Fraunhofer IAIS

Effiziente Betrugserkennung durch Maschinelles Lernen

Neue Potenziale mit »Informed Machine Learning« und »Explainable Artificial Intelligence« erschließen

www.iais.fraunhofer.de/fraunhofer-fraud-detection

Fraunhofer
IAIS

Effiziente Betrugserkennung durch Maschinelles Lernen

Neue Potenziale mit »Informed Machine Learning« und »Explainable Artificial Intelligence« erschließen

Autorinnen und Autoren
Dr. Daniel Trabold
Raoul Blankertz
Lisa Schrader

www.iais.fraunhofer.de/fraunhofer-fraud-detection

Das Fraunhofer IAIS

Als Teil der größten Organisation für anwendungsorientierte Forschung in Europa ist das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS mit Sitz in Sankt Augustin bei Bonn eines der führenden Wissenschaftsinstitute auf den Gebieten Künstliche Intelligenz, Maschinelles Lernen und Big Data in Deutschland und Europa. Mit seinen rund 300 Mitarbeitenden unterstützt das Institut Unternehmen bei der Optimierung von Produkten, Dienstleistungen, Prozessen und Strukturen sowie bei der Entwicklung neuer digitaler Geschäftsmodelle. Damit gestaltet das Fraunhofer IAIS die digitale Transformation unserer Arbeits- und Lebenswelt.

Inhalt

1. Executive Summary

Immer mehr alltägliche Anwendungen profitieren von Künstlicher Intelligenz (KI) und ihrer Fähigkeit, komplexe Probleme zu lösen. Die meisten KI-Verfahren nutzen für diese Zwecke Methoden des Maschinellen Lernens (ML), d. h. intelligente Algorithmen, die Daten auswerten und selbstständig aus ihnen lernen. Zu den bekanntesten Anwendungen zählen u. a. Sprach- und Gesichtserkennungssysteme, Empfehlungssysteme oder die medizinische Bilderkennung zur Unterstützung von Diagnosen. Auch in der intelligenten Betrugserkennung (Fraud Detection) wird Maschinelles Lernen bereits gewinnbringend eingesetzt. Fehlende Transparenz und mangelnde Genauigkeit schränkten die Anwendbarkeit bisher jedoch gelegentlich ein. Neueste Entwicklungen im Maschinellen Lernen ermöglichen es, fachliches Expertenwissen einzubinden und Transparenz zu schaffen, wodurch die Effizienz der Betrugserkennung deutlich gesteigert wird. Dieses Whitepaper beschäftigt sich damit, wie Betrug aktuell mit datengetriebenen Methoden erkannt wird und welche Potenziale sich durch die neuesten Methoden des vom Fraunhofer IAIS geprägten »Informed Machine Learning« und der internationalen Forschung an »Explainable Artificial Intelligence« für die datengetriebene Betrugserkennung ergeben.

1.1 Einleitung

Die Entwicklungen der letzten Jahre im Maschinellen Lernen führen zu immer besseren Modellen mit steigender Präzision und Trefferquote in der Analyse von Daten. Vor allem mit Methoden des Maschinellen Lernens mit tiefen neuronalen Netzen (Deep Learning) hat die Entwicklung der Künstlichen Intelligenz an Fahrt aufgenommen. Bei einer sehr großen Menge an Trainingsdaten erzielen Deep-Learning-Modelle hohe Trefferquoten, sind jedoch oft nicht transparent und interpretierbar. Solche Black-Box-Modelle können nicht ohne Weiteres in Bereichen eingesetzt werden, in denen das Verständnis darüber unabdingbar ist, wie genau die Entscheidungen getroffen werden. Darüber hinaus ist es sehr schwierig, wichtiges fachliches Expertenwissen in diese Modelle zu integrieren, weswegen dieses häufig ungenutzt bleibt.

Diese Nachteile der klassischen Verfahren des Maschinellen Lernens spielen auch eine Rolle für die Betrugserkennung. Primäres Ziel ist es hier, eine hohe Erkennungsrate der Betrugsfälle bei gleichzeitig hoher Genauigkeit zu erreichen, wofür das Erfahrungswissen von Analyst*innen eine entscheidende Rolle spielt. Gleichzeitig soll ein Modell für den Zweck der Betrugserkennung auch Transparenz über seine Entscheidungen liefern. Nur so kann im Review, bei dem die identifizierten Betrugsfälle manuell geprüft werden, die Entscheidung des Modells effizient nachvollzogen und dokumentiert werden. Es gibt viele wissenschaftliche Ansätze für den Einsatz von Methoden des Maschinellen Lernens in der Betrugserkennung [1, 2] sowie praktische Erfolge im realen Einsatz, wie zum Beispiel bei der Erkennung von Kreditkartenbetrug [3]. Traditionellen Methoden des Maschinellen Lernens fehlt oft die Möglichkeit, Expertenwissen systematisch zu integrieren und einen Fokus auf die Nachvollziehbarkeit ihrer Entscheidung zu setzen. Mit neuen Methoden aus der Wissenschaft, wie dem vom Fraunhofer IAIS geprägten »Informed Machine Learning« (Informed ML [4]) und der internationalen Forschung an »Explainable Artificial Intelligence« (Explainable AI [5]), wird die Vereinbarkeit von hoher Genauigkeit, Integration von Expertenwissen und Erklärbarkeit ermöglicht. Mit Informed Machine Learning lernen Verfahren nicht nur datengetrieben, sondern aus Daten und Expertenwissen zugleich. Dies erhöht die Genauigkeit und kann Transparenz schaffen. Zudem können mit Explainable AI Modelle so konstruiert werden, dass sie zu jeder Entscheidung wertvolle und nachvollziehbare Begründungen liefern, die zum Beispiel in einer Überprüfung benötigt werden.

2. Betrugsdelikte in der Wirtschaft

Im Deliktsbereich der Wirtschaftskriminalität versteht man unter Betrug (im Englischen »Fraud« und in der Fachsprache der Wirtschaftsprüfung »dolose Handlung«) vorsätzliche und verschleierte Handlungen, um einen ungerechtfertigten Vorteil zu erlangen.

2.1 Interner und externer Betrug

Betrug im Unternehmen hat viele verschiedene Facetten und wird grob in internen Betrug (**Internal Fraud**) und externen Betrug (**External Fraud**) unterschieden. Internal Fraud (oder auch **Occupational Fraud**) wird von Personen innerhalb des geschädigten Unternehmens begangen (Innentäter). Beispiele dafür sind Unterschlagung, Korruption und Bilanz-Manipulation. Im Gegensatz zum Internal Fraud greifen Betrügende beim External Fraud das Unternehmen von außen an. Darunter fallen unter anderem Online-Zahlungsbetrug, Erstellung von Fake-Accounts und Account-Diebstahl. Das Risiko der Betrugsarten variiert stark je nach Branche. Telekommunikationsgesellschaften sehen sich u. a. mit Abrechnungsbetrug und Roaming-Betrug konfrontiert [6], wohingegen im Finanzdienstleistungssektor Kreditkartenbetrug, Geldwäsche und Rogue Trading (unautorisierter Handel) vorkommen [7].

2.2 Wirtschaftlicher Schaden für Unternehmen

Das Ausmaß von Betrugsdelikten ist für Unternehmen nur schwer zu beziffern, da solche Wirtschaftsstraftaten häufig ohne polizeiliche Beteiligung bearbeitet werden und somit nicht in Statistiken erfasst werden. Hinzu kommt ein geringes Anzeigeverhalten aufgrund eines befürchteten Image-Verlustes [8]. Verschiedene Studien versuchen daher abzuschätzen, welcher Schaden durch Betrug in Deutschland entsteht. Das Bundeskriminalamt (BKA) kommt in seinem Lagebericht dabei auf ein Gesamtschadensvolumen von 728 Millionen Euro für »Wirtschaftskriminalität bei Betrug« (siehe Abb. 1) [8]. Durch nicht entdeckte und nicht gemeldete Fälle ist die Dunkelziffer jedoch sehr viel höher. Zu dem rein finanziellen Schaden für die Unternehmen kommen oft noch schwer zu beziffernde Reputationsverluste und andere Folgeschäden hinzu. Laut verschiedener Studien sind 30 bis 47 Prozent aller Unternehmen weltweit betroffen, die jeweils Gesamtschäden von bis zu mehreren Millionen Euro erleiden [9, 10].

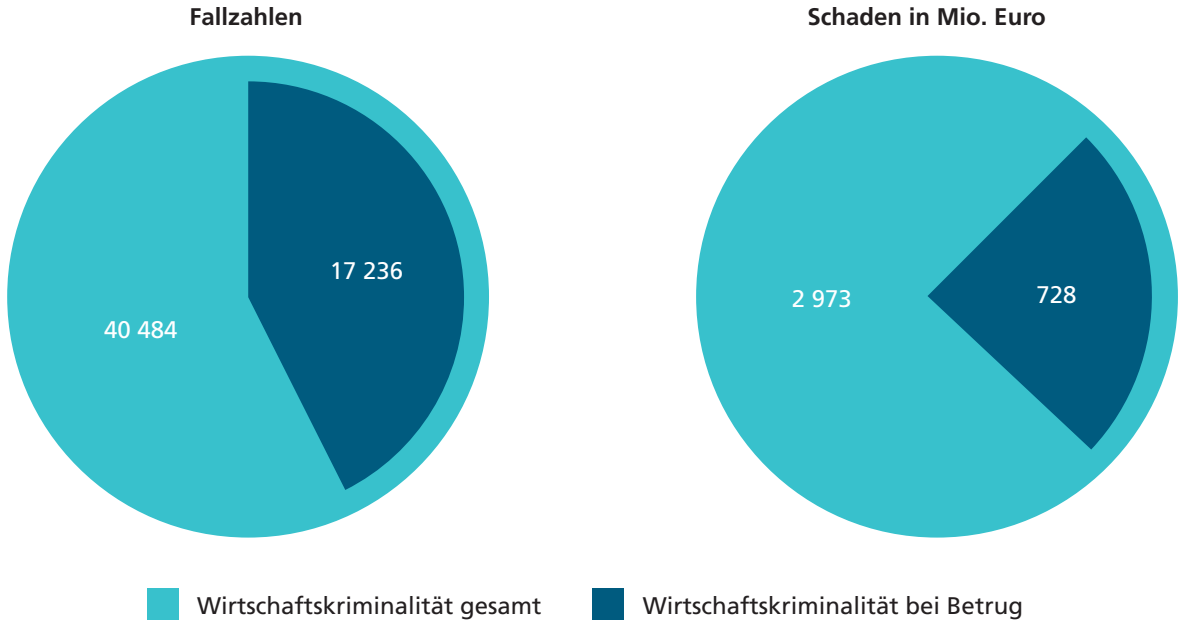

Abbildung 1: Wirtschaftskriminalität: Vergleich Fallzahlen und wirtschaftlicher Schaden (in Millionen Euro) allgemein mit Betrugsfällen im Besonderen [8].

3. Datengetriebene Betrugserkennung

Um sich vor Betrugsschäden zu schützen, empfiehlt es sich für Unternehmen verschiedene Maßnahmen zu ergreifen. Neben Maßnahmen, die Betrug allgemein erschweren, wird oft zusätzlich daraufgesetzt, Betrug frühzeitig zu identifizieren und so den Betrugsfall zu verhindern oder den Schaden zu minimieren. Dabei spielen datengetriebene Methoden eine große Rolle. Eine Studie der weltgrößten Organisation zur Bekämpfung von Wirtschaftskriminalität, der Association of Certified Fraud Examiners, hat ermittelt [11], welche datengetriebenen Techniken für die Betrugserkennung in Unternehmen zum Einsatz kommen (siehe Abb. 2). Demzufolge nutzen die meisten Unternehmen eher klassische Methoden wie die automatisierte Meldung von Anomalien oder zuvor definierten Regelverstößen. Vergleichsweise selten werden bisher Methoden der Künstlichen Intelligenz wie das Maschinelle Lernen eingesetzt. Allerdings gibt hier jedes vierte der befragten Unternehmen an, derlei Methoden in den nächsten zwei Jahren einsetzen zu wollen.

3.1 Rahmenbedingungen

Trotz der Vielzahl an unterschiedlichen Deliktsbereichen und Branchen gibt es einheitliche Anforderungen und Herausforderungen für die datengetriebene Betrugserkennung.

Allgemeine Anforderungen an ML-Modelle:

- **Hohe Trefferquote:** Ein Betrugserkennungssystem sollte möglichst viele Betrugsfälle auch als solche erkennen, da jeder nicht identifizierte Fall zu finanziellem Schaden und Reputationsverlust führen kann.
- **Hohe Präzision:** Je nach Einsatzszenario schließt sich an die automatische Erkennung der Verdachtsfälle eine manuelle Überprüfung an oder Transaktionen werden automatisch gesperrt. In beiden Fällen ist eine hohe Präzision wichtig, um zum Beispiel Fehlalarme zu vermeiden und eine hohe Kundenzufriedenheit zu gewährleisten.

Besondere Herausforderungen bei der Betrugserkennung:

- **Kontinuierlich variierende Betrugsmuster:** Die Muster, nach denen Betrüger*innen vorgehen, ändern sich kontinuierlich und professionelle Betrüger*innen suchen stets nach neuen Möglichkeiten, sich einen Vorteil zu verschaffen. Selbstlernende Systeme sind in der Lage, sich an diese Änderungen kontinuierlich und schnell anzupassen.
- **Unbalancierte Daten:** Es gibt vergleichsweise wenig Betrugsfälle, d. h. unter vielen legitimen Transaktionen sind nur vereinzelt auffällige Transaktionen zu finden. Um die seltenen Betrugsfälle zu identifizieren, bieten sich spezielle Methoden wie zum Beispiel Regellerner oder Datenvorverarbeitungsschritte wie Sampling an.
- **Transparenz:** Bei der Betrugserkennung kann es verschiedene Gründe für die Notwendigkeit von Transparenz geben. Einige Beispiele umfassen Gerichtsfestigkeit, Revisionssicherheit, gesetzliche Vorschriften, Nachvollziehbarkeit bei manueller Überprüfung und Erklärbarkeit von automatisch geblockten Transaktionen.

Für eine gute Balance im Spannungsfeld dieser Anforderungen sind spezielle Verfahren für die Betrugserkennung notwendig.

3.2 Aktuelle Verfahren der Betrugserkennung

Bisher gängige Herangehensweisen an die datengetriebene Betrugserkennung lassen sich in zwei Ansätze aufteilen (siehe auch Abb. 2):

- **Manueller Ansatz:** Zum Beispiel die manuelle Erstellung von feststehenden Regeln, mit denen Transaktionen in einer Datenbank identifiziert werden. Je nach Anwendungsgebiet wird dies auch »Score Cards« oder »Red-Flag-Analyse« genannt.
- **Ansatz mit Maschinellem Lernen:** Der Einsatz von selbstlernenden Verfahren, wie beispielsweise Regellerner, Support-Vector-Maschinen, künstlichen neuronalen Netzen etc. zur Identifizierung von auffälligen Transaktionen.

Der **manuelle Ansatz** bietet einige Vorteile gegenüber vielen klassischen Methoden des Maschinellen Lernens und wird daher in einigen Bereichen, wie zum Beispiel bei der Erkennung von Internal Fraud (vgl. Kapitel 2.1), vorwiegend eingesetzt. Das Expertenwissen der Analyst*innen kommt beim manuellen Ansatz in großem Umfang zum Einsatz, was gleichzeitig eine sehr hohe Transparenz der Vorgehensweise und Entscheidungen gewährleistet. Durch die damit verbundene manuelle Arbeit und sich ändernde Betrugsmuster verbraucht dieser Ansatz allerdings viel Zeit und Ressourcen und kann nicht gut

skaliert werden. Hinzu kommt, dass der manuelle Ansatz in den meisten Fällen nicht an die Genauigkeit von Ansätzen des Maschinellen Lernens herankommt, was zu einer hohen Anzahl falscher Verdachtsfälle und unentdeckter Betrugsfälle führt.

Für eine schnelle Identifizierung kontinuierlich variierender Betrugsmuster bietet sich daher der **Einsatz von Methoden des Maschinellen Lernens** an. Eine Reihe wissenschaftlicher Studien beschäftigt sich mit dem Einsatz von Maschinellem Lernen für die Betrugserkennung [1, 2], wobei untersuchte Methoden von Entscheidungsbäumen über Support-Vector-Maschinen bis hin zu tiefen neuronalen Netzen reichen. Mit einigen dieser Methoden (z. B. Support-Vector-Maschinen, tiefe neuronale Netze) können im Gegensatz zum manuellen Ansatz sehr hohe Trefferquoten und eine hohe Präzision erreicht werden, allerdings keine Transparenz. Letztere kann für Modelle des Maschinellen Lernens entweder durch Interpretierbarkeit (Verständlichkeit des ganzen Modells) oder durch Erklärbarkeit (Nachvollziehbarkeit der einzelnen Ausgaben) erreicht werden. Zusätzlich zu fehlender Transparenz ist bei den klassischen Methoden nicht klar, wie Fachexpert*innen mit den Systemen interagieren und wie sie wertvolles Expertenwissen integrieren können. Methoden, die in Kapitel 4 aufzeigt werden, können diese Nachteile überwinden.

Es zeigt sich in der Praxis, dass selbstlernende Verfahren nicht in allen Deliktsbereichen und Branchen standardmäßig zum Einsatz kommen. Die Gründe dafür liegen in den verschiedenen Herausforderungen der jeweiligen Bereiche. Insbesondere sind die Betrugsszenarien sehr unterschiedlich. Bei vielen Arten des Internal Frauds liegen in der Regel keine annotierten Daten für die Betrugserkennung vor und Modelle müssen manuell oder mit sogenannten unüberwachten Methoden erstellt werden. Im Gegensatz dazu melden beim Kreditkartenbetrug viele Kreditkartenbesitzer*innen selbst die Unregelmäßigkeiten, wodurch Transaktionen mit dem Label »Fraud« versehen werden können. Anschließend können Methoden aus dem Bereich des überwachten Maschinellen Lernens mit diesen historischen Daten trainiert werden. Verfahren können somit die Muster für Betrug anhand bestehender Daten lernen. An den folgenden beiden Beispielen werden die speziellen Herangehensweisen der Betrugserkennung veranschaulicht. Das erste Beispiel zeigt ein Anwendungsszenario für Internal Fraud, in dem Maschinelles Lernen in der Praxis derzeit sehr selten zum Einsatz kommt, es aber vielversprechende Ansätze mit großem Potenzial gibt. Das zweite Beispiel demonstriert den Einsatz von Maschinellem Lernen für die Erkennung von Kreditkartenbetrug.

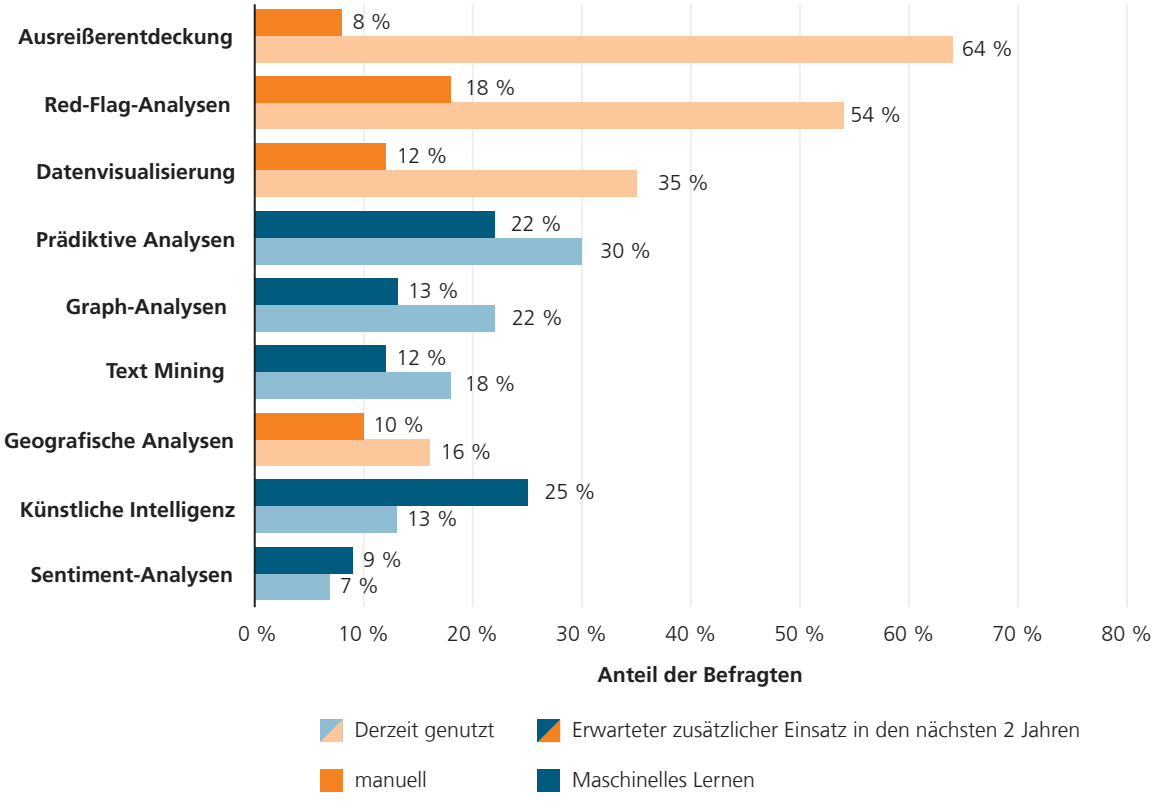

Abbildung 2: Techniken für die datengetriebene Betrugserkennung in Unternehmen weltweit, n = 1055 [11].

3.3 Beispiel: Erkennung von Betrug in Buchhaltungsdaten

Das Ziel der Erkennung von Internal Fraud in Buchhaltungsdaten – z. B. durch eine interne Revision oder Wirtschaftsprüfung – ist es, betrügerische Handlungen eigener Mitarbeitenden gegen das Unternehmen wie zum Beispiel Unterschlagung oder Korruption aufzudecken. Da die Buchhaltungssoftware alle finanziellen Geschäftsprozesse des Unternehmens abbildet, hinterlassen auch betrügerische Handlungen dort ihre Spuren.

Neben Compliance- und Anti-Fraud-Maßnahmen zur Verhinderung von Internal Fraud ist es wichtig, entsprechende dolose Handlungen frühzeitig zu identifizieren, um den Schaden zu minimieren. Verschiedene Studien ermitteln, welche Methoden dabei eingesetzt werden. Die Association of Certified Fraud Examiners kommt in ihrer aktuellen Studie [12] zu dem Ergebnis, dass Internal Fraud in erster Linie (43 Prozent) durch Hinweise u. a. von Mitarbeitenden und Kund*innen entdeckt wird (vgl. Abb. 3). Diese primär passive Methode geht jedoch mit längeren Aufdeckungszeiten und höheren Schadensvolumen einher, wohingegen aktive Methoden wie Audits, Überprüfungen oder gezieltes Monitoring den Schaden deutlich reduzieren [12]. Bei den befragten Unternehmen gehört die proaktive Datenanalyse (sogenannte »Forensic Data Analysis«) bzw. das Monitoring mit 38 Prozent allerdings zu den am wenigsten geläufigen Kontrollmechanismen [12] (siehe auch Abb. 4). Der Umstand, dass viele Fälle nur durch Hinweise oder Zufall erkannt werden und die hohe Anzahl unentdeckter Fälle zeigen, dass es noch großes Potenzial bei der systematischen Identifizierung von Internal Fraud gibt.

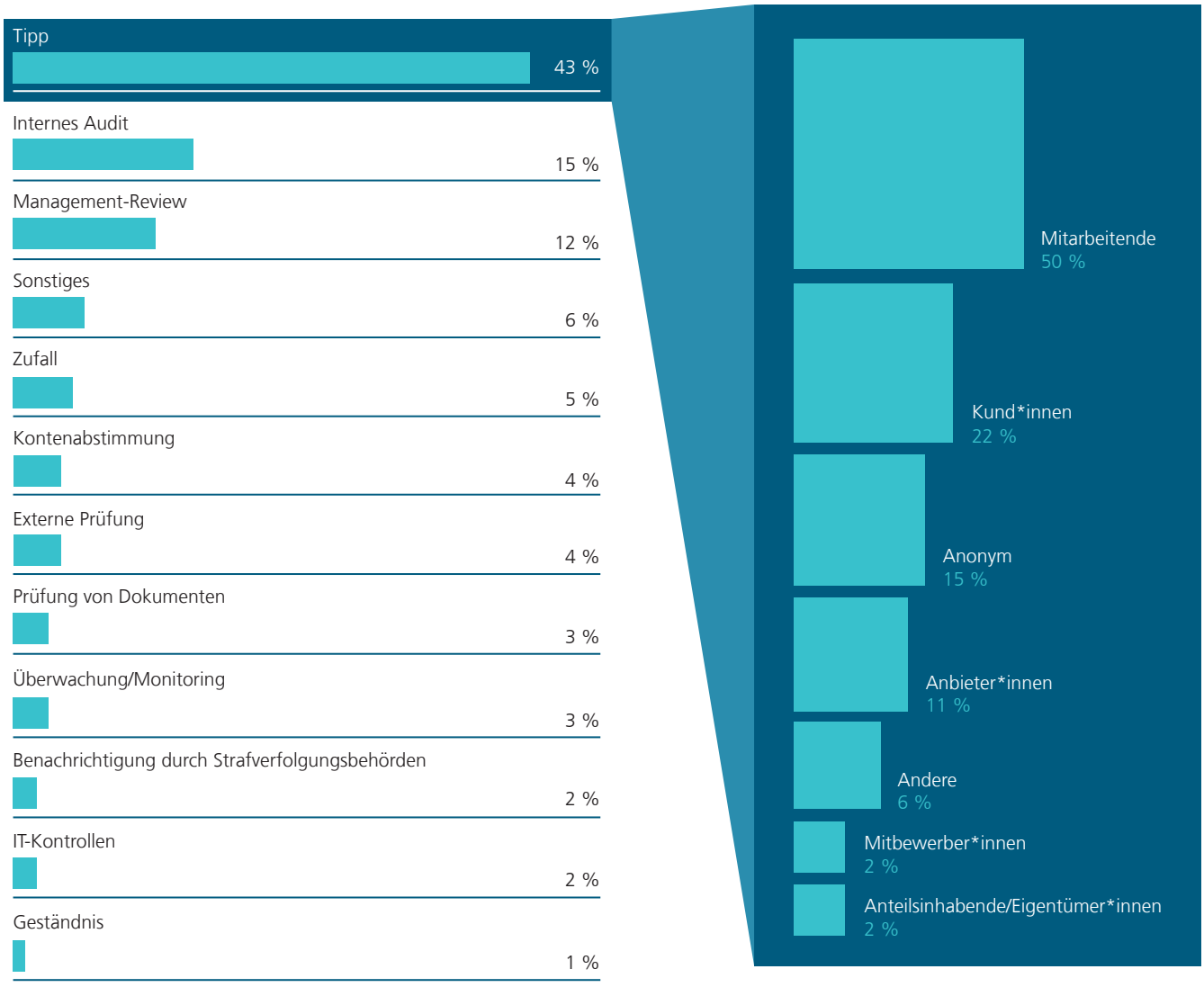

Abbildung 3: Wie Internal Fraud in Unternehmen entdeckt bzw. von wem er gemeldet wird [12].

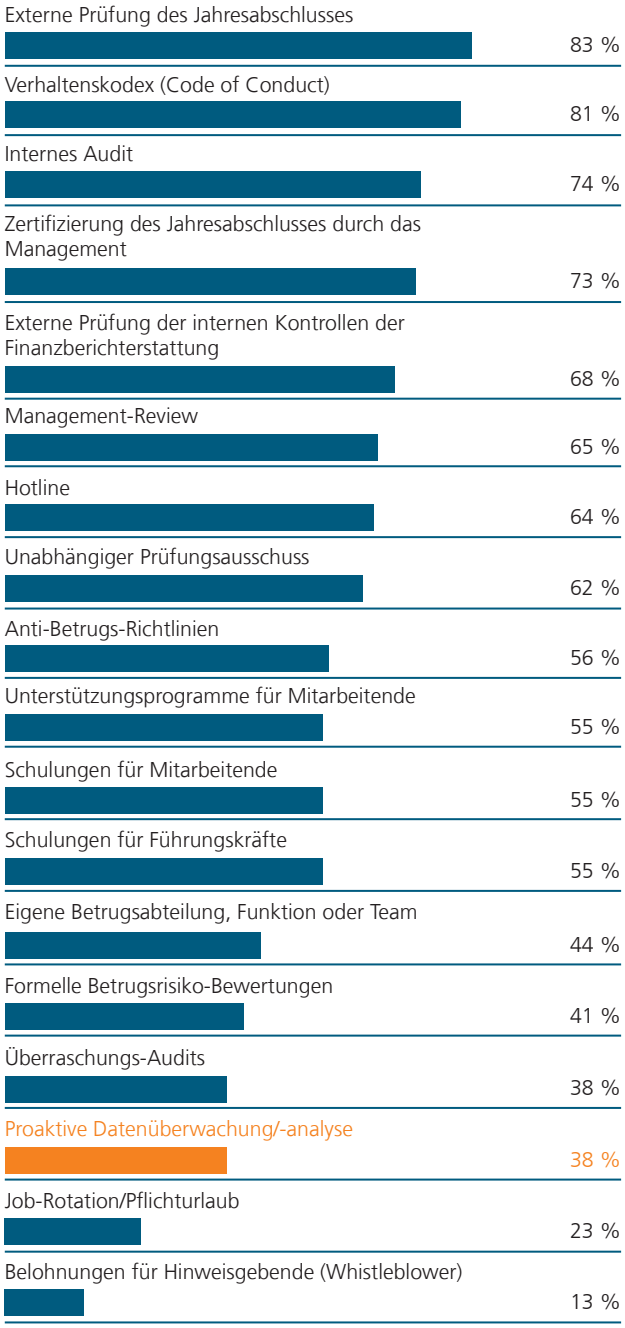

Abbildung 4: Welche Anti-Fraud-Maßnahmen am häufigsten in Unternehmen zum Einsatz kommen [12].

Die etablierte Methode aus dem Bereich »Forensic Data Analysis« beruht auf dem manuellen Ansatz (siehe Kapitel 3.2 »Aktuelle Verfahren der Betrugserkennung«). Dieser nutzt vorher festgelegte und starre Regeln. Dafür erfolgt im Vorfeld der Analyse eine Risikoeinschätzung (Risk Assessment) und die Ausrichtung der Analyse auf bestimmte Unternehmens- und Datenbereiche. Neben hohem manuellem Aufwand besteht das Risiko, dass nicht alle ungewöhnlichen Muster in den Daten entdeckt und damit ggf. betrügerische Handlungen übersehen werden. Hinzu kommt, dass Innentäter*innen die Kontrollen des Unternehmens kennen (bzw. unauffällig austesten können). So können sie neue Muster entwickeln, um dolose Handlungen zu begehen, ohne dabei durch die starren regelbasierten Analysen erkannt zu werden. Diese starren Regeln tendieren dazu, nicht besonders präzise zu sein und erzeugen viele Fehlalarme. Da die Überprüfung von Verdachtsfällen besonders aufwendig ist, trägt die Reduzierung der Fehlalarme direkt zu einer deutlichen Kostensenkung bei und erhöht darüber hinaus das Vertrauen in das System.

Wie in Kapitel 3 schon für den allgemeinen Fall dargestellt wurde, fehlt es Methoden des Maschinellen Lernens, die bisher zur Erkennung von Internal Fraud eingesetzt wurden, entweder an Erklärbarkeit oder an der nötigen Komplexität, um eine hohe Präzision zu erreichen. In neuesten wissenschaftlichen Untersuchungen konnte jedoch gezeigt werden, wie Ansätze mit tiefen neuronalen Netzen für die Erkennung von Internal Fraud in Buchhaltungsdaten eingesetzt werden können, damit gleichzeitig fachlich nachvollziehbare Erklärungen der gefundenen Anomalien geliefert werden [13]. Dieser Ansatz gehört zu dem Gebiet »Explainable Artifical Intelligence«, in dem das Fraunhofer IAIS umfangreiche Expertise hat (siehe Kapitel 4).

3.4 Beispiel: Bekämpfung von Kreditkartenbetrug

Big Data Analytics und Künstliche Intelligenz sind in der Finanzdienstleistungsbranche heute – und noch mehr in naher Zukunft – eine entscheidende Technologiekomponente [14]. Insbesondere bei der Betrugsprävention spielen maschinelle Lernverfahren bereits heutzutage eine große Rolle, wie das folgende Beispiel über Kreditkartenbetrug zeigt.

Für Zahlungskartenbetrug im Allgemeinen gibt es verschiedene Angriffsvektoren, darunter sowohl Straftaten mit echten (gestohlenen) als auch mit gefälschten Zahlungskarten. Der größere Teil des Zahlungskartenbetrugs entfällt auf den Einsatz von gestohlenen Kreditkartendaten im Internet (siehe Abb. 5, »Karte nicht vorhanden«) und ist damit der Cyberkriminalität zuzuordnen. Eine starke Internationalisierung und Arbeitsteilung (z. B. beim Handel mit Kreditkartendaten) stellt die Bekämpfung vor große Herausforderungen.

Neben technischen und prozessualen Vorkehrungen, die das Ausführen von betrügerischen Kreditkartentransaktionen im Vorhinein verhindern oder erschweren (z. B. durch 3D-Secure-Codes), müssen Transaktionen durch den Emittenten (wie z. B. die kartenausgebende Bank) geprüft und autorisiert werden. Hierbei sollen unter anderem diejenigen Transaktionen blockiert werden, die mit hoher Wahrscheinlichkeit betrügerischen Ursprungs sind. Eine spezielle Herausforderung ist dabei der effiziente Umgang mit großen und schnellen Datenströmen und unbalancierten Daten. Der Anteil von Betrug unter den Transaktionen liegt im Promillebereich.

Viele Banken bzw. Dienstleister beschäftigen eigene Abteilungen, die sich mit der Ableitung von Modellen aus historischen Daten befassen, die dann für die automatische Überprüfung zukünftiger Transaktionen eingesetzt werden. Verfahren, die für diese Art der automatischen Überprüfung geeignet sind, können entweder dem manuellen Ansatz folgen oder auf Methoden des Maschinellen Lernens beruhen. Ein manuelles Vorgehen ist starr und aufwendig, da neue Betrugsmuster nur langsam erkannt werden. Es müssen kontinuierlich alte Modelle bewertet werden und ggf. neue erstellt werden. In Bezug auf den Einsatz von Methoden des Maschinellen Lernens für das automatische Blockieren von betrügerischen Kreditkartentransaktionen fokussiert sich die Literatur hauptsächlich auf Black-Box-Modelle [15]. Im Gegensatz zu dem manuellen Ansatz haben diese Methoden zwar meistens eine hohe Genauigkeit und können sich an neue Betrugsmuster automatisch anpassen, jedoch sind diese Modelle nicht interpretierbar und Domain-Experten können ihr Wissen nicht direkt integrieren. Wie im nächsten Abschnitt aufgezeigt wird, lassen sich beide Ansätze in hybriden Verfahren kombinieren und so die Vorteile vereinen.

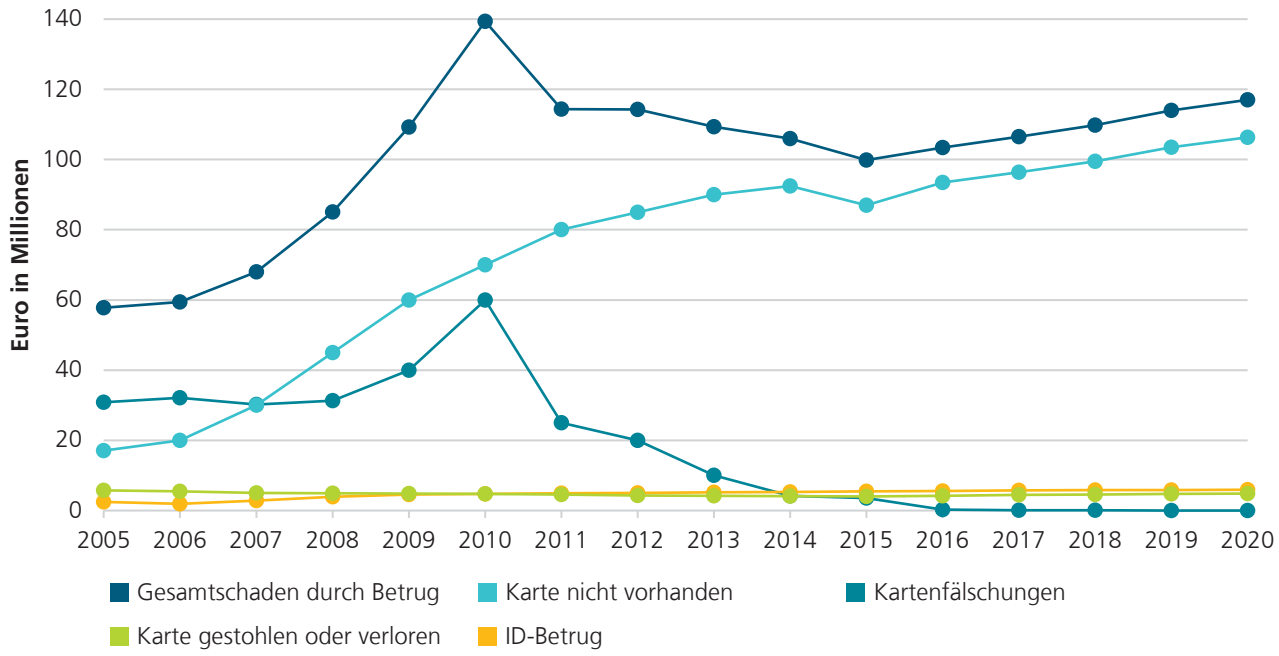

Abbildung 5: Finanzieller Schaden in Deutschland durch Betrug in Millionen Euro von 2005 bis 2020, gestaffelt nach Art des Betrugs [16].

4. Informed Machine Learning und Explainable Artificial Intelligence für die Betrugserkennung

In zwei neuen Forschungsgebieten wird die Entwicklung hybrider Verfahren des Maschinellen Lernens vorangetrieben: **»Informed Machine Learning«** [4] erforscht das Einbinden von Expertenwissen in den Prozess des Maschinellen Lernens, während **»Explainable Artifical Intelligence«** [5] sich mit der Transparenz von Modellen des Maschinellen Lernens beschäftigt. Methoden aus beiden Gebieten werden u. a. schon in der Entwicklung von Sprachdialogsystemen und der Bildverarbeitung erfolgreich eingesetzt. Informed Machine Learning und Explainable Artifical Intelligence bieten ebenso großes Potenzial für die Betrugserkennung, denn sie ermöglichen sowohl die effiziente Identifikation betrügerischer Transaktionen in komplexen Datenstrukturen als auch die Integration von Expertenwissen in diesen Prozess und nachvollziehbare Entscheidungen für die Nutzenden.

Informed Machine Learning bezeichnet alle Verfahren, die es erlauben, bereits verfügbares (Experten-)Wissen in einen Algorithmus des Maschinellen Lernens einfließen zu lassen [4]. Es »beschreibt das Lernen aus einer hybriden Informationsquelle, die aus Daten und Vorwissen besteht. Das Vorwissen ist präexistent und von den Daten getrennt und wird explizit

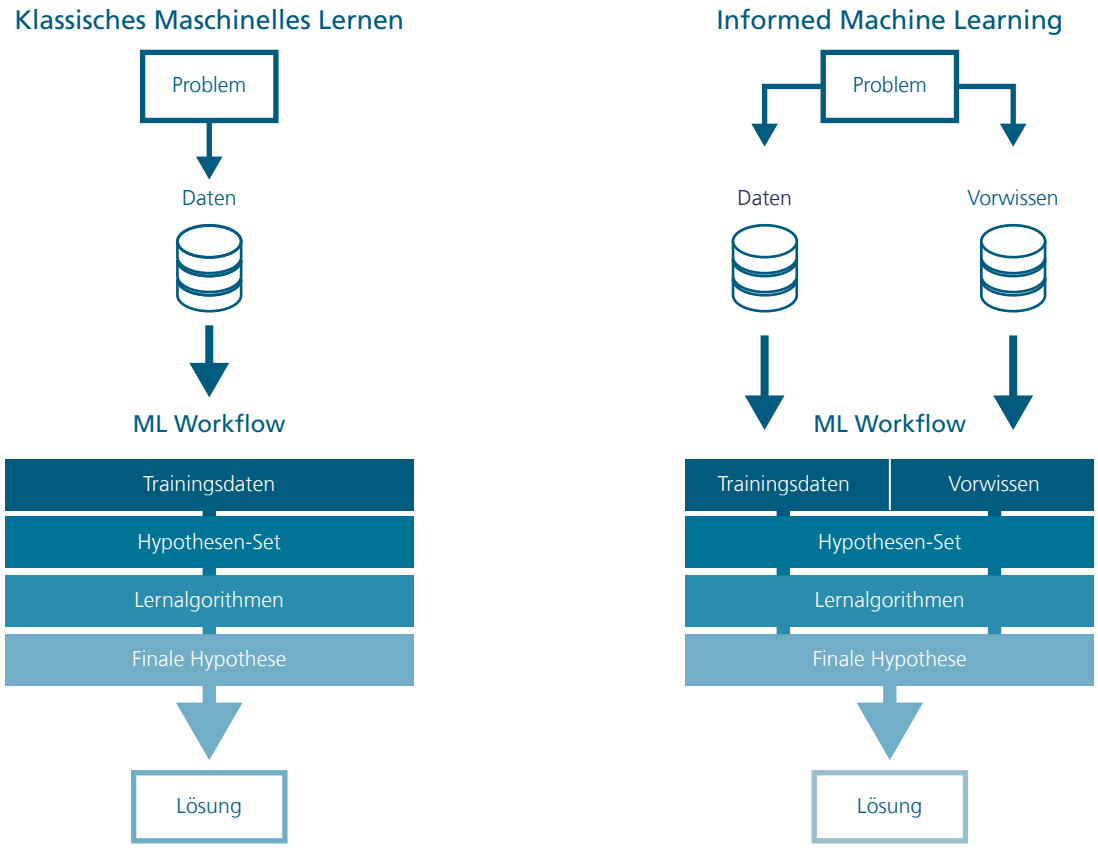

Abbildung 6: Im Vergleich zu klassischen Verfahren des Maschinellen Lernens werden beim Informed Machine Learning Daten und Vorwissen in den Trainingsprozess eingebunden.

in den Workflow des Maschinellen Lernens integriert« (siehe Abb. 6) [4]. Eine Art von Vorwissen, das regelmäßig in den Datenanalyse-Prozess einfließt, ist das Wissen von Expert*innen, die umfangreiche Erfahrung in ihrem jeweiligen Fachgebiet besitzen. In klassischen Verfahren wird dieses Wissen ad-hoc in eine Analyse eingebracht. Demgegenüber integrieren Verfahren des Informed Machine Learnings menschliches Wissen z. B. durch systematische und dynamische Anreicherung der Trainingsdaten. Die am Fraunhofer IAIS entwickelte Software **»Fraunhofer Fraud Detection«** [3] nutzt diesen hybriden Ansatz, indem sie durch Expert*innen definierte Prozessschablonen zur adaptiven Datenanreicherung mit Verfahren zum automatischen Finden von neuen Regeln aus den Daten kombiniert. Dies kapselt einerseits die Komplexität und sorgt dynamisch für optimale Merkmale zur Betrugserkennung.

Explainable Artificial Intelligence bezeichnet all diejenigen Verfahren der Künstlichen Intelligenz, die es Menschen ermöglichen, das »Wie« und »Warum« der Entscheidungsfindung eines Algorithmus zu verstehen. Im Whitepaper des Fraunhofer IAIS zur Vertrauenswürdigkeit von KI heißt es dazu: »KI-Anwendungen, von denen die Rechte und Interessen Dritter betroffen sind, müssen grundsätzlich transparent sein. Transparenz bedeutet die Nachvollziehbarkeit der Arbeitsweise der KI-Anwendung« [17]. Neben der Erfüllung regulatorischer und gesetzlicher Anforderungen dient die Erklärbarkeit von KI-Anwendungen weiteren Zwecken: Sie ermöglicht es, das KI-System zu verbessern und macht es schwieriger, KI-Systeme zu manipulierten Entscheidungen zu verleiten [5]. Nicht zuletzt steigern erklärbare KI-Anwendungen das Vertrauen und die Akzeptanz der Gesellschaft. Hier gibt es drei Arten von Methoden: Ante-hoc-Methoden, die von Natur aus transparent sind, wie Regelsets und Entscheidungsbäume, interaktive Verfahren und Post-hoc-Methoden, die nur eine spezifische Lösung z. B. ein Black-Box-Modell erklären (siehe Abb. 7). Die Betrugserkennung profitiert von Erklärbarkeit, indem entdeckte Muster in einem Review von Fachexpert*innen nachvollzogen und unabhängig von der eingesetzten Methode dokumentiert werden können. Dass solche Verfahren gut für die Betrugserkennung eingesetzt werden können, bestätigen beispielsweise Untersuchungen zum Einsatz von Adversarial Autoencodern für die Erkennung von Internal Fraud [13].

Schließlich lassen sich durch die Kombination von integriertem Vorwissen und erklärbaren Lösungen mit selbstlernenden Verfahren des Maschinellen Lernens die Vorzüge beider Ansätze vereinen. Dadurch ergibt sich ein eindeutiger Vorteil einerseits gegenüber Black-Box-Verfahren und andererseits gegenüber starren, manuell erstellten Regelsets. Für die Betrugserkennung bedeutet dies, dass sich selbstlernende Verfahren dynamisch an veränderte Betrugsmuster in den Daten anpassen können, Expertenwissen eingebunden werden kann und gleichzeitig die Transparenz des Verfahrens den Review der identifizierten Betrugsfälle erleichtert. Das Interesse an derartigen Verfahren belegen die Erfahrungen des Fraunhofer IAIS in Kundenprojekten zu Kreditkartenbetrug und Identitätsbetrug im Onlinehandel.

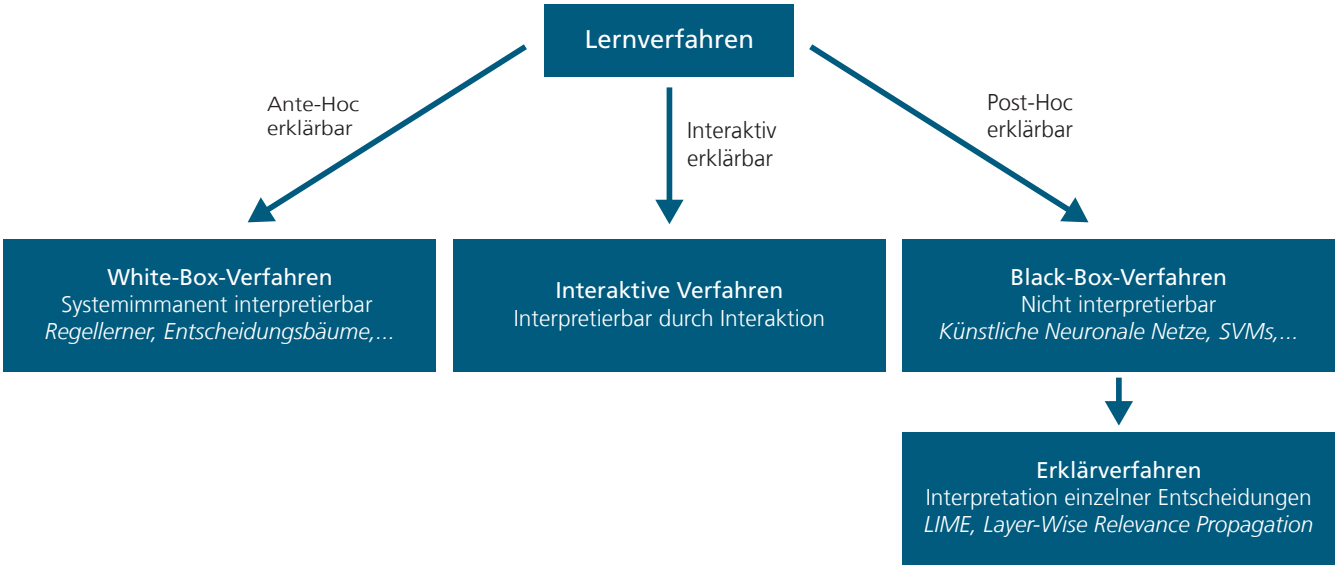

Abbildung 7: Die verschiedenen Erklärmodelle von Explainable Artificial Intelligence im Vergleich [18].

Die bereits erwähnte Software »Fraunhofer Fraud Detection« [3] vereint die Vorzüge beider Ansätze durch Kombination eines transparenten Regellerners mit flexiblen Prozessschablonen zur Integration von Vorwissen. In der Praxis zeigt sich, dass derartige Verfahren zu höherer Effizienz führen. So lernt die Software z. B. anhand aktueller Zahlungsdaten und gängiger Betrugsmuster und findet Regeln, um diese gezielt zu identifizieren. Zudem lassen sich je nach individueller Strategie konkrete Schwerpunkte setzen, um z. B. möglichst viele Betrugsfälle zu erkennen oder möglichst viele Verluste zu vermeiden. Abbildung 8 zeigt, wie die Software konkret vorgeht.

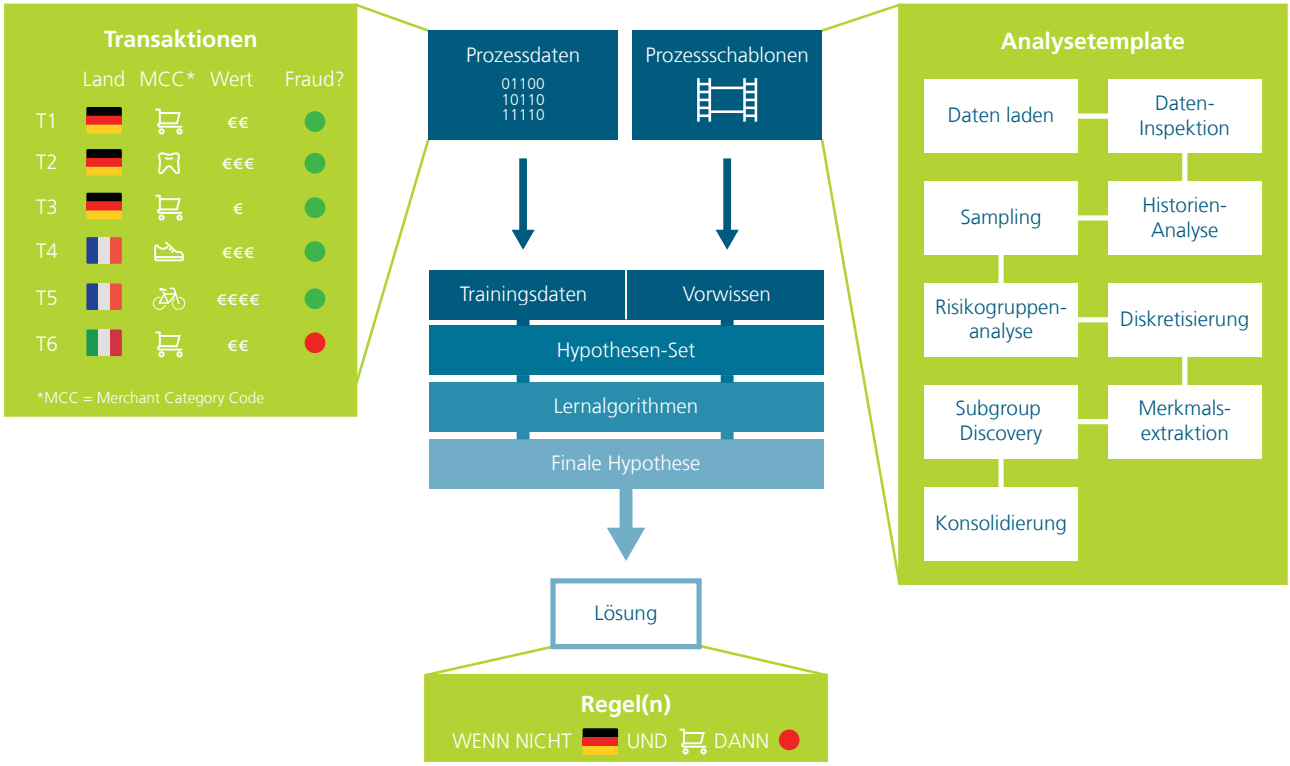

Abbildung 8: Funktionsweise der Software Fraunhofer Fraud Detection am Beispiel typischer Zahlungsdaten und eines Analysetemplates voller Expertenwissen.

5. Fazit und Ausblick

Durch Betrug entstehen große Schäden für betroffene Unternehmen. Trotzdem wird Betrug aktuell nicht oder nur teilweise datengetrieben entdeckt und verhindert. Die Methoden, die dabei zum Einsatz kommen, sind teilweise durch hohen manuellen Aufwand geprägt. Bereits genutzte klassische Ansätze mit Maschinellem Lernen basieren oft auf Black-Box-Methoden, die für den praktischen Einsatz weniger gut geeignet sind. Der hier vorgestellte Ansatz mit Informed Machine Learning und Explainable Artifical Intelligence vereint die Vorzüge der manuellen und der automatischen Betrugserkennung durch Transparenz und hohe Genauigkeit und steigert so die Effizienz der Betrugserkennung.

Mit der Zunahme von digitalen Transaktionen und der technischen Möglichkeiten nimmt auch die Komplexität von Betrugsmustern zu, wie ein aktuelles Beispiel von Deep Fake in Buchhaltungsdaten zeigt [19]. Durch die rasanten Entwicklungen im Bereich der Künstlichen Intelligenz werden die Verfahren algorithmisch immer besser und die Hürde, Methoden des Maschinelles Lernen im Betrieb einzusetzen, sinkt stetig.

Für die Betrugserkennung bedeutet dies, dass es nie leichter war, Maschinelles Lernen einzusetzen. Es ist davon auszugehen, dass die Verbreitung selbstlernender Verfahren für die Betrugserkennung aufgrund der erwähnten Entwicklungen weiter stark zunehmen wird. Aktuell zeigt sich, dass Deliktsbereiche, in denen derlei Verfahren bis jetzt noch nicht umfassend eingesetzt werden konnten, durch Verfahren des Informed Machine Learning und der Explainable Artificial Intelligence künftig stark von Maschinellem Lernen profitieren werden. Neben den konkret dargestellten Szenarien gibt es auch große Potenziale in anderen Use Cases, wie z. B. für die Prävention von Geldwäsche (Anti Money Laundering, AML) oder für die Erkennung von Abrechnungs- oder Versicherungsbetrug. Das Fraunhofer IAIS unterstützt mit Expertise im Bereich Maschinelles Lernen beim Transfer von wissenschaftlichen Erkenntnissen in die Praxis der Betrugserkennung. Für eine frühzeitige Erkennung betrügerischer Handlungen empfehlen sich proaktive Präventions- und Erkennungsmaßnahmen mit fortgeschrittenen Konzepten des Maschinellen Lernens, sog. »Advanced Machine Learning«, auf dem neusten Stand der Technik.

6. Literatur- und Quellenverzeichnis

[1] Aisha Abdallah, Mohd Aizaini Maarof, Anazida Zainal: Fraud detection system: A survey (2016). Journal of Network and Computer Applications, Vol. 68.

[2] Richard J. Bolton, David J. Hand: Statistical Fraud Detection: A Review (2002). Statistical Science, Vol. 17, No. 3, 235–255.

[3] Fraunhofer IAIS: Schützen Sie Ihre Kreditkartenkunden mit unserer Software. URL: https://www.iais.fraunhofer.de/de/geschaeftsfelder/big-data-analytics-and-intelligence/fraunhofer-fraud-detection.html letzter Zugriff am 27.07.2021

[4] Laura von Rueden, Sebastian Mayer, Jochen Garcke, Christian Bauckhage, Jannis Schuecker: Informed Machine Learning - Towards a Taxonomy of Explicit Integration of Knowledge into Machine Learning (2020). ArXiv.

[5] Technische Universität Dresden: Center for Explainable and Efficient AI Technologies (2019). URL: https://www.iais.fraunhofer.de/content/dam/iais/all/doc/CEE-AI_Broschuere_web.pdf letzter Zugriff am 27.07.2021

[6] Gabriel Maciá-Fernández, Pedro García-Teodoro, Jesús Díaz-Verdejo. Fraud in roaming scenarios: An overview (2010). Wireless Communications, IEEE. 16. 88 - 94.

[7] Arjan Reurink: Financial Fraud - A Literature Review (2016). MPIfG Discussion Paper 16/5.

[8] Bundeskriminalamt: Wirtschaftskriminalität - Bundeslagebild 2019 (2020). URL: https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/Lagebilder/Wirtschaftskriminalitaet/wirtschaftskriminalitaet_node.html letzter Zugriff am 27.07.2021

[9] KPMG AG Wirtschaftsprüfungsgesellschaft: Im Spannungsfeld, Wirtschaftskriminalität in Deutschland 2020 (2020). URL: https://hub.kpmg.de/wirtschaftskriminalitaet-in-deutschland-2020-im-spannungsfeld letzter Zugriff am 27.07.2021

[10] PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft: Wirtschaftskriminalität – Ein niemals endender Kampf, PwC's Global Economic Crime and Fraud Survey 2020 (2020). URL: https://www.pwc.de/de/consulting/forensic-services/wirtschaftskriminalitaet-ein-niemals-endender-kampf.pdf letzter Zugriff am 27.07.2021

[11] Association of Certified Fraud Examiners: Anti-Fraud Technology Benchmarking Report (2019). URL: https://www.acfe.com/technology-benchmarking-report.aspx letzter Zugriff am 27.07.2021

[12] Association of Certified Fraud Examiners: Report to the Nations – 2020 Global Study on Occupational Fraud and Abuse (2020). URL: https://www.acfe.com/report-to-the-nations/2020/ letzter Zugriff am 27.07.2021

[13] Marco Schreyer, Timur Sattarov, Christian Schulze, Bernd Reimer, Damian Borth: Detection of Accounting Anomalies in the Latent Space using Adversarial Autoencoder Neural Networks (2019). ArXiv.

[14] Bundesanstalt für Finanzdienstleistungsaufsicht: Big Data trifft auf künstliche Intelligenz – Herausforderungen und Implikationen für Aufsicht und Regulierung von Finanzdienstleistungen (2018). URL: https://www.bafin.de/SharedDocs/Downloads/DE/dl_bdai_studie.html letzter Zugriff am 27.07.2021

[15] A. O. Adewumi, A. A. Akinyelu: A survey of machine-learning and nature-inspired based credit card fraud detection techniques (2017). Int J Syst Assur Eng Manag 8, 937–953.

[16] Fair Isaac Corporation: CNP Fraud Thrives with Escalating E-Commerce (2021) URL: https://www.fico.com/european-fraud/germany letzter Zugriff am 27.07.2021

[17] Fraunhofer IAIS: Vertrauenswürdiger Einsatz von Künstlicher Intelligenz (2019). URL: https://www.iais.fraunhofer.de/content/dam/iais/KINRW/Whitepaper_KI-Zertifizierung.pdf letzter Zugriff am 27.07.2021

[18] Gesellschaft für Informatik: Lexikon: Explainable AI (ex-AI) (2018), URL: https://gi.de/informatiklexikon/explainable-ai-ex-ai letzter Zugriff am 27.07.2021

[19] Marco Schreyer, Timur Sattarov, Bernd Reimer, Damian Borth: Adversarial Learning of Deepfakes in Accounting (2019) ArXiv.

Impressum

Herausgeber
Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
Schloss Birlinghoven
53757 Sankt Augustin

www.iais.fraunhofer.de

Redaktion
Daria Tomala
Silke Loh

Grafik
Angelina Lindenbeck

Layout
Angelina Lindenbeck
Achim Kapusta

Bildnachweise
Titelbild: Samrit 1646/stock.adobe.com (Hintergrund), Thitichaya 715/stock.adobe.com (Schloss)

Kontakt

Fraunhofer-Institut für Intelligente
Analyse- und Informationssysteme IAIS
Schloss Birlinghoven
53757 Sankt Augustin

www.iais.fraunhofer.de

Ansprechpartner:
Dr. Daniel Trabold
Tel.: +49 2241 14-2751
daniel.trabold@iais.fraunhofer.de