

Ersetzt die KI den Abschlussprüfer? Herausforderungen im Rahmen der Aufdeckung von Bilanzdelikten

Prof. Dr. Barbara E. Weißenberger und Leonhard J. Lösse

Executive Summary

Jüngste Fälle wie Wirecard verdeutlichen die Notwendigkeit einer frühzeitigen Aufdeckung von Bilanzmanipulationen im Interesse verschiedener Stakeholder. Wenig überraschend ist daher, dass zahlreiche Ansätze existieren, die mittels Machine Learning versuchen, Unternehmen zu identifizieren, die ihre Finanzberichterstattung manipuliert haben. Entgegen weitläufigen Erwartungen oder Befürchtungen – je nach Perspektive – ist die Aussagekraft solcher Modelle in vielfältiger Hinsicht beschränkt. Insbesondere die Seltenheit von Bilanzfälschungen, die zum Training herangezogen werden können als auch die nicht immer trennscharfe Abgrenzung von Manipulationen limitiert die Leistungsfähigkeit der Modelle. Behindert wird die tatsächliche Verwendung dabei vor allem durch zahlreiche Fehlalarme, sodass der Mehraufwand für weitergehende Analysen regelmäßig den Nutzen übersteigt. Um Modelle dennoch sinnvoll einsetzen zu können, bedarf es Prognosen, die effizient mit fachlicher Expertise beurteilt werden können. Erst die Kombination aus interpretierbaren Prognosen und fachlicher Expertise verspricht, bedeutende Potenziale für eine effiziente Risikobeurteilung realisieren zu können.

Finanzielle Unternehmenstransparenz ist ein Schlüsselement für die Funktionsfähigkeit der Kapitalmärkte. Auf deren Basis können außenstehende Stakeholder informierte Entscheidungen über die Bereitstellung oder auch den Entzug von Eigenkapital treffen, sowie über den Abschluss bzw. die Ausgestaltung von Kooperations-, Kredit-, Arbeits- und Lieferverträgen. Umso schwerer wiegt es, wenn Finanzberichte manipuliert werden und damit ein unzutreffendes Bild der Vermögens-, Finanz- und Ertragslage

zeichnen: Weil ausgewiesenes Vermögen nicht existiert, Wertminderungen und finanzielle Belastungen nicht berücksichtigt oder Umsatzerlöse und mithin Gewinne fingiert sind. Solche deliktischen Bilanzmanipulationen schädigen nicht nur Eigentümer und Vertragspartner, sondern auch das gesamte Wirtschaftssystem. Denn durch den entstehenden Vertrauensverlust in die Aussagekraft der finanziellen Unternehmenstransparenz leidet die Funktionsfähigkeit von Kapital- wie Realgütermärkten.

Strengere Regulierung hilft nur eingeschränkt

Es überrascht deshalb nicht, dass große Bilanzskandale immer wieder zu einer Verschärfung der Regulierung von Rechnungslegung und Unternehmensprüfung führen. So wurde schon 1931 als Reaktion auf zahlreiche Bilanzdelikte der Berufsstand der Wirtschaftsprüfer begründet. Auch auf die gravierende Falschbilanzierung bei dem inzwischen insolventen Zahlungsdienstleister Wirecard reagierte der Gesetzgeber mit dem 2021 in Kraft getretenen Gesetz zur Stärkung der Finanzmarktintegrität (FISG), das strengere Anforderungen als bisher an Unternehmensüberwachung, Abschlussprüferhaftung und Bilanzkontrollen enthält.

Ob sich Bilanzdelikte allerdings allein durch mehr Regulierung bei der Erstellung und Prüfung von Jahresabschlüssen vermeiden lassen, ist fraglich. Denn die Aufdeckung des sich über mindestens vier Jahre hinziehenden Bilanzskandals bei Wirecard wurde lange weder durch den Aufsichtsrat noch durch den Abschlussprüfer oder gar externe Qualitätssicherungs- und Enforcement-Gremien vorangetrieben. Vielmehr waren es zunächst Finanzjournalisten und -analysten, die in den öffentlichen Geschäftszahlen von Wirecard typische Fehlerindikatoren, so genannte ‚red flags‘, identifizierten, wie stark und gleichmäßig wachsende Umsätze mit hohen Margen trotz eines wettbewerbsintensiven Umfelds oder Neukreditaufnahmen bei gleichzeitig umfangreichen Liquiditätspositionen auf ausländischen Treuhandkonten.

Komplementäre datengetriebene Modelle

Dies weist auf das Potenzial einer Lösungsstrategie hin, die deutlich besser konform zu marktwirtschaftlichen Prinzipien geht als weitere regulatorische Eingriffe, nämlich die verbesserte Auswertung veröffentlichter Finanzinformationen durch die Entwicklung und Nutzung geeigneter digitaler Technologien. Ökonomische Akteure werden damit in die Lage versetzt, systematisch aussagekräftige Schlussfolgerungen über die Vertrauenswürdigkeit vorgelegter Abschlüsse zu ziehen und wirtschaftliche Entscheidungen danach auszurichten. Im Mittelpunkt der

Forschungsanstrengungen hierzu steht algorithmisches oder Machine Learning als Teilbereich Künstlicher Intelligenz (KI). Dessen breiter Einsatz ist erst mit der zunehmend einfachen Verfügbarkeit von Software, Rechenleistung und hinreichend großer Datensätze möglich geworden. Vielfältige Anwendungsfälle wie die Betrugsprävention bei Kreditkartentransaktionen zeigen die grundsätzliche Leistungsfähigkeit solcher Technologien.

Stärken und Grenzen von Machine Learning

Eine Stärke algorithmischen Lernens liegt darin, dass keinerlei Regeln für das Erkennen von Bilanzbetrug programmiert werden müssen. Stattdessen entwickelt der Algorithmus diese eigenständig aus der Verarbeitung großer Trainingsdatensätze, die sowohl ordnungsmäßige als auch fehlerhafte Finanzberichte enthalten und so in ein Modell zur Risikoeinschätzung für Bilanzbetrug münden. Explizites Wissen über Zusammenhänge zwischen bestimmten Merkmalen eines Finanzberichts und aktueller oder zu erwartender Bilanzmanipulation ist nicht notwendig, sondern es werden im Gegenteil sogar faktisch unbekannte Zusammenhänge berücksichtigt.

Allerdings hängt die Aussagekraft derartiger Modelle stark von der Qualität der verwendeten Trainingsdaten ab, was gerade für die Vorhersage von Bilanzdelikten herausfordernd ist. Denn diese treten zum einen vergleichsweise selten auf, zum anderen häufig seriell bei denselben Unternehmen. Beides verlangsamt die algorithmischen Lernprozesse der Betrugs-erkennung. Hinzu kommt: Das entwickelte Modell entspricht einer ‚black box‘. Denn wie eine bestimmte Vorhersage im Modell genau entsteht oder welche Faktoren das Ergebnis beeinflussen, ist nicht mehr nachvollziehbar. In vielen Anwendungen ist dies unproblematisch, z.B. wenn Prognosen dennoch hinreichend genau oder Irrtümer nur mit geringen Kosten behaftet sind.

Ineffizienzen in der Anwendung

Bei der Vorhersage von Bilanzbetrug tritt gerade hier ein bisher vernachlässigtes Problem auf, das jetzt erstmals ausführlich von Patrick Vorst von der Universität Maastricht und seinem US-amerikanischen Koautor Messod D. Beneish thematisiert wird (The Cost of Fraud Prediction Errors, erscheint in The Accounting Review): Weil Bilanzdelikte mit großen Schäden einhergehen, werden Modelle typischerweise darauf trainiert, Fälle von Bilanzbetrug (wahr-positive Vorhersagen) möglichst gut zu identifizieren. Das führt statistisch jedoch gleichzeitig zu einer hohen Zahl falsch-positiver Vorhersagen, die im konkreten Fall die Anzahl der tatsächlichen Betrugsfälle um das rund 60fache oder sogar noch mehr übersteigen. Da im Vorhinein unbekannt ist, wann ein solcher ‚Fehlalarm‘ vorliegt, müssen alle Verdachtsfälle untersucht werden.

Dies geht jedoch mit hohen Kosten einher, wenn z.B. Qualitätskontroll- und Enforcement-Gremien deutlich größere Kapazitäten als bisher für Sonderprüfungen aufbauen müssen. Für die von einer falsch-positiven Wertung betroffenen Firmen kommt es zu Reputationsverlust bis hin zu existenzbedrohenden Schäden. Märkte versagen, wenn erfolgversprechende Projekte mit als potenziell deliktisch klassifizierten Firmen aufgrund des entstandenen Vertrauensverlusts nicht mehr umgesetzt werden. Auch Investoren werden geschädigt, wenn die betroffenen Firmen die Kosten von Sonderprüfungen tragen müssen oder vorsorglich gleich ganz aus den Anlageportfolios ausgeschlossen werden. Denn letzteres reduziert zwar das Risiko von Fehlinvestitionen, gleichzeitig wird aber auf die Erträge einer Vielzahl von Unternehmen mit in Wirklichkeit ordnungsmäßiger Finanzberichterstattung verzichtet.

Bedarf interpretierbarer Vorhersagen

Um die Vorteile KI-basierter Modelle dennoch nutzen zu können, wird derzeit an Verfahren für interpretierbares Machine Learning geforscht. Denn weil algorithmische Prognosen nicht aus sich selbst heraus erklärbar sind, werden zusätzliche Schritte notwendig, damit sich Rückschlüsse auf die für eine konkrete Vorhersage relevanten Einflussgrößen ziehen lassen. Dafür werden z. B. die tatsächlich verwendeten Daten gezielt verändert, um die bestimmenden Faktoren der Betrugsvorhersage zu erfassen. Dies ermöglicht auch ein besseres Verständnis dafür, wie ein Modell sich im Bereich konkreter Einzelprognosen verhält.

Erfolg durch Kombination aus Machine Learning und fachlicher Expertise

In der Vorhersage von Bilanzbetrug erwachsen daraus neue und wichtige Anknüpfungspunkte für die menschliche Expertise: Die Genauigkeit der Modelle bleibt unberührt, aber gleichzeitig werden Anhaltspunkte für Plausibilisierung und Begründung in der Klassifikation potenziell deliktischer Firmen geschaffen. Das ist gerade in regulierten Anwendungsfeldern wie der Unternehmenstransparenz bedeutsam, in denen Rechenschaftspflichten eine wichtige Rolle spielen. Gleichzeitig wird die Grenze algorithmischen Lernens deutlich. Gerade weil Bilanzdelikte kontextabhängig sind und auch ein leistungsfähiges Modell nur bekannte Betrugsmuster wiedererkennt, führen Strukturbrüche, z. B. durch veränderte Rechnungslegungsvorschriften, Wirtschafts- und Finanzkrisen oder neuartige Typen von Bilanzdelikten dazu, dass Manipulationen auch in leistungsfähigen Modellen übersehen werden. Solche Änderungsdynamiken müssen weiterhin mit menschlicher Fachexpertise und ganzheitlichem Blick adressiert werden. Gut ausgebildete, unabhängige und kritisch denkende Wirtschaftsprüfer, Rechnungsleger und Controller werden auch in Zukunft gebraucht.

KOMPAKT

Machine Learning als Lösung?

- Machine Learning Modelle identifizieren manipulierte Fälle relativ gut, allerdings mit vielen Fehlalarmen (falsch-positiv).
- Aktuell erscheint die Anwendung der Modelle nicht effizient möglich zu sein.
- Das regulierte Umfeld potenzieller Nutzer erfordert nachvollziehbarere Vorhersagen.
- Die Kombination aus fachlicher Expertise und interpretierbaren datengetriebenen Modellen verspricht den größten Erfolg.

Weiterführende Literatur

Beneish, M.D./Vorst, P. (2022): The Cost of Fraud Prediction Errors. In: *The Accounting Review*, Vol. 97, No. 6, S. 91–121.

Weißenberger, B.E./Förster, G./Bravidor, M./Wesser, M. B. (2019): Wohin führt die Digitalisierung? Auswirkungen auf Wirtschaftsprüfung, Steuerberatung, Finanzfunktion und Hochschullehre. In: *Die Wirtschaftsprüfung*, 72. Jg., Heft 20, S. 1118–1124.

Weißenberger, B.E./ Löße, L.J. (2020): Digitale Unternehmenstransparenz. Wenn Abschlüsse nicht mehr von Menschen, sondern von Algorithmen gelesen werden, in: *Fragestellungen und Perspektiven der Rechnungslegung nach HGB und IFRS. Festschrift zum 65. Geburtstag von Dr. Norbert Lüdenbach*, hrsg. von Jens Freiberg et al., Herne: NWB-Verlag, S. 425–438.



Prof. Dr. Barbara E. Weißenberger
Leonhard J. Löße, M.Sc.

Lehrstuhl für BWL, insb. Controlling und Accounting
 Heinrich-Heine-Universität Düsseldorf
 Universitätsstr. 1 • oeconomicum • 40225 Düsseldorf
 barbara.weissenberger@hhu.de • www.controlling.hhu.de

In unseren Policy-Briefs erläutern wir knapp und prägnant Kernergebnisse aus der betriebswirtschaftlichen Forschung und darauf aufbauend Umsetzungsvorschläge für die Praxis. Damit wollen wir in Unternehmen und Gesellschaft neue Handlungsimpulse geben.

Der vorliegende Policy-Brief wurde in leicht veränderter und gekürzter Form am 19. September 2022 in der Frankfurter Allgemeinen Zeitung in der Rubrik „Der Betriebswirt“ veröffentlicht.

Februar 2023

