

VERSICHERER DER ZUKUNFT – MIT DATA SCIENCE ZUM ERFOLG

Neue Möglichkeiten der Risikobewertung – Artikel 4

Im Rahmen der 2016 in Kraft getretenen Solvency-II-Richtlinien werden spezielle Anforderungen an das Risikomanagement gestellt. Der Risikokontrollprozess sowie die Risikoberichterstattung sind zentrale Elemente des internen Steuerungs- und Kontrollsystems (IKS). Im Rahmen des Risikokontrollprozesses muss sich der Versicherer mit den Themen Risikoidentifikation, Risikoanalyse und -bewertung, Risikosteuerung sowie Risikoüberwachung auseinandersetzen.

Um neue Risiken zu identifizieren sowie bewerten zu können, kommt Data Science eine immer größere Bedeutung zu. In unserem vorangegangenen Artikel der Serie Versicherer der Zukunft – mit Data Science zum Erfolg haben wir die Anwendung von Data Science im Bereich Marketing und Vertrieb näher betrachtet und explizite Anwendungsbeispiele geliefert. Die dort erwähnten Algorithmen und Modelle können auch im Rahmen des Risikokontrollprozesses zum Einsatz kommen, welcher die folgenden Risikokategorien berücksichtigt:¹

- Versicherungstechnisches Risiko
- Marktrisiko
- Kreditrisiko (einschließlich Länderrisiko)
- Operationelles Risiko
- Liquiditätsrisiko
- Konzentrationsrisiko
- Strategisches Risiko
- Reputationsrisiko

Sei es Betrugsrisiko oder das Risiko von Naturkatastrophen, Data Science ermöglicht es, durch die Auswertung von internen sowie externen Daten Risiken besser zu bewerten und neue Risiken zu identifizieren.

Die alltäglich gesammelten Daten bilden bereits eine fundierte Grundlage für Analysen. Im Rahmen von Solvency II sind diese Analysen als Teil des ORSA (Own Risk and Solvency Assessment) verpflichtend. ORSA beschreibt „die Gesamtheit der Prozesse und Verfahren, die eingesetzt werden, um kurz- und langfristige Risiken, denen das Versicherungsunternehmen ausgesetzt ist bzw. ausgesetzt sein könnte, zu identifizieren, zu bewerten, zu kontrollieren, handzuhaben oder zu berichten, und die der Ermittlung der Eigenmittel, die den gesamten Solvabilitätsbedarf sicherstellen sollen, dienen“.²

Zusätzlich können externe Datenquellen herangezogen werden, wodurch der Informationsgehalt weiter gesteigert werden kann. Mit der Verwendung einer Analytics Toolbox können durch die Anwendung von mathematischen sowie statistischen Algorithmen die gespeicherten Daten verarbeitet und verschiedene Ergebnistypen abgeleitet werden (Abb. 1). Diese umfassen neue, bisher unbekannte Risiken, Risikobewertungen oder Key Risk Indicators (KRI).

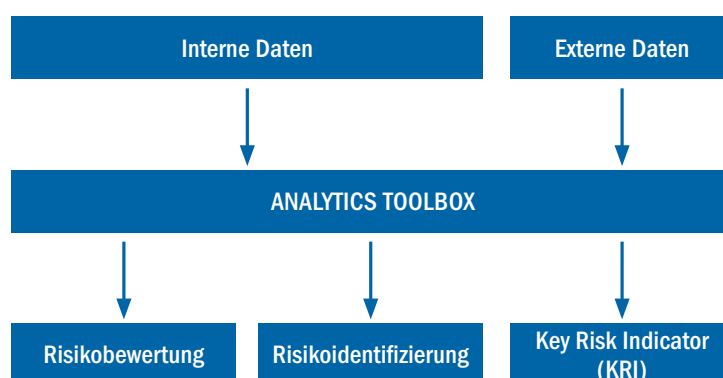


Abbildung 1: Exemplarisches Risikoanalyse-Framework

¹ Rundschreiben 3/2009: Aufsichtsrechtliche Mindestanforderungen an das Risikomanagement (MaRisk VA).

² CEIOPS Issues Paper (IGSRR-09/08): Own Risk and Solvency Assessment (ORSA), 27.05.2008.

Die Tools sowie Algorithmen, welche für die Analysen genutzt werden, sind oftmals kostenfrei verfügbar und können schnell eingesetzt werden. Abhängig vom jeweiligen Analyseziel werden verschiedene Algorithmen verwendet. In unserem Artikel Zielgerichtetes Marketing und Anwendungen im Vertrieb haben wir bereits einige Algorithmen näher besprochen, welche unter anderem für Machine Learning verwendet werden können. Machine Learning ermöglicht Maschinen das Lernen von Gesetzmäßigkeiten anhand von Daten ohne explizite Programmierung. Grundsätzlich werden fünf verschiedene Lerntypen unterschieden. Die zwei am häufigsten zum Einsatz kommenden Lerntypen sind dabei:

Supervised Learning beschreibt das Lernen einer Funktion, welche einen Eingangswert auf eine Zielgröße abbildet, wobei die assoziierte Zielgröße bekannt ist.

Unsupervised Learning beschreibt das Lernen einer Funktion, welche einen Eingangswert auf eine Zielgröße abbildet, wobei die assoziierte Zielgröße nicht bekannt ist.

Es gibt darüber hinaus noch drei weitere Lerntypen: Reinforcement Learning, Active Learning und Semi-Supervised Learning.

Ähnlich wie die Lerntypen können auch die Lernziele in verschiedene Kategorien eingeteilt werden. Hierzu zählen:

Optimierung: Minimierung einer Nutzenfunktion unter Nebenbedingungen

Klassierung: Einordnung von Beobachtungen basierend auf observierten Variablen

Zusammenhangsanalyse: Analyse von Zusammenhängen bei Beobachtungen und Variablen

Vorhersagen: Vorhersage von Szenarien basierend auf historischen Ereignissen

Clustering: Identifizierung von Clustern und deren Grenzen in Daten

Komplexitätsreduktion: Priorisierung von beobachteten Variablen

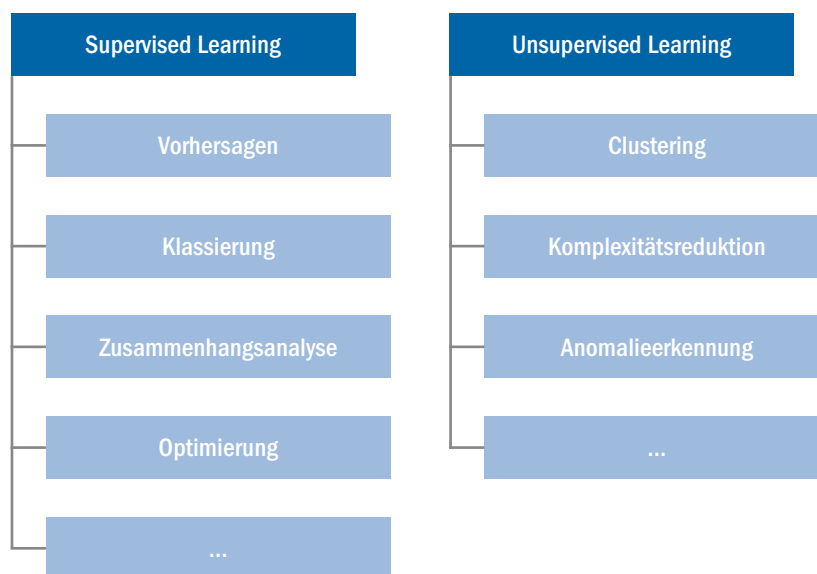


Abbildung 2: Zuordnung der Lernziele zu den Lerntypen Supervised und Unsupervised Learning

Möchte man Machine Learning einsetzen, so kommt es häufig zu der Situation, dass mehrere Algorithmen für ein und dasselbe Lernziel verwendet werden können. Um den für die Situation optimalen Algorithmus zu identifizieren, gilt es dann, die Ergebnisse

gegenüberzustellen und den Algorithmus mit dem besten Resultat auszuwählen. Im Folgenden möchten wir auf zwei Beispiele eingehen, wie man Data Science in der Risikobewertung einsetzen kann.

Schadenvorhersage

Im Rahmen eines Projekts mit einem deutschen KFZ-Versicherer wurden Möglichkeiten evaluiert, um die aktuell verwendeten Modelle zur Schadenvorhersage zu verbessern. Der Versicherer hatte bereits ein fortgeschrittenes Modell im Einsatz, um die Schadenhöhe von KFZ-Unfällen vorherzusagen. Allein durch die Erweiterung der aktuell verwendeten Datenbasis um externe Daten konnte durch den daraus resultierenden höheren Informationsgehalt die Vorhersagegenauigkeit verbessert werden. Im Weiteren konnte durch einen Wechsel des aktuellen Modells

sowie den Einsatz von Gradient Boosting und eines Random Forest eine weitere Verbesserung erzielt werden. Zusammenfassend konnte der Versicherer seine Vorhersagegenauigkeit im Vergleich zu seinen zuvor im Einsatz befindlichen Modellen um 29 % verbessern.³ Die Verwendung eines verbesserten Modells kann unter anderem folgende Vorteile mit sich bringen:

- Niedrigeres Risikokapital
- Niedrige Policenpreise
- Genauere Risikoeinschätzung

Betrugserkennung

Durch Versicherungsbetrug entsteht nach aktuellen Informationen des Gesamtverbands deutscher Versicherer ein jährlicher wirtschaftlicher Schaden von vier bis fünf Milliarden Euro in den Bereichen Kraftfahrt-, Haftpflicht- und Sachversicherung.⁴ Im Schnitt weisen rund 10 % der gemeldeten Schäden Ungereimtheiten auf. Möglichkeiten, das durch Betrug entstehende operationelle Risiko zu minimieren, werden somit dringend benötigt. Die Identifizierung von betrügerischen Schadenmeldungen ist demnach zwingend notwendig. Eine Möglichkeit, um dies zu erreichen, ist der Einsatz von Machine Learning. Durch das Trainieren von Algorithmen unter Verwendung von historischen Daten können neue Schadenmeldungen analysiert und Betrugsfälle identifiziert werden. In diesem Zusammenhang kommen häufig Klassifizierungsalgorithmen, wie Gradient Boosting, Random Forests oder

Regressionsmodelle, zum Einsatz. Neben der reinen Klassifizierung der Meldungen als Betrug oder Nichtbetrug kann auch die Wahrscheinlichkeit bestimmt werden, mit der es sich bei dem gemeldeten Schaden um Versicherungsbetrug handelt. Ein Vorreiter auf diesem Gebiet ist die Lemonade Insurance Corporation, die erfolgreich 18 verschiedene Anti-Betrug-Algorithmen verwendet.⁵

Die Chancen im Riskmanagement, welche sich durch Data Science ergeben, sind vielfältig und erfolgsversprechend. Basis für die Analysen sind zum einen Unternehmensdaten, zum anderen die richtige IT-Infrastruktur. In unserem nächsten Artikel werden wir uns genauer diesen beiden Themen widmen und auf die Herausforderungen, die mit einer Implementierung verbunden sind, eingehen.

Zu den Autoren:



Alexander Riesner
Manager
Office Wien
Praterstraße 31
1020 Wien
E-Mail alexander.riesner@zeb.at



Tobias Holler
Analyst
Office München
Theresienhöhe 13a
80339 München
E-Mail tobias.holler@zeb.de

³ Verbesserung in Prozent bezogen auf den Mean Absolute Error (MAE) des Referenzmodells.

⁴ GDV-Pressemitteilung vom 4. Mai 2017: Fast jede zehnte Schadenmeldung mit Ungereimtheiten.

⁵ <https://www.lemonade.com/claims>