

Bestandes- und Profitabilitätsanalyse im Motorfahrzeugbereich

Eine Versicherung gibt die Möglichkeit, sich mit Hilfe von Prämien gegen Schäden finanziell abzusichern. Die Prämien werden immer kundenspezifischer bestimmt. Dazu werden Prognosemodelle für die Schadenshäufigkeit und die Schadenshöhe benötigt, die auf Kundenmerkmalen beruhen. Bei einer Versicherungsfirma sind dabei ältere Prognosemodelle für Lieferwagen vorhanden. Es stellt sich die Frage, ob überhaupt ein neues Modell erstellt werden muss, oder ob das alte Modell ausreicht. Zusätzlich soll überprüft werden, ob das Merkmal "Region" einen Einfluss auf die Schadenshäufigkeit hat.

Beim zur Verfügung gestellten Datensatz handelte es sich um die Verträge von Kleintransportern. Die Daten stammten aus den Jahren 2016 und 2017. Insgesamt enthielt der Datensatz 524'242 Beobachtungen und 298 Variablen. Der Fokus lag auf der Kollisionskasko, welche aus 91'588 Beobachtungen besteht. Bei der Datenaufbereitung ging es unter anderem darum, aus fachlicher Sicht relevante Merkmale (auch Variablen genannt) zu identifizieren und geeignete Merkmale zu konstruieren. So wurden raumgliedernde Bfs-Merkmale hinzugefügt, weil die vorhandenen geografischen Merkmale den "Stadt-Land"-Effekt ungenügend oder zu detailliert erfassten.

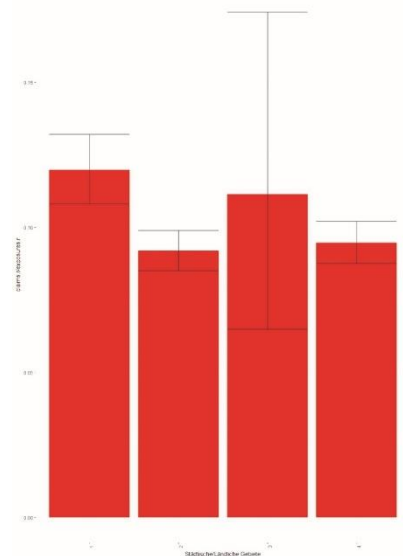
Diese Arbeit fokussiert sich auf ein Vorhersagemodell für die Schadensanzahl pro Vertrag und Zeitperiode. Innerhalb einer Zeitperiode haben alle Merkmale einen konstanten Wert. Das Vorhersagemodell beruht auf einer entsprechend aufgesetzten Poisson-Regression. Die für das Vorhersagemodell relevanten Merkmale wurden über ein schrittweises Selektionsverfahren ausgewählt. Zwei raumgliedernde Merkmale erwiesen sich als relevant, die "Gemeindetypologie 2012" sowie "Städtische/Ländliche Gebiete 2000".

Einige Merkmale, welche im alten Vorhersagemodell wichtig sind, sind beim neuen Modell nicht mehr statistisch signifikant. Dafür enthält es 11 zusätzliche Merkmale. Eine out-of-sample Validierung war leider nicht machbar. Einerseits ist die Zeitperiode von zwei Jahren zu kurz, um weiter unterteilt zu werden, andererseits hat es zu wenig Schadensfälle, um die Daten in eine Lern- und Teststichprobe zu unterteilen.



Diplomierende
Lukas Schneider
Manuel Walser

Dozierende
Andreas Ruckstuhl
Yves-Laurent Grize



Die Grafik zeigt die Schadenshäufigkeit, aufgeteilt in die vier Gruppen des raumgliedernden Merkmals „Städtische/Ländliche Gebiete“. Die schwarzen Linien zeigen die Konfidenzintervalle an. Wie man sieht, kann man den zweiten und vierten Wert vom ersten unterscheiden. Daher sind diese statistisch signifikant verschieden. Der dritte Wert ist zu keiner anderen Gruppe statistisch signifikant verschieden, da sein Konfidenzintervall zu weit ist.