



Wie die Lebensversicherung Data Science nutzt

Seit jeher benötigen die Versicherer Daten über die zu versichernden Risiken, um diese zu messen und zu bewerten. Erst dadurch wird es möglich, einen angemessenen Preis für den Risikoschutz zu ermitteln. In der Lebensversicherung ist diese Aufgabe besonders komplex, da die Daten oftmals nicht aktuell und nicht umfangreich genug oder nicht vollständig digital verfügbar sind. Zudem sind viele Informationen erforderlich, um belastbare Aussagen zu treffen.

Die Verfahren zur Analyse der Daten wurden im Laufe der Zeit immer weiter verfeinert. In diesem Sinne stellen auch die Entwicklungen in dem Bereich Big Data und Künstliche Intelligenz (KI) für die Versicherungswirtschaft einen konsequenten nächsten Entwicklungsschritt dar. Durch die voranschreitende Digitalisierung der Versicherungsindustrie werden Verfahren aus dem Bereich Big Data und KI zunehmend relevanter, da ältere Ansätze eine bessere Datenlage nicht vollends ausschöpfen können. Im Folgenden werden drei Anwendungsbeispiele für Big-Data-Methoden beschrieben, die charakteristisch für das Lebensversicherungsgeschäft sind. Weitere Anwendungsbeispiele sowie detaillierte Beschreibungen zu wichtigen Methoden und Tools werden im Fachpapier „Big Data in der Lebensversicherung“ der Deutschen Aktuarvereinigung e.V. (DAV) beschrieben, das in Kürze veröffentlicht wird.

Es gibt nicht das Machine-Learning-Modell

Machine Learning (ML) ist ein Teilbereich der KI. Es beschreibt die Lösung eines Problems durch einen Algo-

rithmus, der in einem automatisierten Lernprozess auf Basis von Trainingsdaten erstellt und anschließend mit Validierungsdaten optimiert wird. Hierbei gibt es aber nicht „das Machine-Learning-Modell“. Vielmehr existiert eine Vielzahl höchst unterschiedlicher Modelle, die sich sowohl in ihrer Struktur als auch in ihrer Modellgüte, Schnelligkeit und Interpretierbarkeit unterscheiden. Prominente Vertreter sind zum Beispiel „Random Forest“, „Boosted Trees“, „neuronale Netze“ oder „Support Vector Machines“. Die Wahl der Modellklasse ist je nach Ziel und Sachverhalt zu treffen, auch innerhalb der Modellklassen gibt es durch verschiedene Parameter Kalibrierungsmöglichkeiten. Bei den Big-Data-beziehungsweise KI-Methoden wird neben kommerziellen Systemen vermehrt Open-Source-Technologie verwendet. Hierbei gewinnt insbesondere die Programmiersprache Python im Zusammenhang mit Machine Learning an Bedeutung.

Stornoprävention

Die Identifizierung von Kunden mit einem erhöhten Stornopotenzial stellt ein typisches Beispiel für ein sogenanntes „überwachtes Lernproblem“ dar. Eine besondere Rolle spielen bei diesem Beispiel historische Daten zum Training und zur Validierung. Der Vorteil dieses Machine-Learning-Modells gegenüber herkömmlichen Stornomodellen ist insbesondere die verbesserte Vorhersagegüte durch die erhöhte Anzahl von Merkmalen bei gleichzeitiger Kontrolle, wodurch es nicht zu einer Fehlklassifizierung kommt. So können neben Vertrags- und Kundendaten insbesondere Produkt- und Vermittlerinfor-

mationen in die Modellierung aufgenommen werden. Zudem ermöglicht die Berücksichtigung von zusätzlichen Informationen eine bessere Steuerung der Stornopräventionsmaßnahmen.

Underwriting – Kürzung der Gesundheitsfragen

Beim Abschluss einer Lebensversicherung ist oftmals ein umfangreicher Gesundheitsfragebogen auszufüllen. Ein großer Umfang an Informationen verbessert die Einschätzung des zu versichernden Risikos und ermöglicht dem Versicherer die Berechnung einer auskömmlichen Prämie. Der Kunde bevorzugt jedoch möglichst wenige, einfache und angenehme Fragen. Möchte ein Unternehmen seinen Gesundheitsfragebogen kürzen, muss es prüfen, auf welche Fragen verzichtet werden kann, ohne dass dadurch die Güte der Risikoeinschätzung wesentlich negativ beeinträchtigt wird. Machine-Learning-Verfahren können dabei helfen, solche Fragen zu identifizieren. Als Datengrundlage dienen die ausgefüllten Gesundheitsfragebögen. Für dieses Beispiel kann sich etwa ein ML-Baumverfahren wie „Random Forest“ eignen. Die vom Machine-Learning-Modell angegebene Relevanz der Merkmale (feature importance) kann Experten bei der Kürzung der Fragen helfen. So kann im nächsten Schritt ein Machine-Learning-Modell auf Grundlage von Trainingsdaten erstellt werden, bei dem die Merkmale mit geringer Relevanz ausgeschlossen wurden. Gibt es nur kleine Unterschiede in der Vorhersagegüte beider Modelle, spricht dies dafür, dass auf diese Fragen verzichtet werden könnte.

Aktuarielle Projektion

Traditionelle Projektionsmodelle benötigen viel Zeit und Ressourcen für ihre Vorhersagen. Mithilfe von Machine-Learning-Verfahren können sogenannte „Proxy-Modelle“ (Stellvertreter-Modelle) für Ad-hoc- oder „What-if“-Analysen erstellt werden. Das Machine-Learning-Modell wird in diesem Beispiel mithilfe von künstlich hergestellten (synthetischen) Daten trainiert und validiert, die vorher vom ursprünglichen aktuariellen Modell erzeugt werden müssen. Bei ausreichendem Training und Validierung ist das Modell in der Lage, deutlich schnellere Vorhersagen zu treffen als das Grundmodell. Es weist allerdings dafür in gewissen, vorher festgelegten Bereichen, akzeptierbare Abweichungen vom Grundmodell auf. Zudem kann das erstellte Machine-Learning-Modell durch die Untersuchung der gelernten Regeln bei der Vorhersagemodelloptimierung helfen, zum Beispiel durch neue Erkenntnisse über die Zusammenhänge verschiedener Versicherungsprodukte.

Antidiskriminierung und Datenschutz

In der öffentlichen Debatte kommt oftmals der Eindruck auf, dass es für den Einsatz von Algorithmen neuer Antidiskriminierungsvorgaben bedarf. Was unter Diskriminierung zu verstehen und somit verboten ist, ist allerdings bereits gesetzlich geregelt und gilt unabhängig vom Medium, also auch für den Einsatz von Machine-Learning-Algorithmen oder allgemein KI. In der Versicherungswirtschaft ist das Prinzip der sachgerechten Differenzierung verankert. Das heißt, Personen werden unterschiedlich behandelt, wenn sie Unterschiede bei relevanten Merkmalen aufweisen. Bei ungewollter unterschiedlicher Behandlung ist es grundsätzlich nicht ausreichend, das diskriminierende Kriterium selbst nicht in die Modellierung eingehen zu lassen, da es dennoch über andere Kriterien zu mittelbarer Diskriminierung kommen kann. Zu dessen Vermeidung sollten explizite Verfahren angewendet werden, wie die Umgewichtung der Trainingsdaten. Auch für den Datenschutz existieren bereits hinreichende Vorgaben. Die seit 2018 EU-weit geltende Datenschutzgrundverordnung regelt den Umgang und die Verarbeitung personenbezogener Daten ausführlich.

Fazit

KI-Methoden sind nächster Evolutions-schritt

Big-Data- und KI-Methoden stellen den nächsten Entwicklungsschritt der statistischen Methoden in der Versicherungswirtschaft dar. Eine vollständig automatisierte Entscheidungsfindung der Modelle ist jedoch bislang nicht absehbar. In der Praxis werden die Modelle eher zur Unterstützung einer menschlichen Entscheidung dienen. Die neuen Methoden können bei zahlreichen Schritten im Lebensversicherungsgeschäft unterstützen. Wegen einer potenziell schwierigen Datenlage besteht aber in großen Teilen der Branche Handlungsbedarf. Die Erhebung der Daten durch Fragebögen geschieht häufig nur einmalig, nicht extensiv und nicht abgestimmt auf die weitere Verarbeitung mit Big-Data-Methoden. Weiterführend könnten externe Datenquellen genutzt werden und Erhebungsverfahren ausgeweitet sowie angepasst werden. Zudem sind bestehende Daten oftmals nicht vollständig digital vorhanden, hier sind Digitalisierungsschritte notwendig. Big-Data- und KI-Methoden haben das Potenzial, eine große Unterstützung für die Branche zu werden, dieses lässt sich ohne eine hinreichende Datenlage nur schlecht ausschöpfen.