werden.

### Ziel

Bei mehr als 40.000 Umfrageteilnehmern gestaltet sich eine händische Zuordnung der Antworten zu passenden, vordefinierten Kategorien wie Produktauswahl, Lage oder Preis-Leistungsverhältnis sehr zeitaufwendig. Aus diesem Grund suchte OBI nach einer Lösung zur automatischen Kategorisierung von offenen Antworten.

### Lösung

In einem ersten Schritt erfolgte eine Aufbereitung der von OBI übermittelten Antworten durch eoda. Zur Optimierung der Prognosegüte wurden ähnliche Terme mithilfe einer Wortstammfunktion verdichtet, orthographisch falsche Begriffe einem passenden Term zugeordnet und sinnfreie Worte exkludiert. Zudem wurden relevante Terme, die häufig in Kombination miteinander auftreten, unter einem Meta-Term zusammengefasst und so in das Modell implementiert.

Im Anschluss an die Datenaufbereitung wurde mit dem Data Mining Algorithmus *Gradient Boosting Machine* ein Modell gebildet, welches individuelles Kundenfeedback automatisiert der passenden Antwortkategorie zuordnet. Dieses Modell wurde auf die offenen Antworten angewendet und die Ergebnisse mit der manuellen Zuordnung von OBI verglichen.

### Ergebnis

Das von eoda entwickelte Modell erzielt eine treffsichere Klassifikation der Antworten. Mit seiner hohen Genauigkeit stellt das Modell eine wertvolle Unterstützung zur effizienten Fragebogenauswertung dar. Antworten auf offene Fragen müssen somit nicht länger manuell kategorisiert werden, sondern können automatisch den entsprechenden Antwortklassen zugeordnet werden. Dies führt zu einer deutlichen Verringerung des Arbeits- und Zeitaufwands bei hoher Qualität.