9
AI: Science over Fiction

Erkennung von Versicherungsbetrug mit künstlicher Intelligenz

Faktenpapier

Aus der Serie: AI: Science over Fiction

www.bitkom.org

bitkom

Herausgeber

Bitkom
Bundesverband Informationswirtschaft,
Telekommunikation und neue Medien e. V.
Albrechtstraße 10 | 10117 Berlin
T 030 27576-0
bitkom@bitkom.org
www.bitkom.org

Verantwortliches Bitkom-Gremium

AK Artificial Intelligence

Projektleitung

Dr. Nabil Alsabah | Bitkom e. V.

Autoren

Prof. Dr. Martin Spindler | Universität Hamburg
Dr. Heinrich Kögel | Economic AI

Satz & Layout

Katrin Krause | Bitkom e. V.

Titelbild

© Mat Reding | unsplash.com

1 Betrugserkennung ist eine Schlüsselkompetenz in Versicherungsunternehmen

»Management von Schadensfälle inklusive Betrugserkennung ist eine Kernkompetenz von Versicherungsunternehmen und Banken.«

Das Schadenmanagement und insbesondere die Erkennung von Versicherungsbetrug ist von herausragender Bedeutung für Versicherungsunternehmen. Laut einer Studie des Gesamtverbands der Deutschen Versicherungswirtschaft (GDV) ist nahezu jeder zehnte Schadensfall ein Betrugsfall.[1] Davon wird jedoch nur ein geringer Anteil aufgedeckt. Dies führt zu Schäden in Milliardenhöhe für Versicherungsunternehmen und letztlich auch für die Gemeinschaft aller Versicherten.

Versicherungsbetrug stellt in allen Versicherungssparten ein Problem dar, von der Schaden- und Unfallversicherung über die private und gesetzliche Krankenversicherung bis hin zur Lebensversicherung. Dabei gibt es unterschiedliche Arten von Versicherungsbetrug. So kann der Versicherungsnehmer beispielsweise durch fingierte Schadensmeldungen und Rechnungen unrechtmäßige Zahlungen vom Versicherungsunternehmen verlangen. Dies kann auch in Zusammenarbeit mit einer dritten Partei erfolgen. Derartiger Versicherungsbetrug ist häufig in der KFZ-Versicherung anzutreffen. In der Krankenversicherung wiederum hat der Leistungserbringer einen Anreiz, durch überhöhte Rechnungen bzw. durch Abrechnung von nicht erbrachten Leistungen zusätzliche Einnahmen zu generieren. Im Folgenden definieren wir Versicherungsbetrug sehr weit gefasst: von Tippfehlern in Rechnungen, die zum Auszahlen eines falschen Betrags durch das Versicherungsunternehmen führen, bis hin zu absichtlichen Betrugshandlungen, wie etwa die weiter oben ausgeführten Beispiele.

Eine effiziente Schadensbearbeitung und vor allem die Bekämpfung von Versicherungsbetrug ist für die Versicherungsbranche von außerordentlicher Bedeutung. Aufgrund der Vielzahl möglicher Betrugsszenarien und der zumeist großen Anzahl an zu bearbeitenden Fällen, stellt die Betrugserkennung allerdings seit jeher eine große Herausforderung für Versicherungsunternehmen dar. Zur Erkennung von Versicherungsbetrug kommen in der Praxis zumeist regelbasierte Systeme zum Einsatz, die mit manueller Inspektion und Intuition von Schadenssachbearbeiterinnen und Schadenssachbearbeitern kombiniert werden. Dieser Ansatz ist typischerweise nicht nur mit hohen Kosten verbunden, sondern auch zeitintensiv und anfällig für Fehler.

Durch die fortschreitende Digitalisierung entlang des gesamten Schadenprozesses liegen Versicherungsunternehmen heutzutage immer mehr Informationen in digitaler Form vor. Diese Entwicklung ermöglicht es der Versicherungsbranche, ihre Ansätze zur Betrugserkennung fundamental zu verbessern. Durch die Anwendung neuer Methoden aus dem Bereich des Maschinellen Lernens und der künstlichen Intelligenz (KI) können Betrugsmuster Algorithmus-basiert aus den Daten gelernt werden. Anschließend können diese zur automatisierten Betrugserkennung beim Eintreffen neuer Schadensfälle verwendet werden. Daraus ergeben sich Kostenersparnisse, nicht nur durch eine bessere Betrugserkennung, sondern auch durch ein schnelleres und effizienteres Handling von Schadensfällen. Diese Erfahrung hat auch Economic AI in gemeinsamen Projekten mit Versicherungen gemacht. Die frei werdenden Kapazitäten der Sachbearbeiterinnen und Sachbearbeiter können für andere Tätigkeiten verwendet werden, bei

1 Artikel »Du Lügst!« im Magazin »Positionen« des GDV, Ausgabe 3/2018, Seiten 24–26. ↗**https://www.gdv.de/resource/blob/40078/d6d85908fca5267199504650a1759dc7/positionen-ausgabe-2018-data.pdf**

denen menschlicher Input sehr wichtig ist, wie beispielsweise bei der Betreuung von Kunden und deren Bindung an das Unternehmen.

Die im Folgenden vorgestellten Methoden können nicht nur zur Betrugserkennung im Versicherungswesen verwendet werden. Sie können auch in anderen Bereichen zum Einsatz kommen, wie bei der Erkennung von Kreditkartenbetrug, Geldwäsche, Abrechnungsbetrug oder Finanzbetrug, da diese ähnliche Charakteristika aufweisen.

2 Versicherungsdaten weisen besondere Merkmale auf

»Nahezu jeder zehnte Schadensfall ein Betrugsfall.«

Bei der Analyse von Versicherungsdaten zur Betrugserkennung ergeben sich drei Herausforderungen. Um gute Ergebnisse zu erzielen, müssen diese beim Trainieren der Modelle und Algorithmen berücksichtig werden.

1. **Unstrukturierte Daten**. Die Daten in Versicherungs- bzw. Schadensdatenbanken liegen häufig in sogenannter »unstrukturierter« Form vor. Die Daten weisen somit keine formale Struktur auf und die Felder, die beispielsweise pro Schadenmeldung ausgefüllt sind, variieren von Fall zu Fall stark. Häufig sind auch viele Felder gar nicht ausgefüllt. Im Englischen spricht man hierbei von sogenannten »missing Data«. Außerdem gibt es Freitextfelder, in denen die Sachbearbeiterin oder der Sachbearbeiter Informationen in nicht standardisierter Form eintragen kann. Dies sind häufig Bilder vom Unfallort oder Schaden sowie Unfall- und Polizeiberichte. Krankenversicherungsdaten liegen häufig auch in unstrukturierter Form vor, da eingereichte Rechnungen eine unterschiedliche Anzahl an Abrechnungsposten aufweisen. Aufgrund ihrer fehlenden Struktur kann die Analyse von unstrukturierten Daten mittels traditioneller Methoden, wie Linearer Regression, an ihre Grenzen stoßen. Wie im nächsten Abschnitt näher erläutert, eignen sich allerdings neuere Methoden des Deep Learning gut, um unstrukturierte Daten zu analysieren. Die Aufbereitung der Daten bleibt dabei jedoch nach wie vor eine Herausforderung.

2. **Ungleichgewicht zwischen Betrugs- und Nichtbetrugsfällen**. Typisch für Versicherungsbetrug ist, dass aus datenanalytischer Perspektive nur ein verhältnismäßig geringer Anteil der Schadensfälle betrügerischer Natur ist und dass davon wiederum nur ein kleiner Anteil tatsächlich als Betrug erkannt wurde. Dies hat zur Folge, dass die Anzahl an Schadensfällen, die in den Daten als »Betrugsfall« klassifiziert ist, relativ zur Gesamtzahl aller Schadensfälle nur sehr gering ist. In der Statistik spricht man hierbei von sogenannten »unbalanced Data«, da es ein starkes Ungleichgeweicht zwischen Betrugs- und Nichtbetrugsfällen gibt. Dieses Ungleichgewicht erschwert das Trainieren von Algorithmen. Durch entsprechende statistische Modellierung kann hier jedoch Abhilfe geschaffen werden. Verbunden mit dieser Thematik ist auch die Frage, wie man die Performance des Algorithmus zur Betrugserkennung bewertet. Insbesondere stellt sich die Frage, wie die fälschliche Einstufung eines Nichtbetrugsfalls als Betrugsfall und wie das Übersehen eines tatsächlichen Betrugsfalls

bewertet werden soll. Hier ist die Wahl von geeigneten Zielfunktionen beim Trainieren der Algorithmen wichtig.

3. **Nicht entdeckte Betrugsfälle**. Bei erkannten Betrugsfällen besteht in aller Regel eine hohe Sicherheit, dass es sich tatsächlich um Betrug handelt. Es gibt somit sehr wenige Fälle im Datensatz, die fälschlicherweise als Betrug eingestuft wurden. Umgekehrt verhält es sich jedoch zumeist derart, dass es mit großer Sicherheit eine Vielzahl an Betrugsfällen gibt, die unentdeckt sind und fälschlicherweise als Nichtbetrug eingestuft wurden. Selbstverständlich sollen automatisierte Verfahren auch bisher unerkannte Betrugsfälle entdecken. Zu diesem Zweck werden spezielle Verfahren ergänzend zu Methoden des Deep Learning benötigt, die im nächsten Abschnitt ebenfalls erläutert werden.

3 Deep Learning ist der neue Standard zur Betrugserkennung

»Deep Learning setzt den Standard bei der datengetriebenen Erkennung von Betrugsfällen.«

Um Betrugsmuster datengetrieben zu lernen und damit neue Schadensfälle in Betrug oder Nichtbetrug klassifizieren zu können, müssen Algorithmen auf vorhandenen Daten trainiert und evaluiert werden. Um den im letzten Abschnitt ausgeführten Herausforderungen von unstrukturierten und unbalancierten Daten zu begegnen, eignen sich insbesondere Methoden des Deep Learnings. Zur Erzielung guter Ergebnisse müssen jedoch maßgeschneiderte Deep Learning Modelle entwickelt werden. Deep Learning ist ein Teilbereich des Maschinellen Lernens und gehört zu den Methoden des »supervised Learning«. Diese Methoden nutzen den Umstand, dass in den Daten für jeden Fall ein »Label« existiert, welches angibt, ob es sich um einen Betrugsfall oder Nichtbetrugsfall handelt. Anhand der weiteren im Datensatz verfügbaren Informationen versuchen die Algorithmen dann Muster zu erlernen, wie typische Betrugsfälle aussehen. Es können unterschiedliche Architekturen für Deep Learning Modelle gewählt werden. In aktuellen Forschungsarbeiten haben wir sogenannte Transformer Modelle und Embeddings entwickelt, die sehr gute Ergebnisse bei der Betrugserkennung in der praktischen Anwendung geliefert haben.

Wie bereits erwähnt, gibt es für gewöhnlich viele unentdeckte Betrugsfälle in Versicherungsdaten. Da diese nicht entdeckt wurden, können Methoden des »supervised Learning« jedoch auch keine typischen Betrugsmuster zu deren automatisiertem Auffinden erlernen. Zum Aufdecken solcher nicht erkannter Betrugsfälle können Methoden des sogenannten »unsupervised Learning« herangezogen werden. Diese Methoden suchen nach »ungewöhnlichen« Fällen in den Daten, die in bestimmten Dimensionen von der Mehrzahl der anderen Fälle stark abweichen. Sind solche Fälle identifiziert, können diese einer ausführlichen Prüfung durch Sachbearbeiter und Sachbearbeiterinnen unterzogen werden und das »Label« für den jeweiligen Fall gegebenenfalls von Nichtbetrug auf Betrug korrigiert werden. Sogenannte »Autoencoder« und »Variational Autoencoder« eignen sich besonders gut, um derartige Anomalien in den Daten zu entdecken. In aktuellen Forschungsarbeiten haben wir in Zusammenarbeit mit Kollegen von dem renommierten russischen Forschungsinstitut Skoltech solche Methoden zur Betrugserkennung entwickelt.

4 Deep Learning Modelle müssen erklärbar sein

»Die Vorhersagen der maschinellen Lernmodelle müssen von den Sachbearbeitern verstanden werden, um den Fällen nachzugehen.«

Beim Maschinellen Lernen liegt der Fokus traditionell auf dem Erzielen einer möglichst hohen Vorhersagegüte. Im Falle von Betrugserkennung bedeutet dies, dass eine möglichst präzise Klassifikation von Betrugs- und Nichtbetrugsfällen angestrebt wird. Fehlklassifikationen sollen somit möglichst vermieden werden.

Selbstverständlich ist eine genaue Erkennung von Betrug sehr wichtig, damit datengetriebene, automatisierte Systeme einen Mehrwert gegenüber den traditionellen, regelbasierten Ansätzen bieten. Eine genaue Betrugserkennung stellt jedoch lediglich den ersten Schritt dar, denn wenn ein neu gemeldeter Schadensfall oder eine neu eingereichte Rechnung als Betrugsfall klassifiziert wurde, muss der Sachbearbeiter oder die Sachbearbeiterin anschließend Nachforschungen anstellen, um den Betrug plausibel nachweisen zu können. Für eine gezielte Fallanalyse ist im zweiten Schritt deshalb wichtig zu wissen, welche Variablen dazu geführt haben, dass ein Fall als betrügerisch klassifiziert wurde. Anders formuliert bedeutet das, es muss ersichtlich sein, welche Konstellation in den Daten zur Einstufung als Betrug geführt hat.

Bei Deep Learning Modellen handelt es sich typischerweise um sogenannte »black Box« Verfahren. Zwar liefern die Modelle häufig sehr präzise Klassifikationen, jedoch ist zumeist nicht klar verständlich, wie eine bestimmte Klassifizierung eigentlich zu Stande kam. Der Grund für diese Schwierigkeit liegt darin begründet, dass Deep Learning Modelle mit Hilfe von sehr komplexen Transformationen der Daten zu ihren Ergebnissen gelangen. Die Transformationen werden mittels neuronaler Netze erzeugt. Im Bereich der künstlichen Intelligenz gibt es eine junge Forschungsrichtung, die unter den Namen »interpretable AI« und »explainable AI« an einer besseren Interpretierbarkeit von Deep Learning Modellen arbeitet. Aufbauend auf diesen neuen Forschungsarbeiten können auch Modelle zur Betrugserkennung besser interpretierbar gemacht werden. Das führt zu einer erheblichen Erleichterung des Claims-Management-Prozesses und macht Deep Learning Systeme praktisch einsetzbar.

Im Vergleich zu reinen »black Box« Verfahren stellen interpretierbare Modelle bereits einen enormen Fortschritt dar. Über die reine Interpretierbarkeit hinausgehend, sind in der Praxis allerdings häufig auch kausale Fragestellungen von Interesse: »Was für einen Effekt hat das Ändern einer Inputvariable auf die Zielvariable?«. Im Versicherungsbereich könnte beispielsweise von Interesse sein, wie sich die Höhe des monatlichen Einkommens eines Versicherten auf die Betrugswahrscheinlichkeit auswirkt. Derartige Fragestellungen können typischerweise nicht mit den bisher diskutierten Modellen beantwortet werden. Stattdessen muss auf sogenannte »kausale« Modelle zurückgegriffen werden, die in dem Artikel »Raus aus der Black Box« aus dem Jahr 2018 für Versicherungsanwendungen diskutiert werden (siehe Literaturverzeichnis).

5 KI-Systeme müssen in den Schadenmanagement-Prozess integriert werden

Bei Systemen der künstlichen Intelligenz stellt sich in der Praxis stets die Frage, wie diese in die bisherigen Prozesse eines Unternehmens integriert werden können. Ein neues System muss nicht nur für sich selbst genommen erfolgreich funktionieren, sondern auch einen Nutzen in der Zusammenarbeit mit Menschen stiften.

Im Schadenmanagement sollte das Ziel von automatisierten, datengetriebenen Systemen zur Betrugserkennung sein, dass sie die Schadenmeldungen in Echtzeit überwachen und Alarm schlagen, wenn sie einen Fall als Betrug einschätzen. Im Anschluss kann dann der angezeigte Fall darauf untersucht werden, ob es sich tatsächlich um einen Betrug handelt. Erhärten sich die Indizien, können nächste Schritte eingeleitet werden.

Ein weiterer wichtiger Aspekt bei der Integration von KI-Systemen in Unternehmen ist, dass die Systeme regelmäßig neu trainiert und gegebenenfalls beispielsweise neue Variablen aufgenommen werden müssen. Dies ist erforderlich, damit sich ändernde oder auch neue (Betrugs-) Muster erkannt werden können.

KI-Systeme zur Betrugserkennung stellen ein Werkzeug dar, das die Sachbearbeitung bei automatisierbaren Tätigkeiten entlasten soll. Dadurch können Sachbearbeiter sich verstärkt der gründlichen Prüfung von verdächtigen Fällen widmen und andere Tätigkeiten ausführen, die nicht von Maschinen übernommen werden können. Um das Potenzial von KI-Systemen auch tatsächlich voll ausschöpfen zu können, zeigt dieser Abschnitt, dass der Schadenmanagement-Prozess bei der Integration von automatisierten Betrugserkennungssystemen ebenfalls angepasst werden muss.

6 Geldwäsche, Kreditkarten- und Abrechungsbetrug stellen weitere Anwendungsfelder dar

In vielen anderen Bereichen gibt es ähnliche Datenstrukturen, wie im Versicherungsbereich. Dort können prinzipiell die gleichen Methoden zur Aufdeckung von betrügerischem Handeln eingesetzt werden. Im Bankenbereich ist beispielsweise die Verhinderung von Geldwäsche eine große Herausforderung. Banken sind gesetzlich verpflichtet, in diesem Bereich aktiv zu sein. Bei Finanzdienstleistern spielt Kreditkartenbetrug eine große Rolle und im Einzelhandel kann Betrug, beispielsweise in Zusammenhang mit Rechnungstellungen, ebenfalls große Schäden verursachen. Diese Beispiele zeigen, dass auch in vielen weiteren Bereichen ein großes Potenzial besteht, Deep Learning und künstliche Intelligenz gewinnbringend einzusetzen.

»Geldwäsche, Kreditkartenbetrug oder Abrechnungsbetrug können mit den gleichen Methoden analysiert und letztlich verhindert werden.«

7 Fazit und Ausblick

In diesem Beitrag wurde dargestellt, wie Betrug im Versicherungsbereich durch den Einsatz von künstlicher Intelligenz effizienter aufgedeckt werden kann. Durch die Digitalisierung stehen Versicherungsunternehmen immer größere Datenmengen zur Verfügung. Diese bieten ein großes Potenzial, KI-Systeme zur Betrugserkennung einzusetzen. Daten im Versicherungsbereich weisen jedoch einige Besonderheiten auf, die bei der Entwicklung von KI-Modellen bzw. Deep Learning Modellen berücksichtigt werden müssen. In aktuellen Forschungsarbeiten gemeinsam mit Kooperationspartnern und in anwendungsorientierten Projekten haben wir für diese Herausforderungen Lösungen entwickelt, die im Vergleich zu herkömmlichen Modellen eine bessere Performance liefern.

»Um auch von Nutzen im betrieblichen Alltag zu sein, müssen die KI-Modelle in die Abläufe und IT-Systeme integriert werden.«

Künstliche Intelligenz zur Betrugsabwehr kann zu Kosteneinsparungen und zu einem effizienteren Schadenbearbeitungsprozess führen. Unsere Erfahrung ist, dass KI-Systeme traditionelle Ansätze zur Betrugserkennung in der Versicherungsbranche übertreffen.

Es gibt Bemühungen, die aktuellen Systeme noch weiter zu verbessern. Momentan wird daran geforscht, dass auch soziale Netzwerkstrukturen zur Betrugserkennung herangezogen werden können: In der KFZ-Versicherung sind zum Beispiel Fälle bekannt, in denen Anwälte und Werkstätten gemeinsam Arbeiten, um Versicherungsbetrug zu begehen und in Krankenversicherungen kann es passieren, dass Ärzte und Apotheken kollusiv zusammenarbeiten. Mit neuesten Methoden können auch derartige Strukturen in Modellen berücksichtigt werden. Dadurch kann die Gemeinschaft der Versicherten noch besser geschützt und die Versicherungsbeiträge effizienter genutzt werden.

8 Literaturverzeichnis

1. Farbmacher, H., Löw, L., und M. Spindler (2020): »An Explainable Attention Network for Fraud Detection in Claims Management«, im Erscheinen beim Journal of Econometrics.

2. Fursov, I., Zaytsev, A., Khasyanov, R., Burnaev, E., und M. Spindler (2019): »Sequence Embeddings Help to Identify Fraudulent Cases in Healthcare Insurance«, arXiv: 1910.03072.

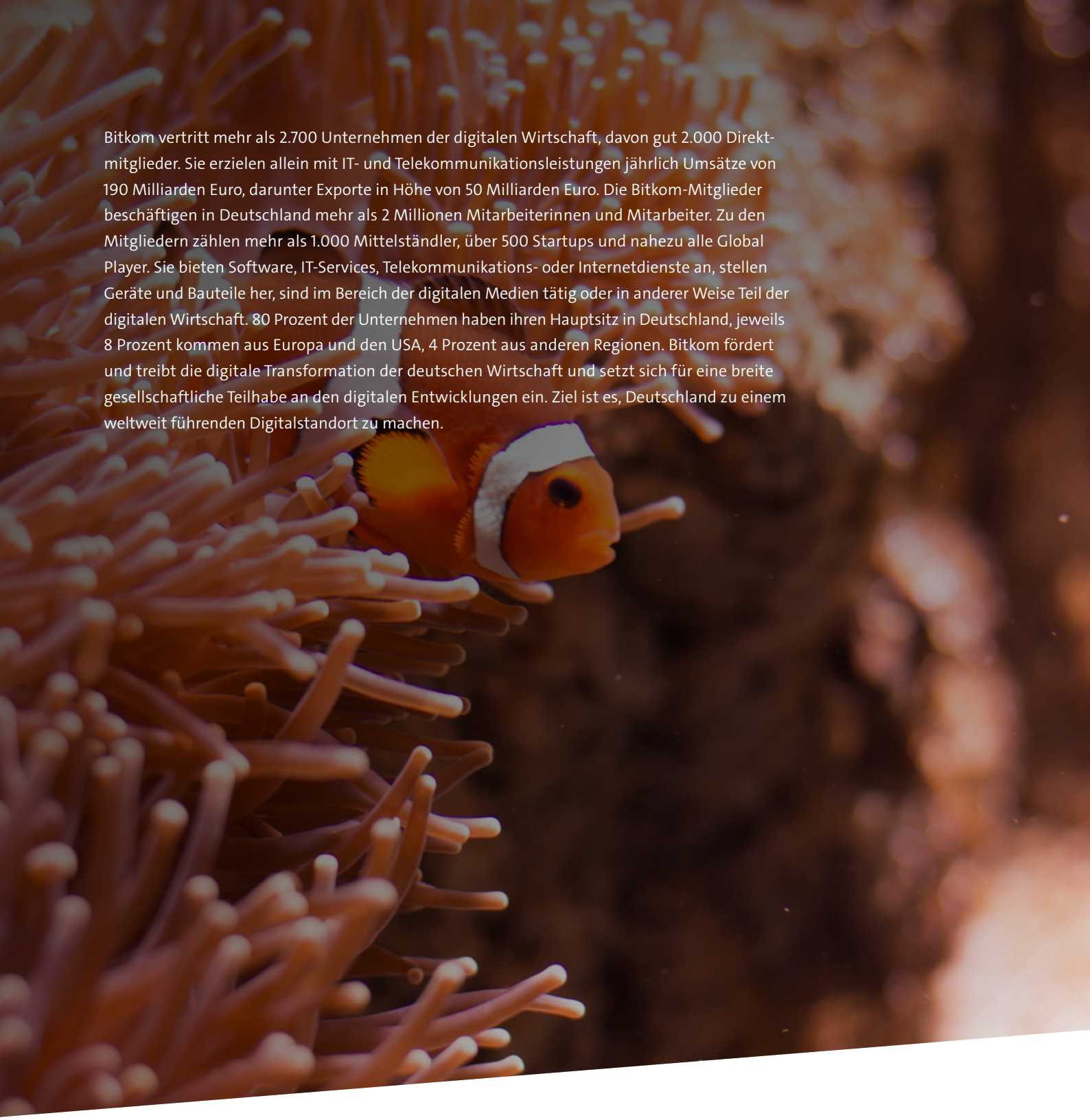
3. Löw, L., Spindler, M., und E. Brechmann: »A Self-Attention Network for Hierarchical Data Structures with an Application to Claims Management«, arXiv: 1808.10543.

4. Spindler, M. (2018): »Raus aus der Black Box«, Versicherungswirtschaft (73), 70–72.

Autoren

Martin Spindler
Martin Spindler ist Professor für Statistik und Maschinelles Lernen an der Universität Hamburg. In der Forschung beschäftigt er sich mit den Grundlagen von Maschinellem Lernen und der Anwendung auf betriebliche Probleme. Zusammen mit führenden internationalen Forschern hat er das Unternehmen Economic AI gegründet, um Unternehmen bei der Anwendung und Entwicklung von KI zu unterstützen und einen Wettbewerbsvorteil durch maßgeschneiderte forschungsbasierte Lösungen zu sichern.

Heinrich Kögel
Heinrich Kögel ist Data Science Manager bei Economic AI. Er ist in der Entwicklung von Algorithmen tätig und führt Projekte im Bereich Data Science durch. Zuvor forschte Heinrich Kögel auf dem Gebiet der empirischen Wirtschaftswissenschaften u.a. bei der Max-Planck-Gesellschaft und verbrachte einen Forschungsaufenthalt an der Harvard University. Er hat in Quantitativer Ökonomik an der Ludwig-Maximilians-Universität München promoviert.



Bitkom vertritt mehr als 2.700 Unternehmen der digitalen Wirtschaft, davon gut 2.000 Direktmitglieder. Sie erzielen allein mit IT- und Telekommunikationsleistungen jährlich Umsätze von 190 Milliarden Euro, darunter Exporte in Höhe von 50 Milliarden Euro. Die Bitkom-Mitglieder beschäftigen in Deutschland mehr als 2 Millionen Mitarbeiterinnen und Mitarbeiter. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 500 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig oder in anderer Weise Teil der digitalen Wirtschaft. 80 Prozent der Unternehmen haben ihren Hauptsitz in Deutschland, jeweils 8 Prozent kommen aus Europa und den USA, 4 Prozent aus anderen Regionen. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem weltweit führenden Digitalstandort zu machen.

Bundesverband Informationswirtschaft,
Telekommunikation und neue Medien e.V.

Albrechtstraße 10
10117 Berlin
T 030 27576-0
F 030 27576-400
bitkom@bitkom.org
www.bitkom.org

bitkom