



# Machine Learning und empirische Rechnungslegungsforschung: Einige Erkenntnisse und offene Fragen

Thorsten Sellhorn

Eingegangen: 22. Dezember 2019 / Angenommen: 12. März 2020 / Online publiziert: 23. März 2020  
© Der/die Autor(en) 2020

**Zusammenfassung** Im Zuge der digitalen Transformation von Wirtschaft und Gesellschaft ergeben sich zunehmend Anwendungsfelder für Ansätze des maschinellen Lernens nicht nur in der Rechnungslegungspraxis, sondern auch in der betriebswirtschaftlichen Forschung auf diesem Gebiet. Der nachfolgende Beitrag diskutiert selektiv einige Einsatzgebiete von Machine-Learning-Ansätzen in der Unternehmensberichterstattung, der Abschlussprüfung sowie der Unternehmensanalyse und -bewertung. Zudem werden aktuelle und potenzielle Anwendungen in der empirischen Forschung aufgezeigt sowie limitierende Faktoren diskutiert.

**Schlüsselwörter** Digitalisierung · Rechnungslegung · Abschlussprüfung · Maschinelles Lernen · Prognose · Kausaleffekte · Regulierung

**JEL-Klassifikation** M41 · O33

---

T. Sellhorn (✉)  
LMU München, München, Deutschland  
E-Mail: [sellhorn@bwl.lmu.de](mailto:sellhorn@bwl.lmu.de)

## Machine Learning and Empirical Accounting Research: Some Findings and Open Questions

**Abstract** The digital transformation of economy and society increasingly fosters machine learning applications in a wide range of areas. Accounting practice and empirical accounting research are no exception. This article selectively discusses machine learning approaches in corporate reporting, auditing, analysis, and valuation. In addition, current and potential applications in empirical research are shown, and limiting factors are discussed.

**Keywords** Machine learning · Financial reporting · Digital transformation · Causal effects · Prediction

### 1 Einleitung

Die digitale Transformation, so ist in der rasant wachsenden populärwissenschaftlichen Literatur zum Thema zu lesen, biete einerseits immense Chancen für diejenigen, die sich schnell anpassen, gehe aber andererseits auch mit großen Risiken und Bedrohungen für die Zauderer einher.<sup>1</sup> Dieser Markt spielt mit der Sorge vieler Wirtschaftsakteure, wichtige Trends und Weichenstellungen zu verpassen und damit zum Opfer derjenigen zu werden, die schneller sind. Dabei geht es u. a. um technische Grundlagen und Visionen, z. B. bei künstlicher Intelligenz und insbesondere maschinellem Lernen<sup>2</sup>, die wachsende Bedeutung (und bisher weitgehende bilanzielle Vernachlässigung) immaterieller Vermögenswerte<sup>3</sup>, Plattform-Geschäftsmodelle und Monopolbildung<sup>4</sup> sowie Automatisierung und ihre sozialen Folgen.<sup>5</sup> Teilweise wird suggeriert: Die digitale Transformation setzt Wirtschaft und Gesellschaft unter Zugzwang. Wir gestalten sie nicht aktiv, sondern sie droht uns zu geschehen.<sup>6</sup> Auch in der Schmalenbach-Gesellschaft haben sich in den letzten Jahren nahezu alle Arbeitskreise diesem Thema gewidmet; eine Übersicht bieten Krause und Pellens (2017).<sup>7</sup>

Der nachfolgende Beitrag diskutiert eine subjektive und damit selektive Auswahl dieser Implikationen auf den Gebieten der Unternehmensberichterstattung, der Abschlussprüfung sowie der Unternehmensanalyse und -bewertung und wendet sich

<sup>1</sup> Vgl. etwa Brynjolfsson und McAfee (2014), Ford (2015) oder Kelly (2017).

<sup>2</sup> Vgl. etwa Agrawal et al. (2018), die Machine-Learning-Algorithmen als Hochleistungs-Prognosemodelle sehen.

<sup>3</sup> Vgl. etwa Lev und Gu (2016); Lev (2019); Lev (2018).

<sup>4</sup> Vgl. etwa McAfee und Brynjolfsson (2017).

<sup>5</sup> Vgl. etwa Schwab (2017).

<sup>6</sup> So auch Ballwieser und Hachmeister (2019, S. 13): „empfundene Bedrohung“.

<sup>7</sup> Vgl. mit Bezug zum Finanz- und Rechnungswesen insb. die Beiträge der Arbeitskreise Digital Finance (S. 243–265), Finanzierung (S. 267–283), Finanzierungsrechnung (S. 285–300), Externe Unternehmensrechnung (S. 301–317), Externe und Interne Überwachung der Unternehmung (S. 321–336), Integrated Reporting (S. 337–353) und Corporate Governance Reporting (S. 355–373) in Krause und Pellens (2017).

damit an ein sowohl praktisch als auch akademisch interessiertes Fachpublikum. Der Fokus liegt dabei auf dem maschinellen Lernen (*Machine Learning*, nachfolgend kurz: ML), einem Bündel technischer Ansätze auf dem Gebiet der künstlichen Intelligenz, die ein wesentlicher Treiber der digitalen Transformation sind.<sup>8</sup> ML-Algorithmen sind statistische Modelle, die in großen Mengen strukturierter und unstrukturierter (Massen-) Daten Muster erkennen können.<sup>9</sup> Überwachte ML-Algorithmen konstruieren mathematische Modelle auf der Basis von Trainingsdaten (gegebene Paare von Ein- und Ausgaben), um Vorhersagen, Entscheidungen oder Klassifikationen zu treffen, ohne für diese Aufgabe explizit programmiert worden zu sein. Unüberwachte ML-Algorithmen erzeugen statistische Modelle, die Eingaben beschreiben und Kategorien sowie Zusammenhänge erkennen. So teilen etwa sog. Clustering-Verfahren die Eingabedaten in mehrere Kategorien (z. B. thematische Blöcke in Textdokumenten) ein, die sich durch bestimmte Muster voneinander unterscheiden. ML-Algorithmen werden in einer Vielzahl von Anwendungen eingesetzt, in denen es schwierig oder unmöglich ist, einen konventionellen Algorithmus zur effektiven Ausführung der Aufgabe zu entwickeln.<sup>10</sup> Auch auf dem Gebiet der Rechnungslegung wirken ML-Ansätze tief in die Praxis hinein und werden zu einem zunehmend wichtigen Gegenstand der Rechnungslegungsforschung (nachfolgend kurz: RLF). Sie verändern zudem als wichtige Quelle neuer Daten und empirischer Verfahren auch die Methodik ebendieser RLF selbst.<sup>11</sup>

Der Beitrag ist wie folgt aufgebaut. Abschn. 2 umreißt die Erkenntnisinteressen der empirischen RLF und diskutiert Kontexte, in denen ML-Ansätze eine Rolle spielen können. Abschn. 3 betrachtet, wie ML-Ansätze die in der Rechnungslegung eingesetzten Systeme und Prozesse verändern und damit zu wichtigen Erfahrungsobjekten der empirischen RLF werden. In Abschn. 4 werden ML-Ansätze als Instrumente empirischer RLF diskutiert, die in den Dienst deskriptiver, kausaler und prognoseorientierter Fragen gestellt werden können. Abschn. 5 erörtert die vielfach noch mangelnde Nachvollziehbarkeit und Transparenz algorithmischer Prognosen, Klassifikationen und Entscheidungen als möglichen limitierenden Faktor für das

---

<sup>8</sup> Andere Facetten, die hier nicht betrachtet werden, sind z. B. die digitale Erfassung und zentrale Bündelung analoger Daten aus heterogenen Vorkontexten, die Prozessautomatisierung durch sog. *Software Bots*, also regelbasierte Softwareprogramme, denen Expertenwissen fest einprogrammiert wurde und die repetitive Tätigkeiten ausführen können (auch: *Robotic Process Automation*, RPA), oder die Blockchain-Technologie.

<sup>9</sup> Vgl. zu diesem Absatz [de.wikipedia.org/wiki/Maschinelles\\_Lernen](https://de.wikipedia.org/wiki/Maschinelles_Lernen) und [en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning).

<sup>10</sup> Dazu zählen „automatisierte Diagnoseverfahren, Erkennung von Kreditkartenbetrug, Aktienmarktanalysen, Klassifikation von Nukleotidsequenzen, Sprach- und Texterkennung“ ([de.wikipedia.org/wiki/Maschinelles\\_Lernen](https://de.wikipedia.org/wiki/Maschinelles_Lernen)) sowie Bilderkennung, E-Mail-Filterung und autonome Systeme (vgl. [en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)).

<sup>11</sup> Nicht Gegenstand dieses Beitrags ist der Einsatz von ML-Algorithmen im Rahmen der in der Rechnungslegung abzubildenden Sachverhalte und (Kern-) Geschäftsmodelle (z. B. Google-Suchalgorithmus). Als wesentliche Werttreiber vieler Unternehmen stellen diese immateriellen Vermögenswerte die Rechnungslegung und Unternehmensbewertung jedoch vor enorme Herausforderungen. Vgl. hierzu etwa Lev und Gu (2016), Lev (2018, 2019) sowie Ballwieser und Hachmeister (2019). Ausgeklammert bleiben hier zudem die ebenfalls vielfältigen Einsatzgebiete und Implikationen im internen Rechnungswesen (z. B. *Predictive Analytics*) sowie auf dem Gebiet der Besteuerung (z. B. RPA; vgl. Cooper et al. 2019).

weitere Vordringen von ML-Algorithmen. Fazit und Ausblick (Abschn. 6) beschließen den Beitrag.

## 2 Konzeptioneller Rahmen und Forschungsfragen

### 2.1 Erkenntnisziele der empirischen Rechnungslegungsforschung (RLF)

Wie kann sich die RLF als Forschung auf den Gebieten der Unternehmensberichterstattung<sup>12</sup>, der Abschlussprüfung sowie der Unternehmensanalyse und -bewertung dem Vordringen von ML-Ansätzen empirisch nähern bzw. diese zur Erweiterung und Verfeinerung der eigenen Methodik nutzen? Empirische Studien auf dem Gebiet der Rechnungslegung verfolgen die Erkenntnisziele Deskription (Beschreibung), Erklärung oder Prognose. *Deskriptive Studien* sammeln systematisch Fakten über ein relevantes Phänomen und/oder dessen statistischen Zusammenhang mit anderen Phänomenen. Hierzu gehört deskriptive Inferenz als der Versuch, Informationen über unbeobachtete Fakten (z. B. eine Population) aus beobachteten Fakten (z. B. aus einer Stichprobe) abzuleiten.<sup>13</sup> Erste Interviewstudien beschreiben etwa den Einsatz von *Robotic Process Automation* (RPA; vgl. Cooper et al. 2019) oder *Data Analytics* (Gepp et al. 2018) in Rechnungswesen und Abschlussprüfung. Auch die Leistungsfähigkeit von ML-Algorithmen für die Prognose von Rückstellungen ist mittlerweile Gegenstand erster wissenschaftlicher Studien (vgl. etwa Ding et al. 2020).

Nach Watts und Zimmerman (1986, S. 2), „the objective of accounting theory is to explain and predict accounting practice“. Entsprechend kommen in der RLF erklärende und Prognoseansätze vor. *Erklärende Studien* testen kausale Theorien, wollen also Phänomene als Ursachen und Wirkungen oder Determinanten und Konsequenzen durch kausale Inferenz miteinander verbinden.<sup>14</sup> Die Herausforderungen der kausalen Inferenz in nicht-experimentellen empirischen Studien, einschließlich der empirischen Rechnungslegungsforschung, sind bekannt (z. B. Larcker und Rusticus 2007; Gassen 2014; Gow et al. 2016). Kausale Erkenntnisziele können sich entweder auf einen untersuchten primären Effekt (z. B. die Determinanten des Einsatzes von ML-Algorithmen in der Rechnungslegung oder ihre Auswirkungen auf deren Qualität) oder auf einen sekundären, moderierenden Effekt (z. B. die Auswirkungen verschiedener Governancestrukturen auf die Determinanten des Einsatzes von ML-Algorithmen in der Rechnungslegung) beziehen. Ein bedeutender Teil der empirischen RLF ist explizit an solchen sekundären kausalen Effekten interessiert, z. B. an den Auswirkungen von Regulierung. So ist etwa anzunehmen, dass neue

<sup>12</sup> Verstanden als Abbildung der finanziellen Situation des Unternehmens nach außen in Form des (geprüften) Jahres- bzw. Konzernabschlusses und -lageberichts sowie ggf. weiterer verpflichtender und freiwilliger Maßnahmen der finanziellen und zunehmend auch nicht-finanziellen Berichterstattung.

<sup>13</sup> Vgl. etwa King et al. (1994, S. 34). Shmueli (2010, S. 291) charakterisiert deskriptive Modellierung als die kompakte Zusammenfassung oder Darstellung einer Datenstruktur.

<sup>14</sup> Bloomfield et al. (2016, S. 351) sprechen in diesem Zusammenhang von *Attribution*: „Having identified an association, researchers then seek to attribute the association to the causal factors they have specified.“.

Transparenzvorschriften wie etwa die Verpflichtung bestimmter kapitalmarktorientierter Unternehmen zur Abgabe einer nicht-finanziellen Erklärung durch das CSR-Richtlinie-Umsetzungsgesetz oder die Einführung von IFRS 16 *Leasingverhältnisse* die für die Rechnungslegung relevanten Prozesse und IT-Systeme beeinflussen dürften.

Schließlich analysieren *Prognosestudien*, ob die Variation in Variablen  $X$  zukünftige Werte von  $X$  oder anderen Variablen  $Y$  vorhersagt (d. h. mit diesen assoziiert ist). Shmueli (2010, S. 291) versteht Prognosemodellierung als den Prozess der Anwendung eines statistischen Modells oder Data-Mining-Algorithmus auf Daten zum Zwecke der Vorhersage neuer oder zukünftiger Beobachtungen. ML-Algorithmen sind prädestiniert für den Einsatz auf diesen Forschungsgebieten.

Neben diesen eher klassischen sozialwissenschaftlichen Forschungsansätzen befindet sich insbesondere im Kontext der digitalen Transformation der Wirtschaft auch in der RLF zunehmend der konstruktions- bzw. gestaltungsorientierte Forschungsansatz (*Design Science*; vgl. grundlegend Hevner et al. 2004) im Vordringen. Diese Form der anwendungsorientierten Forschung ist in der Wirtschaftsinformatik seit langem verbreitet, deutlich weniger (bisher) jedoch in der RLF. Ziel der Design Science ist die Entwicklung und Evaluation von Artefakten (im hier interessierenden Kontext im Wesentlichen ML-Algorithmen), die dazu in der Lage sind, ein klar definiertes Problem effektiv und effizient zu lösen.<sup>15</sup>

## 2.2 Machine Learning in der empirischen RLF

Zur Strukturierung der Elemente von empirischen Studien in der Rechnungslegung wird hier das *Predictive Validity Framework* nach Libby (sog. „Libby Boxes“; Libby 1981) genutzt (vgl. auch Libby et al. 2002) und die Frage gestellt, an welchen Stellen dieses gedanklichen Rahmens der Einsatz von ML-Algorithmen eine Rolle spielen kann.

In der (empirischen) RLF interessieren oft kausale Fragen: Wirkt eine Größe  $X$  kausal auf eine Größe  $Y$ ? Eine valide Antwort auf Fragen dieses Typs hat mehrere Voraussetzungen. Erstens wird eine kausale Theorie in Form eines Erklärungsmodells benötigt, das eine kausale Beziehung zwischen zwei Größen postuliert.<sup>16</sup> Zum einen könnte der Einsatz von ML-Algorithmen in Rechnungslegung und Abschlussprüfung als abhängige Variable interessieren. Was bestimmt z. B. den Einsatz von ML-Algorithmen in der Finanzorganisation von Unternehmen oder die Investitionen von Wirtschaftsprüfungsgesellschaften in ML-gestützte datenanalytische Prüfungstechnologien? Wirkt Regulierung hier als investitionshemmend (etwa weil bestehende Prüfungsstandards datenanalytische Ansätze noch nicht hinreichend decken) oder

<sup>15</sup> Eine Diskussion von Design-Science-Ansätzen im Bereich der Abschlussprüfungsforschung findet sich in Kogan et al. (2019), der das Sonderheft „Data Analytics Forum“ der Fachzeitschrift „Accounting Horizons“ 2019 einleitet.

<sup>16</sup> Leicht abweichende Ansprüche sind an Prognosefragen vom Typ „Lassen sich zukünftige Ausprägungen einer Größe  $Y$  mithilfe heutiger Ausprägungen einer Größe  $X$  vorhersagen?“ zu stellen. Hier wird nicht notwendigerweise eine Kausalbeziehung unterstellt, sondern lediglich ein stabiler statistischer Zusammenhang, welcher natürlich bei Vorliegen einer kausalen Theorie überzeugender ist.

-fördernd (etwa weil neue Rechnungslegungsstandards IT-Investitionen *de facto* erzwingen)?

Zum anderen könnte der ML-Einsatz selbst als möglicher Einflussfaktor (unabhängige Variable) auf andere interessierende Größen im Fokus stehen. So wäre im Bereich der Unternehmensberichterstattung zu fragen, wie sich ein wesentlich auf ML-Algorithmen beruhendes Geschäftsmodell (z. B. der Google-Suchalgorithmus) im Vergleich zu „herkömmlichen“ Geschäftsmodellen auf die Entscheidungsnützlichkeit der Rechnungslegung (die wesentliche immaterielle Werte nach wie vor ausblendet) oder auf die Berichterstattung über wesentliche Leistungsindikatoren im Lagebericht (Fokus auf nichtfinanzielle Größen wie Daten und Mitarbeiter-Knowhow) auswirkt. Was passiert, wenn Unternehmen in ihren Finanz- und Rechnungslegungsprozessen zunehmend ML-gestützte Verfahren anwenden? Werden die Informationssysteme dadurch „feiner“, steigt die „Qualität“ der Rechnungslegung oder sinken die Kosten freiwilliger Offenlegung durch neue Informationen und geringere Bereitstellungskosten? Im Kontext der Abschlussprüfung wäre zu untersuchen, wie der Einsatz von ML-gestützten Datenanalysen in der Abschlussprüfung Größen wie Prüfungshonorare, Prüfungsqualität, die Struktur des Prüfermarktes oder Arbeitsangebot, -nachfrage und -bedingungen beeinflusst.

Zweitens erfordern valide empirische Studien die überzeugende empirische Operationalisierung der unabhängigen und abhängigen Variablen (Konstruktvalidität). Bildet der ML-Einsatz von Wirtschaftsakteuren selbst den Forschungsgegenstand, stehen hier ggf. Datenverfügbarkeitsprobleme im Weg, da dieser von Externen nur begrenzt beobachtbar ist. Jedoch mag der Einsatz von ML-Algorithmen weniger selbst Gegenstand des Forschungsinteresses sein, sondern dem Forscher bei der empirischen Messung anderer interessierender Phänomene helfen. So bieten etwa ML-Ansätze im Bereich der automatisierten Textanalyse enorme Potenziale, die bereits an anderer Stelle beschrieben wurden (vgl. etwa Loughran und McDonald 2016; Menacher et al. 2017; Sellhorn et al. 2019 sowie unten Abschn. 4.1).

ML-gestützte Verfahren helfen, drittens, auch bei der Schätzung der statistischen Assoziation zwischen interessierenden Variablen (Validität der statistischen Inferenz). Die automatisierte Extraktion von Daten, die früher aufwändig handerhoben werden mussten, durch *Web Scraping* und automatisierte Textanalyse ermöglicht das Arbeiten mit größeren Stichproben und steigert damit die Trennschärfe bzw. Teststärke statistischer Analysen (*Statistical Power*).

Besonderes Potenzial entfalten ML-Algorithmen, viertens, im Kontext von Prognosemodellen, z. B. für Ereignisse wie Insolvenzen oder wesentliche Bilanzierungsfehler (vgl. dazu unten Abschn. 3.2). Diese Eigenschaft als *Prediction Machines* (Agrawal et al. 2018) macht sie aber auch dort nützlich, wo im Kontext von kausalen Fragestellungen kontrafaktische Szenarien zu schätzen sind (interne Validität). So sind beispielsweise bei Ereignisstudien Annahmen darüber notwendig, wie sich der Aktienkurs des betrachteten Unternehmens in Abwesenheit des interessierenden Informationsereignisses entwickelt hätte. Hier können ML-gestützte Prognosemodelle zum Einsatz kommen (vgl. etwa Arnott et al. 2019). Bei der Identifikation von kausalen Effekten helfen ML-Algorithmen potenziell auch durch Ansätze zur automatisierten Auswahl von Kontrollvariablen für Regressionsanalysen (vgl. zu den ökonomischen Einsatzmöglichkeiten von ML-Ansätzen Abschn. 4).

### 3 Machine Learning als Erfahrungsobjekt der RLF

Der Einsatz von ML-Ansätzen in der Rechnungslegung, inkl. Unternehmensberichterstattung, Abschlussprüfung sowie Unternehmensanalyse und -bewertung, geht derzeit, soweit ersichtlich, über erste Pilotprojekte noch nicht sehr weit hinaus.<sup>17</sup> Dennoch lassen sich weitreichende Potenziale erkennen. In diesem Abschnitt werden diese Potenziale anhand dreier Forschungsbereiche beispielhaft illustriert.

#### 3.1 ML-gestützte Schätzungen

ML-Ansätze sind prinzipiell einsetzbar für die Verbesserung von Prognosen und Schätzungen für Bilanzansatz und -bewertung sowie deren Plausibilisierung im Kontext der Abschlussprüfung. Solche Schätzungen gehen in vielfältige Bilanz- und GuV-Positionen ein. So enthalten die Anhänge von IFRS-Konzernabschlüssen Abschnitte, in denen das bilanzierende Unternehmen wesentliche Schätzungen und Beurteilungen durch das Management auflistet. Der Abschlussprüfer kommentiert diese regelmäßig in seinen Ausführungen zu besonders wichtigen Prüfungssachverhalten (sog. *Key Audit Matters*, KAMs) im Bestätigungsvermerk, und Enforcement-Institutionen wie die Deutsche Prüfstelle für Rechnungslegung (DPR) und die European Securities Markets Authority (ESMA) machen diese regelmäßig zum Gegenstand ihrer Prüfungen.<sup>18</sup> Zu diesen wesentlichen Schätz- und Ermessenssachverhalten zählen u. a. der Goodwill und dessen Werthaltigkeitstest, aktivierte Entwicklungskosten, die Nutzungsdauern von Vermögenswerten des Anlagevermögens, Rückstellungen für leistungsorientierte Pensionsverpflichtungen und Prozessrisiken, die beizulegenden Zeitwerte von nicht auf aktiven Märkten gehandelten Finanzinstrumenten und als Finanzinvestition gehaltenen Immobilien sowie Posten aus der Umsatzrealisierung bei langfristiger Auftragsfertigung.

Aufgrund ihrer hohen ökonomischen Bedeutung und den mit ihrer Bewertung verbundenen bilanzpolitisch nutzbaren Ermessensspielräumen kommt diesen Sachverhalten in Forschung und Praxis hohe Aufmerksamkeit zu. Hieraus ergibt sich aus der Sicht von Abschlussprüfern (vgl. Marten et al. 2015, S. 301) und Adressaten, aber durchaus auch aus Erstersicht, ein Bedarf nach unabhängigen, unverzerrten Schätzalgorithmen als Ersatz oder zumindest Beurteilungsmaßstab für Managementschätzungen (vgl. auch Ding et al. 2020). ML-Algorithmen kommen hier als Wegbereiter einer stärkeren Objektivierung von Schätzungen in Betracht. Während frühere Ansätze computergestützter Entscheidungsunterstützung primär auf regelbasierten (Experten-) Systemen aufbauten, wird das größte Potenzial in der heutigen Zeit sinkender Informationsverarbeitungskosten und steigender (Massen-) Datenverfügbarkeit zunehmend in selbstlernenden Systemen gesehen, die auf der Basis von Trainingsdaten ein Prognosemodell entwickeln und dessen Schätzfehler minimieren. Mithilfe dieses Prognosemodells werden dann auf der Basis von verfügbaren

<sup>17</sup> Für das Rechnungswesen deutscher Unternehmen vgl. etwa KPMG (2017, 2018 und 2019).

<sup>18</sup> Vgl. zuletzt DPR, Prüfungsschwerpunkte 2020, veröffentlicht am 18.11.2019; [https://www.frep.info/docs/pressemitteilungen/2019/20191118\\_pm.pdf](https://www.frep.info/docs/pressemitteilungen/2019/20191118_pm.pdf).

Inputdaten Schätzungen zukünftiger Ausprägungen der interessierenden Größen generiert.

Das Training von ML-Algorithmen für diese Aufgabe erfordert die Verfügbarkeit der jeweiligen Schätzungen bzw. Soll-Werte sowie der anschließenden Realisierungen bzw. Ist-Werte. Während diese Werte unternehmensintern regelmäßig vorliegen dürften,<sup>19</sup> stellt ihre aus externer Sicht i. d. R. fehlende Beobachtbarkeit eine Hürde für die RLF dar. Für die Entwicklung und Überprüfung von ML-Schätzalgorithmen sind empirische RW-Forscher daher auf die wenigen Fälle beschränkt, in denen die Datenverfügbarkeit hinreichend gut ist, oder sie sind auf Unternehmenskooperationen angewiesen, die ihnen Zugang zu internen Daten verschaffen.

Ein Beispiel für Forschung auf diesem Gebiet ist der Beitrag von Ding et al. (2020), die mithilfe verschiedener ML-Algorithmen die Höhe von Rückstellungen und Schadenaufwendungen im Versicherungsgeschäft schätzen.<sup>20</sup> Die algorithmischen Schätzungen erweisen sich im Vergleich zu Managementschätzungen – und insbesondere in Kombination mit diesen – als insgesamt genauer und weniger stark mit bilanzpolitischen Anreizen assoziiert.

### 3.2 ML-gestützte Peer-Group-Bestimmung

Der Bestimmung von Peer Groups, also von Unternehmen mit vergleichbarem Rendite-Risiko-Profil, kommt nicht nur im Kontext der Unternehmensberichterstattung, sondern auch in der Abschlussprüfung sowie der Unternehmensanalyse und -bewertung eine große Bedeutung zu. Unternehmen suchen Benchmarks für die Herleitung von Renditezielen und den anschließenden Soll-Ist-Vergleich sowie für Zwecke der relativen Performancebeurteilung, Abschlussprüfer betrachten gegenwärtige und potenzielle Wettbewerber ihrer Mandanten bei der Risikoanalyse im Rahmen der Entwicklung der Prüfstrategie (vgl. Marten et al. 2015, S. 276; Kogan et al. 2019) und in der Unternehmensanalyse und -bewertung werden Peer Groups für die Ermittlung von kapitalmarktbasierten Beta-Faktoren, Kapitalkostensätzen und Bewertungs-Multiples benötigt (vgl. Dierkes und Sumpelmann 2019). In der RLF werden Peer Groups u. a. als Kontrollgruppe für die Abschätzung kausaler Effekte herangezogen (vgl. Ding et al. 2020).

Herkömmliche Ansätze wie der Rückgriff auf Branchenklassifikationen wie die US-amerikanische *Standard Industrial Classification* (SIC) oder das *North American Industry Classification System* (NAICS) sind aufgrund ihrer Statik nicht geeignet, dynamische Marktentwicklungen zu berücksichtigen, und eignen sich zudem häufig nicht zweckübergreifend. In der Literatur finden sich daher mittlerweile verschiedene Ansätze, die den Prozess der Peer-Group-Bestimmung mithilfe von ML-Algorithmen zu automatisieren, zu flexibilisieren sowie präziser und intersubjektiv nachvollziehbarer zu machen suchen. So verwenden Ding et al. (2019) K-Median-

<sup>19</sup> Ausnahmen sind z. B. die (auch unternehmensintern nicht ohne Weiteres *ex-post* beobachtbaren) Nutzungsdauern von Vermögenswerten des Anlagevermögens oder erzielbaren Beträge von goodwill-tragenden zahlungsmittelgenerierenden Einheiten.

<sup>20</sup> Zum Einsatz kommen lineare Regression (Minimierung des quadrierten Prognosefehlers), *Random Forests* (Entscheidungsbäume), *Gradient Boosting Machine* sowie künstliche neuronale Netze.



Clustering, einen automatisierten Peer-Selection-Ansatz, der die Ähnlichkeit der Finanzkennzahlen von Unternehmen mit Hilfe eines unbeaufsichtigten maschinellen Lernalgorithmus (*Unsupervised ML*) erfasst. Der Clustering-Algorithmus teilt Beobachtungen dergestalt in sinnvolle Gruppen ein, dass die Beobachtungen in derselben Gruppe einander ähnlich sind und sich von Beobachtungen in anderen Gruppen unterscheiden. Auf diese Weise werden Unternehmen mit ähnlichen Finanzkennzahlen identifiziert, die annahmegemäß auf ähnliche Geschäftsmodelle und -tätigkeiten hindeuten.

Kennzahlenanalyse hat in der RLF eine lange Tradition, etwa um die Insolvenzwahrscheinlichkeit von Unternehmen zu beurteilen (vgl. schon Beaver 1966; Altman 1968; zum Einsatz künstlicher neuronaler Netze zu diesem Zweck vgl. schon Baetge 1998). Zudem werden Finanzkennzahlenmodelle verwendet, um wesentliche Bilanzierungsfehler (*Misstatements*) zu erkennen (vgl. etwa Beneish 1997). Dementsprechend testen Ding et al. (2020) die Leistungsfähigkeit ihrer Machine-Learning-gestützten Peer-Group-Bestimmung in diesen beiden Anwendungsbereichen und können zeigen, dass ihr Ansatz bestehende Insolvenz- und Misstatement-Prognosemodelle stärker verbessert als herkömmliche, statische Verfahren der Peer-Group-Bestimmung, die auf Branchenzugehörigkeit und Unternehmensgröße basieren.

Auf einem ähnlichen ML-Algorithmus wie bei Ding et al. (2020) basieren die dynamischen Ansätze zur Peer-Group-Bestimmung von Hoberg und Phillips (2010, 2016). Im Unterschied zu Ding et al. (2020) verwenden Hoberg und Phillips jedoch keine Finanzkennzahlen, sondern bestimmen die Ähnlichkeit der Produktangebote von Unternehmen durch die automatisierte Analyse von textbasierten Produktbeschreibungen in den Geschäftsberichten (Form 10-K) US-amerikanischer Unternehmen. So schließen sich Unternehmen, deren Produktbeschreibungen sich stärker ähneln, z. B. häufiger und erfolgreicher in M&A-Transaktionen zusammen (Hoberg und Phillips 2010). Zudem erklären die so gebildeten Peer Groups zahlreiche Wettbewerbs- und Investitionsdynamiken auf Produktmärkten.<sup>21</sup>

### 3.3 ML-Ansätze in der Abschlussprüfung

Die US-amerikanische Fachzeitschrift „Accounting Horizons“ widmete der digitalen Transformation der Abschlussprüfung in den Jahren 2015 und 2019 jeweils ein Sonderheft.<sup>22</sup> Neben dem bereits erwähnten Beitrag von Ding et al. (2020) zur ML-gestützten Peer-Group-Bestimmung auf der Basis von Finanzkennzahlen (vgl. Abschn. 3.2) finden sich in dem 2019er „Data Analytics Forum“ weitere Beiträge, die z. T. ML-gestützte Anwendungen für die Abschlussprüfung vorschlagen. Sun (2019) bietet einen Überblick über Funktionsweise und mögliche Einsatzgebiete von künstlichen neuronalen Netzen (*Deep Neural Networks*) im Hinblick auf die Bereiche der Informationsgewinnung aus halb- (Textdaten) und unstrukturierten

<sup>21</sup> Einen weiteren Beitrag zur digitalisierten Peer-Group-Bestimmung leisten Dierkes und Sumpelmann (2019), die ausgewählte webbasierte Unternehmensbewertungstools hinsichtlich der Brauchbarkeit der in ihnen enthaltenen Verfahren zur Bestimmung von Referenzunternehmen testen.

<sup>22</sup> Vgl. *Accounting Horizons*, 29. Jahrgang, Heft 3 (Forum: Big Data) sowie 33. Jahrgang, Heft 3 (Data Analytics Forum).

(Bild-, Audio- und Video-) Daten für Zwecke der Geschäftsmodell-, Markt- und Risikoanalyse sowie der Unterstützung von Prüfungsurteilen mit halb- und unstrukturierten Daten als Prüfungsnachweisen. Der Beitrag schließt mit einer Diskussion von derzeit noch bestehenden Umsetzungsproblemen, insbesondere mangelnder Datenverfügbarkeit (bisher nicht hinreichend bestehende Audit Data Warehouses, die die Daten aller Mandanten enthalten), der Notwendigkeit, für jede Aufgabenstellung ein separates neuronales Netz zu trainieren und eine hinreichend große Menge an (oft manuell) klassifizierten Trainingsdaten zu generieren, sowie verbesserungsbedürftiger IT-Kenntnisse der Abschlussprüfer.

Eine konkrete Anwendung von ML-basierten Verfahren der automatisierten Textanalyse (vgl. dazu im Überblick Sellhorn et al. 2019) schlagen Yan und Moffitt (2019) für den Bereich der Analyse von Vertragsdaten vor. Kunden-, Leasing-, Kredit- und sonstige Verträge stellen eine wesentliche Informationsquelle im Rahmen der Abschlussprüfung vor, werden aber auf Grund der zu ihrer detaillierten Analyse notwendigen Zeit und Expertise bisher i. d. R. lediglich stichprobenartig ausgewertet (vgl. für die US-Prüfungsstandards AICPA 2008). Das von den Verfassern vorgeschlagene *Contract Analytics Framework* (CAF) umfasst die folgenden sechs Bereiche (vgl. Yan und Moffitt 2019, S. 112): (1) Dokumentenmanagement (Import und Verwaltung von Verträgen aus verschiedenen analogen und digitalen Quellen); (2) Inhaltsidentifikation (Identifizierung und Extraktion von prüfungsrelevanten Informationen, einschließlich Text, numerischen Daten und Tabellen); (3) *Cutoff Testing* (zeitliche Informationen zu Vertragsdauern und -fristen); (4) Datenabgleich (Validierung automatisiert extrahierter Daten mit bestehenden Vertragsdatenquellen wie ERP-Systemen); (5) Versionsvergleich (Erkennung von Abweichungen zwischen Vertragsversionen sowie von einer ursprünglichen Vorlage); und (6) Prüfungshandlungen (Auflistung von Prüfungshandlungen, die auf Daten des *Contract Analytics Framework* zugreifen). Probleme in der Umsetzung des CAF bestehen u. a. dort, wo Verträge handschriftliche Zusätze und Änderungen enthalten, deren fehlerfreie maschinelle Erfassung derzeit noch Probleme bereitet.

No et al. (2019) schlagen ML-Ansätze im Rahmen von Einzelfallprüfungen vor. Vor allem in Prüffeldern mit Massentransaktionen kommt hier regelmäßig die klassische Stichproben- bzw. Auswahlprüfung zur Anwendung (vgl. etwa Marten et al. 2015, S. 351). Wird von dieser Vorgehensweise mithilfe datenanalytischer Ansätze zu Gunsten einer Vollerfassung des gesamten (transaktionalen) Datenbestandes eines bestimmten Prüffeldes abgewichen, kommt es häufig zur Entdeckung einer hohen Anzahl von Auffälligkeiten (*Exceptions, Outliers*), die auf Falschdarstellungen hindeuten und daher vom Prüfer näher darauf untersucht werden müssen, ob ihnen tatsächlich Fehler zu Grunde liegen (vgl. No et al. 2019). Das manuelle Abarbeiten all dieser Auffälligkeiten, deren Anzahl diejenige, die im Rahmen von Stichprobenprüfungen auftreten, um ein Vielfaches übersteigt, ist regelmäßig nicht wirtschaftlich durchführbar. Daher steht der Prüfer vor dem Problem, aus den zahlreichen Auffälligkeiten diejenigen auswählen zu müssen, denen mit hoher Wahrscheinlichkeit ein Fehler zu Grunde liegt. Das von No et al. (2019) vorgeschlagene Multidimensional Audit Data Selection (MADS) Konzept soll Prüfer in die Lage versetzen, die erkannten Auffälligkeiten automatisiert anhand des ihnen inhärenten Risikos zu priorisieren. Anwendungsbereiche erkennen No et al. (2019) dort, wo schon heute

datenanalytische Verfahren in der Abschlussprüfung zur Anwendung kommen, nämlich vor allem bei Buchungs- und Hauptbuchanalysen sowie in geringerem Umfang bei der Umsatzanalyse, der Prozessanalyse und der Derivatebewertung.

Die hier dargestellten Beiträge beruhen sämtlich auf der Idee der ML-gestützten Analyse des gesamten Datenbestandes eines Prüffeldes durch den Abschlussprüfer. Die entwickelten Anwendungsvorschläge sind teilweise noch nicht flächendeckend umsetzbar, da beispielsweise die benötigten Daten nicht (oder nicht in der benötigten digitalen Form) vorliegen und der derzeit bestehende regulatorische Rahmen der Prüfungsstandards nach wie vor von einem risikoorientierten Prüfungsansatz mit Stichproben ausgeht, innerhalb dessen die Verwertung von datenanalytisch gewonnenen Prüfungsnachweisen unklar ist. Das International Auditing and Assurance Standards Board (IAASB) steht dem Einsatz von Datenanalysetools in der Abschlussprüfung vorsichtig positiv gegenüber (IAASB 2017). Allerdings sei die Qualität der Datengrundlage entscheidend (Gefahr des „Garbage in, garbage out“).

Dementsprechend steht die Forschung zur tatsächlichen Verbreitung dieser Ansätze sowie ihrer Einflussfaktoren und Auswirkungen, etwa auf Prüfungshonorare und -qualität, bisher noch am Anfang. Erste Interviewstudien gehen der Frage nach, wie Datenanalysen und Prozessautomatisierung Rechnungslegung und Abschlussprüfung realiter bereits verändert haben. Vielfach werden hier aber Meinungen, Einschätzungen und Erwartungen mindestens ebenso abgefragt wie konkrete Aktivitäten und Fakten. Einzelne Studien gewähren hier erste Einblicke in die Anwendung von Robotics Process Automation, (Big) Data Analytics und ML. So befragen Cooper et al. (2019) Wirtschaftsprüfer zur Nutzung von Robotic Process Automation (RPA) bzw. Software-Bots. Die Befragten sehen eine recht hohe Durchdringung in den Steuer- und Advisory-Abteilungen, gefolgt von der Abschlussprüfung selbst. Während RPA zu wahrgenommenen Effizienzgewinnen und Qualitätssteigerungen führe, gehe sie bisher nicht mit einem Verlust an Arbeitsplätzen einher. Auch sinkende Prüfungshonorare seien bisher nicht zu beobachten – wohl aber schwierigere Verhandlungen mit Mandanten über das zu vergütende Stundenvolumen. Cao et al. (2015) ziehen für Big Data Analytics ähnliche Schlüsse, während Earley (2015) eine verschärfte Erwartungslückenproblematik heraufziehen sehen. Gepp et al. (2018) beurteilen in ihrem Überblicksbeitrag die Abschlussprüfung in puncto Technologieinsatz nicht als Pionier, sondern erkennen die größten Potenziale für ML-gestützte Big-Data-Analysen in den Bereichen Insolvenz-, Misstatement- und Aktienkursprognose.

## 4 Einsatzgebiete von Machine Learning-Ansätzen in der RLF

### 4.1 Datengenerierung

In der Finanz- und RLF finden ML-Ansätze primär bei der Verdichtung und Datenextraktion aus großen Mengen unstrukturierter Texte Anwendung.<sup>23</sup> Sellhorn et al. (2019) unterscheiden dabei erstens Ansätze für die automatisierte Beschaffung und

---

<sup>23</sup> Vgl. statt Vieler Loughran und McDonald (2016).

Aufbereitung großzahliger Daten sowie zweitens die automatisierte Extraktion konkreter Informationen und latenter Signale aus großen Textkorpora. Im ersten Fall geht es zunächst um das automatisierte Herunterladen und Speichern großer Textmengen aus dem Internet. So erlaubt die Website der US-Börsenaufsicht SEC (EDGAR; Electronic Data Gathering and Retrieval) den automatisierten Zugriff auf die dort eingereichten Publizitätsinstrumente.<sup>24</sup> Zweitens können sowohl überwachte als auch unüberwachte ML-Algorithmen beispielsweise für die Klassifikation von Texten und deren Zuordnung zu thematischen Bereichen genutzt werden.<sup>25</sup> So nutzen etwa Hoberg und Philips (2016) die Geschäftsmodellbeschreibungen in US-amerikanischen Geschäftsberichten (*Form 10-K*), um für US-Unternehmen zeitlich variable Branchenzuordnungen zu ermitteln und relevante Wettbewerber zu identifizieren.

## 4.2 Unterstützung kausaler Inferenz

Wie in Abschn. 2.1 diskutiert, sind zahlreiche Fragen in der RLF kausaler Natur. Dies gilt insb. für die Auswirkungen von Änderungen in Rechnungslegungs- und anderen relevanten Regeln. Dazu zählen beispielsweise der Effekt der verpflichtenden IFRS-Einführung in der Europäischen Union (vgl. etwa Brüggemann et al. 2013) auf die Kapitalkosten oder die Auswirkungen der verpflichtenden CSR-Berichterstattung auf den CO<sub>2</sub>-Ausstoß der jeweils betroffenen Unternehmen. Auch im Kontext von Ereignisstudien interessieren kausale Fragen, insbesondere: Wie wäre der Aktienkurs ohne das fragliche Informationsereignis verlaufen? Oft wird unterstellt, dieser hätte sich in dessen Abwesenheit nach dem Marktmodell oder dem CAPM entwickelt.

Die Herausforderungen kausaler Inferenz in nicht-experimentellen Beobachtungsstudien sind bekannt und wurden auch für die RLF bereits intensiv diskutiert (vgl. etwa Larcker und Rusticus 2007; Gassen 2014; Gow et al. 2016). Das konzeptionelle Problem besteht darin, dass die kontrafaktische Entwicklung der interessierenden abhängigen Variablen nicht beobachtbar ist. Wie hätten sich die Kapitalkosten in Abwesenheit der verpflichtenden IFRS-Einführung bzw. der CO<sub>2</sub>-Ausstoß ohne verpflichtende CSR-Berichterstattung entwickelt? Die *X*-Variable kann i. d. R. nicht ohne Weiteres wie in einem kontrollierten Zufallsexperiment randomisiert variiert und ihr Effekt beobachtet werden. Über die kontrafaktische Entwicklung der *Y*-Variable sind daher Annahmen nötig. Dieses Problem der nicht-beobachtbaren kontrafaktischen Entwicklung wird auch als „fundamental problem of causal inference“ bezeichnet.<sup>26</sup>

<sup>24</sup> Vgl. für weitere Informationen [www.sec.gov/edgar/searchedgar/accessing-edgar-data.htm](http://www.sec.gov/edgar/searchedgar/accessing-edgar-data.htm). Bei einfachem „web scraping“ kommen nicht zwingend ML-Verfahren zum Einsatz, wohl jedoch im Rahmen anderer Extraktionsroutinen. Bei der *Named Entity Recognition* z. B. „prognostiziert“ ein ML-Algorithmus, ob es sich bei einer Zeichenfolge um einen Eigennamen (z. B. Unternehmen, Person, Land) handelt. Der Verfasser dankt *Harm H. Schütt* für diesen Hinweis.

<sup>25</sup> Vgl. mit weiteren Nachweisen Sellhorn et al. (2019).

<sup>26</sup> Athey (2015, S. 5). Es ist jedem geläufig, der – gefragt, ob die eingenommenen homöopathischen Globuli denn gegen die Halsschmerzen geholfen hätten – den Impuls verspürt zu antworten: „Das weiß ich nicht. Ich weiß ja nicht, wie es mir jetzt ginge, wenn ich sie nicht eingenommen hätte.“

ML-Algorithmen können – obwohl für Prognosen optimiert – auch Erklärungsmodelle verbessern, die für kausale Fragestellungen geschätzt werden.<sup>27</sup> Denn kausale Erklärungsmodelle lassen sich differenzieren in den Einfluss der kausalen Variable auf die abhängige Variable (z. B. den Effekt einer Intervention wie der verpflichtenden IFRS-Einführung auf die Kapitalkosten der betroffenen Unternehmen) sowie weitere auf die abhängige Variable wirkende Größen, die für die Untersuchung des interessierenden Kausaleffekts idealerweise konstant gehalten werden sollen. ML-Algorithmen eignen sich besonders dazu, die statistische Assoziation dieser anderen Faktoren mit der abhängigen Variablen – und damit die nicht-beobachtbare kontrafaktische Entwicklung der abhängigen Variablen in Abwesenheit des interessierenden Kausaleffekts – zu modellieren. Denn hierin liegt im Grunde ein Prognoseproblem, für das ML-Algorithmen optimiert sind. Solche, oft auf Basis großer Datenmengen ML-gestützt prognostizierten, „Was wäre wenn“-Szenarien werden dann mit der tatsächlichen Entwicklung verglichen, um den kausalen Effekt der Intervention zu schätzen.

Eine weitere empirische Herausforderung (nicht nur) bei kausalen Fragestellungen betrifft die Robustheit der Ergebnisse in Bezug auf den Einfluss verschiedener Modellspezifikationen (vgl. schon Leamer 1983; Leamer und Leonard 1983). Athey und Imbens (2015) schlagen einen ML-Ansatz zur systematischen Robustheitsanalyse vor, der Schätzungen des interessierenden Kausaleffekts automatisiert für eine Vielzahl von Modellspezifikationen erstellt. Die Standardabweichung der Punktschätzungen über die betrachteten Modellspezifikationen hinweg dient dann als Robustheitsmaß für den interessierenden Kausaleffekt.

Mullainathan und Spiess (2017) diskutieren einen weiteren Anwendungsfall, nämlich wie ML-Algorithmen z. B. im Kontext von Instrumentvariablen-schätzungen bei *Correlated Omitted Variable Bias* zu valideren kausalen Schlüssen beitragen können. In einer typischen Anwendung wählt der Forscher zunächst einen Satz von Variablen aus, die als exogen angenommen werden, und verwendet dann zweistufige Least-Squares (2SLS) oder ähnliche Schätzverfahren zum Schätzen der Koeffizienten im Regressionsmodell.<sup>28</sup> Interessiert beispielsweise der kausale Einfluss von Offenlegungsqualität ( $X$ ) auf die Kapitalkosten ( $Y$ ), wird auf der ersten Stufe der Instrumentvariablen-schätzung die interessierende unabhängige  $X$ -Variable, Offenlegungsqualität, als Funktion einer Instrumentvariablen und anderer Variablen modelliert (vgl. Larcker und Rusticus 2010). Aus den Koeffizienten werden angepasste Werte für Offenlegungsqualität,  $\hat{x}$ , abgeleitet, die auf der zweiten Stufe als erklärende Variable für die Kapitalkosten verwendet werden. Besonders in kleinen Stichproben neigt die erste Stufe zu Überanpassung (*Overfitting*). Dies führt dazu, dass die  $\hat{x}$  stark mit  $x$  korreliert sind, der *Correlated Omitted Variable Bias* also nicht hinreichend korrigiert wird. Versteht man die erste Stufe jedoch als Prognoseproblem, können hier ML-Algorithmen angewendet werden, die robuster gegen Überanpassung sind, indem sie etwa für die Modellierung der  $X$ -Variablen und die Ableitung der angepassten Werte  $\hat{x}$  separate Datensätze verwenden (vgl. Mullainathan und Spiess 2017, S. 100).

<sup>27</sup> Vgl. zum Folgenden Athey (2015).

<sup>28</sup> Einen Überblick für die RLF bieten Larcker und Rusticus (2010).

Eine weitere Anwendungsmöglichkeit für ML-Algorithmen zur Unterstützung kausaler Schlüsse liegt u. a. in der Schätzung von *Propensity Scores*, um die Vergleichbarkeit von Behandlungs- und Kontrollgruppe zu erhöhen. Allgemeiner: Wenn ein ML-gestütztes Prognosemodell die Zuordnung zum *Treatment* auf der Basis von *Pre-Treatment* Faktoren vorhersagen kann, ist die wichtige Voraussetzung der *Covariate Balance* nicht gegeben. Oder: Wenn sich die Zuordnung zum *Treatment* auf der Basis von *Post-Treatment Outcomes* „vorhersagen“ lässt, muss das *Treatment* einen Effekt gehabt haben (Mullainathan und Spiess 2017, S. 101).

### 4.3 Prognoseprobleme

Überwachte (*Supervised*) ML-Ansätze haben ihre wesentlichen Stärken bei der Entwicklung und Schätzung von Prognosemodellen (vgl. etwa Athey 2015); sie werden daher auch als „prediction machines“ bezeichnet.<sup>29</sup> Diese Verfahren prognostizieren eine abhängige Variable  $Y$  („Dieses Foto enthält ein/kein menschliches Gesicht“) auf der Basis einer Vielzahl von unabhängigen Variablen  $X$  (Pixel, aus denen das Foto besteht). Trainiert auf der Basis einer Anzahl von Menschenhand als  $X=0$  bzw.  $Y=1$  kategorisierter Fotos, minimiert der Algorithmus den Prognosefehler des Prognosemodells. ML-Ansätze eignen sich daher für Anwendungsbereiche in der RLF, bei denen es um Prognose- oder Klassifikationsprobleme geht (vgl. Mullainathan und Spiess 2017). Dies betrifft z. B. Tests der Effizienzmarkthypothese. Hier sollte es kein Prognosemodell geben, das verlässlich und systematisch künftige Aktienrenditen auf der Basis vergangener Kursbewegungen (schwache Informationseffizienz) oder öffentlich verfügbarer Daten (halbstarke Informationseffizienz) vorhersagen kann. Gelingt dies doch, liegt eine neue Anomalie vor. Arnott et al. (2019) diskutieren den Einsatz ML-gestützter Prognosemodelle zur datengetriebenen Entwicklung renditeträchtiger Tradingstrategien (vgl. dazu Sellhorn et al. 2019). Die in Ding et al. (2020) diskutierten ML-gestützten Schätzalgorithmen (vgl. oben Abschn. 3.1) lassen sich prinzipiell zur Aufdeckung bilanzpolitischer Verzerrungen einsetzen, wenn nämlich die tatsächlichen Schätzungen systematisch und mit bilanzpolitischen Anreizen konsistent von den algorithmisch prognostizierten Werten abweichen.

Effektstudien zu Regulierungsmaßnahmen sind primär an den kausalen Auswirkungen der betreffenden Interventionen interessiert. Diese Problematik wurde in Abschn. 4.2 diskutiert. Staatliches Handeln kann aber auch mit Prognoseproblemen konfrontiert sein. Mullainathan und Spiess (2017) nennen das Beispiel eines Richters, der das Risiko abschätzen muss, ob ein in Untersuchungshaft einsitzender Verdächtiger nach einer möglichen Entlassung untertaucht oder gar weitere Straftaten begeht. Hierin liegt prinzipiell ein Prognoseproblem, das anhand von Daten über Untersuchungshäftlinge mithilfe von ML-Algorithmen adressiert werden kann. In der Rechnungslegungspraxis stellen sich ähnliche Fragen, z. B. im Kontext der Auswahl des Abschlussprüfers, der die höchste Prüfungsqualität bei gegebenem Honorar verspricht.

---

<sup>29</sup> So der Titel eines aktuellen Buches (Agrawal et al. 2018).

Auch Regulierungsfragen haben gelegentlich einen Prognoseaspekt.<sup>30</sup> Angenommen, verpflichtende CSR-Berichterstattung soll den CO<sub>2</sub>-Ausstoß senken helfen, indem Unternehmen Anreize zu nachhaltigerem Handeln erhalten. Die Frage, wie CSR-Berichterstattung kausal auf den CO<sub>2</sub>-Ausstoß wirkt, erfordert ein kausales Erklärungsmodell. Relevant ist aber auch, wie hoch der CO<sub>2</sub>-Ausstoß in der Zukunft insgesamt sein wird. Denn CSR-Berichterstattung und andere Maßnahmen der Nachhaltigkeitsregulierung verursachen Kosten. Ob diese sich lohnen, hängt auch vom erwarteten CO<sub>2</sub>-Ausstoß in der Zukunft und dessen ökonomischen Folgen ab. Dieser wird aber – neben der CSR-Berichterstattung – von vielen weiteren Faktoren beeinflusst. ML-Ansätze können helfen, diese anderen Faktoren zu berücksichtigen, um letztlich den interessierenden Kausaleffekt der CSR-Berichterstattung isolieren zu können. Dabei steht jedoch stets die Frage im Raum, wie stabil die ML-gestützten Prognosemodelle im Zeitablauf sind.

## 5 (Mangelnde) Transparenz von Machine Learning-Ansätzen

Als ein möglicher limitierender Faktor für das Vordringen von ML-Algorithmen wird die vielfach noch mangelnde Nachvollziehbarkeit und Transparenz algorithmischer Prognosen, Klassifikationen und Entscheidungen diskutiert. Denn die entsprechenden Modelle werden durch einen Algorithmus direkt aus Daten erstellt, so dass Menschen (und sogar die Entwickler der Algorithmen) nicht verstehen können, wie Variablen zu Prognosen und Entscheidungen kombiniert werden. Selbst bei Kenntnis aller Inputgrößen können ML-Prognosemodelle so komplizierte Funktionen dieser Variablen sein, dass Menschen nicht verstehen können, über welche funktionalen Zusammenhänge diese Variablen miteinander verbunden sind, um eine endgültige Prognose zu erstellen (vgl. Rudin und Radin 2019).

Diese mangelnde Transparenz kostet Vertrauen. So fragen Mullainathan und Spiess (2017, S. 103): „What factors determine faith in the algorithm?“ Im Kontext von Unternehmensberichterstattung und Abschlussprüfung stellen Kokina und Davenport (2017) fest: „However, machine learning and deep learning neural networks, for example, are often ‚black boxes‘ that are difficult or impossible to understand and interpret, even for technical experts. Until such technologies are made more transparent, it may be difficult for regulatory bodies, accounting firms, and audited organizations to turn over decisions and judgments to them.“ Auch weisen Dierkes und Sümpelmann (2019, S. 190) darauf hin, dass die webbasierte Unternehmensbewertung „durch die Digitalisierung nicht zu einer Black Box werden“ darf, der es „an einer ausreichenden theoretischen Fundierung mangelt“, sondern „bei gleichem Ressourceneinsatz zu einer Verbesserung der Qualität“ führen sollte.

Die Black-Box-Problematik wird allerdings als unterschiedlich schwerwiegend eingestuft. So wurden moderne Techniken des maschinellen Lernens ursprünglich primär für Entscheidungen mit niedrigem Risiko entwickelt, bei denen individuelle Entscheidungen, etwa im Kontext von Online-Werbung und Websuche, das menschliche Leben nicht zu stark beeinflussen (vgl. Rudin und Radin 2019). So weist

<sup>30</sup> Kleinberg et al. (2015) sprechen hier von *Prediction Policy Problems*.

Sun (2019, S. 90) darauf hin, dass viele Werbetreibende der „Blackbox“ hinter Google AdWords vertrauen, ohne den Algorithmus im Detail verstehen zu können. Allerdings verlassen sich Menschen zunehmend nicht nur in ihren Konsum- und Informationsnutzungsentscheidungen zunehmend auf ML-basierte Such- und Empfehlungsalgorithmen, sondern vertrauen sich und ihr Leben ML-basierten Autopilot- und Fahrassistenzsystemen teilweise sogar vollständig an. Entscheidend ist offenbar in vielen Fällen die Funktionalität i.S. einer hohen und stabilen (wahrgenommenen) Klassifikations- bzw. Prognosegüte – weniger die Nachvollziehbarkeit der Klassifikationen und Prognosen.<sup>31</sup> Zudem ist festzuhalten, dass auch die Äußerungen menschlicher Experten einen Black-Box-Charakter haben können, da auf tiefem Erfahrungswissen gründende, oft intuitiv getroffene Prognosen, Klassifikationen und Entscheidungen für den menschlichen Akteur ebenfalls kaum verbal beschreib- und begründbar (und damit auch nicht in Programmcode ausdrückbar) sind.

Dennoch hat der Bedarf nach Nachvollziehbarkeit und Transparenz dazu geführt, dass zunehmend ML-Algorithmen entwickelt werden, deren Prognosen, Klassifikationen und Entscheidungen zumindest teilweise erklärbar sind (sog. *Explainable AI*; XAI). Hier kommt häufig der konstruktionsorientierte Forschungsansatz (*Design Science*; vgl. Abschn. 2.1) zum Einsatz. Anwendungsbeispiele finden sich in den Bereichen Prozessmanagement bzw. *Process Mining* (vgl. etwa Evermann et al. 2017, S. 136–138; Rehse et al. 2019; Mehdiyev und Fettke 2020) sowie im an die RLF angrenzenden Bereich der finanzwirtschaftlichen Forschung. So entwickeln Li et al. (2019) ein Rahmenkonzept für die intuitiv nachvollziehbare Interpretation von ML-gestützten Wechselkursprognosen.

Fraglich ist in diesem Kontext, ob zwischen der Nachvollziehbarkeit und Transparenz von ML-Algorithmen einerseits und ihrer Klassifikations- bzw. Prognosegüte andererseits ein Gegensatz besteht, eine höhere Leistungsfähigkeit also mit der Notwendigkeit erkaufte wird, einer Blackbox zu vertrauen. So haben Fortschritte bei Deep-Learning-Algorithmen für die Bildverarbeitung (*Computer Vision*) in den letzten Jahren zu der weit verbreiteten Annahme geführt, dass die genauesten Modelle für ein bestimmtes datenwissenschaftliches Problem inhärent intransparent und kompliziert sein müssen (vgl. Rudin und Radin 2019). Diese weit verbreitete Ansicht speist sich aus dem Umstand, dass die Komplexität eines Algorithmus häufig mit seiner Leistungsfähigkeit positiv, mit seiner Nachvollziehbarkeit und Transparenz aber negativ korreliert ist.

Dass dieser Tradeoff zwischen Leistungsfähigkeit und Interpretierbarkeit nicht notwendigerweise besteht, argumentieren Rudin und Radin (2019) in einem Bericht über die erste *Explainable Machine Learning Challenge*, einen 2018 erstmals durchgeführten Wettbewerb mit dem Ziel, komplexe und hochleistungsfähige Blackbox-ML-Algorithmen für ein gegebenes Problem und einen gegebenen Datensatz zu entwickeln und diese anschließend erklärbar zu machen. So sei beispielsweise für kriminologische Anwendungen mehrfach nachgewiesen worden, dass komplizierte

---

<sup>31</sup> So berichten Rudin und Radin (2019) von einer Befragung, in der Probanden zu wählen hatten zwischen einem menschlichen Herzchirurgen, der eine hochkomplexe Operation zwar detailliert erklären konnte, seine Patienten aber einem 15 %-igen Sterblichkeitsrisiko aussetzte, und einem intransparenten Roboter mit einem nur 2 %-igen Risiko. Die ganz überwiegende Mehrheit entschied sich für den Roboter.



Blackbox-Modelle auf der Basis moderner ML-Methoden zur Vorhersage zukünftiger Straffälligkeit nicht genauer sind als sehr einfache Vorhersagemodelle auf der Grundlage von Alter und Vorstrafen. Ähnliches gelte in mehreren Bereichen des Gesundheitswesens sowie anderen Bereichen, in denen es um lebensverändernde Entscheidungen geht. Auch hier gingen transparentere Berechnungen nicht auf Kosten der Genauigkeit (vgl. Rudin und Radin 2019). Die Idee, dass simple Modelle häufig verblüffend hohe Erklärungs- und Prognosekraft und zudem große Vorteile hinsichtlich Betriebskosten und Nachvollziehbarkeit aufweisen, liegt auch dem u. a. von Gigerenzer vertretenen Konzept der sog. *Fast-and-Frugal Trees* zugrunde (vgl. etwa Artinger et al. 2015). Hierbei handelt es sich um heuristische Entscheidungsbäume auf der Basis einiger weniger Variablen, die insb. bei Entscheidungen unter Unsicherheit Vorteile gegenüber komplexen statistischen Verfahren aufweisen.

## 6 Fazit und Ausblick

Dieser Beitrag verfolgte das Ziel, einige mögliche Implikationen einer bestimmten Facette der digitalen Transformation – nämlich des Einsatzes von ML-Ansätzen für die Lösung von Klassifikations- und Prognoseproblemen – für die Praxis und Forschung in den Bereichen Unternehmensberichterstattung, Abschlussprüfung sowie Unternehmensanalyse und -bewertung aufzuzeigen und zur Diskussion zu stellen. Dabei wurde argumentiert, dass ML nicht nur die in Rechnungslegung und Abschlussprüfung den tatsächlichen Verhältnissen entsprechend abzubildenden und in der Unternehmensanalyse und -bewertung zu bewertenden Sachverhalte verändert, sondern auch zunehmend tiefer die in Rechnungslegung und Abschlussprüfung selbst zum Einsatz kommenden Systeme und Prozesse prägt. Zudem wurde diskutiert, welche Potenziale ML-Ansätze für die RLF bieten – primär im Rahmen der kostengünstigen und replizierbaren Generierung neuer Forschungsdaten, aber auch im Kontext ökonomischer Erklärungs- und Prognosemodelle.

Was bleibt nun festzuhalten? In der empirischen RLF spielen ML-Ansätze als Untersuchungsgegenstand erst seit relativ kurzer Zeit eine nennenswerte Rolle. Mit im Wesentlichen beschreibender Methodik werden Einsatz und Verbreitung von ML-Ansätzen in Rechnungswesen und Abschlussprüfung untersucht. Da die öffentliche Datenverfügbarkeit begrenzt ist, herrschen derzeit kleinzahlige Befragungsstudien vor. Entsprechend sind bisher nur wenige Erkenntnisse zu den Einflussfaktoren und Auswirkungen des Einsatzes von ML-Ansätzen gesichert.

Deutlich weiter in die empirische RLF vorgedrungen sind ML-Ansätze als Methode zur Datengewinnung. Hier steht seit über zehn Jahren insbesondere die Extraktion von Signalen aus großen, unstrukturierten Textkorpora durch *Natural Language Processing* im Fokus. Auch zur Prognose bzw. Erkennung von Ereignissen wie Bilanzbetrug oder Insolvenz sind ML-Ansätze bereits seit Längerem fester Bestandteil des empirischen Instrumentariums. Erst seit Neuestem werden ML-Algorithmen zur Schätzung zukünftiger Bilanzpositionen verwendet und hinsichtlich ihrer Prognosepräzision mit menschlichen Vorhersagen verglichen.

Der zukünftigen Forschung auf diesem Gebiet stehen vielfältige Wege offen, die hier nur subjektiv und selektiv angerissen werden können. Zum einen wäre weiter

zu ergründen, von wem und warum sowie mit welchen Zielen und Folgen ML-Algorithmen in Rechnungswesen und Abschlussprüfung eingesetzt werden. Wirken zum Beispiel neue Transparenzvorschriften wie IFRS 16 *Leasingverhältnisse* als Katalysatoren für die Einführung ML-gestützter IT-Lösungen in den Finanzbereichen der betroffenen Unternehmen, etwa zur automatisierten Erfassung von Vertragsdaten? Werden damit primär Effizienz- oder (auch) Qualitätssteigerungsziele verfolgt (und erreicht)? Welche Rolle spielen Performance und Transparenz für die Einsatzmöglichkeiten, Verbreitung und Akzeptanz von ML-Ansätzen? Welchen Einfluss hat der Einsatz von ML-Ansätzen in der Abschlussprüfung auf Marktstruktur, Prüfungsqualität, Prüferhonorare sowie die Anforderungen an und Arbeitsbedingungen für zukünftige Berufsangehörige?<sup>32</sup>

Aber auch aus forschungsmethodischer Sicht ist das Potenzial von ML-Ansätzen noch längst nicht ausgeschöpft. Stetig steigende Rechnerleistung, Speicherkapazität und Datenverfügbarkeit erlauben die Entwicklung immer leistungsfähigerer Algorithmen. Vielfältige Datenquellen neben unstrukturierten Texten – etwa Audio-, Video- oder Sensordaten – sind in der RLF noch weitgehend ungenutzt. Schließlich sind die empirischen Fragestellungen der RLF heute zu einem überwiegenden Teil kausaler Natur. Das erhebliche Potenzial von ML-Ansätzen, Prognoseprobleme im Bereich der RLF zu adressieren, liegt damit zurzeit noch weitgehend brach. Wie stehen zum Beispiel diverse Aspekte ökologischer und sozialer Nachhaltigkeit von Unternehmen im Zusammenhang mit zukünftigen finanziellen Zielgrößen? Welche Prognosen über die zukünftige Unternehmensentwicklung erlauben die zahlreichen Facetten unternehmerischer Offenlegungsentscheidungen? Das Aufspüren auffälliger Datenmuster und Zusammenhänge mithilfe von ML-Algorithmen dürfte zukünftig zunehmend nicht nur praxisrelevante Erkenntnisse liefern, sondern auch vielfältige Anstöße für die Theoriebildung in der RLF bieten.

**Danksagung** Ich danke Harm H. Schütt, einem anonymen Gutachter, Alfred Wagenhofer (Herausgeber) sowie den Organisatoren und Teilnehmern der Jahrestagung 2019 der Wissenschaftlichen Kommission Rechnungswesen im Verband der Hochschullehrer für Betriebswirtschaft für wertvolle Anregungen. Gefördert durch die Deutsche Forschungsgemeinschaft (DFG): Projekt-ID 403041268 – TRR 266.

**Funding** Open Access funding provided by Projekt DEAL.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

<sup>32</sup> Vgl. zu ersten empirischen Erkenntnissen für die Rechts-, Steuer- und Wirtschaftsberatungsbranche etwa Warning et al. (2020).

## Literatur

- Agrawal, Ajay, Joshua Gans, und Avi Goldfarb. 2018. *Prediction machines: the simple economics of artificial intelligence*. Boston: Harvard Business Press.
- Altman, Edward I. 1968. Financial ratios, Discriminant analysis and the prediction of corporate bankruptcy. *The Journal Of Finance* 23(4):589–609.
- American Institute of Certified Public Accountants (AICPA). 2008. Audit sampling. AU section 350. <https://pcaobus.org/Standards/Auditing/pages/au350.aspx>. Zugegriffen: 5. März 2020.
- Arnott, Rob, R. Harvey Campbell, und Harry Markowitz. 2019. A backtesting protocol in the era of machine learning. *The Journal of Financial Data Science* 1(1):64–74.
- Artinger, Florian, Malte Petersen, Gerd Gigerenzer, und Jürgen Weibler. 2015. Heuristics as adaptive decision strategies in management. *Journal of Organizational Behavior* 36(1):33–52.
- Athey, Susan. 2015. Machine learning and causal inference for policy evaluation. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 5–6. Sydney: ACM.
- Athey, Susan, und Guido Imbens. 2015. A measure of robustness to misspecification. *American Economic Review* 105(5):476–480.
- Baetge, Joerg. 1998. Das objektive, Bilanzpolitik konterkarierende, ganzheitliche System zur Bilanzbonitätsbeurteilung BP-14. In *Empirische Methoden zur Früherkennung von Unternehmenskrisen*, 17–28. Wiesbaden: VS.
- Ballwieser, Wolfgang, und Dirk Hachmeister. 2019. Der Zusammenhang von Digitalisierung und Unternehmensbewertung. In *Digitalisierung und Unternehmensbewertung*, Hrsg. W. Ballwieser, D. Hachmeister, 11–33. Stuttgart: Schäffer-Poeschel.
- Beaver, William H. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research* 4:71–111.
- Beneish, Messod D. 1997. Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance. *Journal of Accounting and Public Policy* 16(3):271–309.
- Bloomfield, Robert, Mark W. Nelson und Eugene Soltes. 2016. Gathering Data for Archival, Field, Survey, and Experimental Accounting Research. *Journal of Accounting Research* 54(2):341–395.
- Brüggemann, Ulf, Jörg-Markus Hitz und Thorsten Sellhorn. 2013. Intended and Unintended Consequences of Mandatory IFRS Adoption: A Review of Extant Evidence and Suggestions for Future Research. *European Accounting Review* 22(1):1–37.
- Brynjolfsson, Erik, und Andrew McAfee. 2014. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. New York: WW Norton & Company.
- Cao, Min, Roman Chychyla, und Trevor Stewart. 2015. Big Data analytics in financial statement audits. *Accounting Horizons* 29(2):423–429.
- Cooper, Lauren A., Kip D. Holderness Jr., Trevor L. Sorensen, und David A. Wood. 2019. Robotic process automation in public accounting. *Accounting Horizons* 33(4):15–35.
- Dierkes, Stefan, und Johannes Sümpelmann. 2019. Digitalisierte Peer-Group-Bestimmung und Beta-Anpassung. In *Digitalisierung und Unternehmensbewertung*, Hrsg. W. Ballwieser, D. Hachmeister, 173–192. Stuttgart: Schäffer-Poeschel.
- Ding, Kexing, Baruch Lev, Xuan Peng, Sun Ting, und Miklos A. Vasarhelyi. 2020. *Machine learning improves accounting estimates*. available at SSRN 3253220.
- Ding, Kexing, Xuan Peng, und Yunsen Wang. 2019. A machine learning-based Peer selection method with financial ratios. *Accounting Horizons* 33(3):75–87.
- DPR. 2020. Prüfungsschwerpunkte 2020. [https://www.frep.info/docs/pressemitteilungen/2019/20191118\\_pm.pdf](https://www.frep.info/docs/pressemitteilungen/2019/20191118_pm.pdf). Zugegriffen: 5. März 2020.
- Earley, Christine E. 2015. Data analytics in auditing: opportunities and challenges. *Business Horizons* 58(5):493–500.
- Evermann, Joerg, Jana R. Rehse, und Peter Fettke. 2017. Predicting process behaviour using deep learning. *Decision Support Systems* 100:129–140.
- Ford, Martin. 2015. *Rise of the robots: technology and the threat of a jobless future*. New York: Basic Books.
- Gassen, Joachim. 2014. Causal inference in empirical archival financial accounting research. *Accounting, Organizations and Society* 39(7):535–544.

- Gepp, Adrian, Martina K. Linnenluecke, Terence J. O'Neill, und Tom Smith. 2018. Big data techniques in auditing research and practice: current trends and future opportunities. *Journal of Accounting Literature* 40:102–115.
- Gow, Ian D., David F. Larcker, und Peter C. Reiss. 2016. Causal inference in accounting research. *Journal of Accounting Research* 54(2):477–523.
- Hevner, Alan R., Salvatore T. March, Jinsoo Park, und Sudha Ram. 2004. Design science in information systems research. *MIS Quarterly* 28(1):75–105.
- Hoberg, Gerard, und Gordon Phillips. 2010. Product market synergies and competition in mergers and acquisitions: a text-based analysis. *Review of Financial Studies* 23(10):3773–3811.
- Hoberg, Gerard, und Gordon Phillips. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124(5):1423–1465.
- International Auditing and Assurance Standards Board. 2017. Exploring the growing use of technology in the audit, with a focus on data analytics. <https://www.iaasb.org/publications-resources/exploring-growing-use-technology-audit-focus-data-analytics>. Zugegriffen: 5. März 2020.
- Kelly, Kevin. 2017. *The inevitable: understanding the 12 technological forces that will shape our future*. New York: Penguin.
- King, Gary, Robert O. Keohane und Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, und Ziad Obermeyer. 2015. Prediction policy problems. *American Economic Review* 105(5):491–495.
- Kogan, Alexander, Brian W. Mayhew, und Miklos A. Vasarhelyi. 2019. Audit data analytics research—an application of design science methodology. *Accounting Horizons* 33(3):69–73.
- Kokina, Julia, und Thomas H. Davenport. 2017. The emergence of artificial intelligence: How automation is changing auditing. *Journal of Emerging Technologies in Accounting* 14(1):115–122.
- KPMG. 2017. Digitalisierung im Rechnungswesen. <https://hub.kpmg.de/digitalisierung-im-rechnungswesen-2017>. Zugegriffen: 5. März 2020.
- KPMG. 2018. Digitalisierung im Rechnungswesen 2018. <https://hub.kpmg.de/digitalisierung-im-rechnungswesen-2018>. Zugegriffen: 5. März 2020.
- KPMG. 2019. Digitalisierung im Rechnungswesen 2019. <https://hub.kpmg.de/digitalisierung-im-rechnungswesen-2019>. Zugegriffen: 5. März 2020.
- Krause, Stefan und Bernhard Pellens. 2017. Betriebswirtschaftliche Implikationen der digitalen Transformation. In *ZfbF-Sonderheft 72/12*, Hrsg. Stefan Krause, Bernhard Pellens. 1–373. Wiesbaden: Springer-Gabler.
- Larcker, David F. und Tjomme O. Rusticus. 2007. Endogeneity and Empirical Accounting Research. *European Accounting Review* 16(1):207–215.
- Larcker, David F., und Tjomme O. Rusticus. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* 49(3):186–205.
- Leamer, Edward E. 1983. Let's take the con out of econometrics. *Modelling Economic Series* 73:31–43.
- Leamer, Edward, und H. Leonard. 1983. Reporting the fragility of regression estimates. *The Review of Economics and Statistics* 65(2):306–317.
- Lev, Baruch. 2018. The deteriorating usefulness of financial report information and how to reverse it. *Accounting and Business Research* 48(5):465–493.
- Lev, Baruch. 2019. Ending the accounting-for-intangibles status quo. *European Accounting Review* 28(1):1–24.
- Lev, Baruch, und Feng Gu. 2016. *The end of accounting and the path forward for investors and managers*. Hoboken: John Wiley & Sons.
- Li, Yimou, David Turkington, und Alireza Yazdani. 2019. Beyond the black box: an intuitive approach to investment prediction with machine learning. *The Journal of Financial Data Science* 2(1):61–75.
- Libby, Robert. 1981. *Accounting and human information processing: theory and applications*. Englewood Cliffs: Prentice-Hall.
- Libby, Robert, Robert Bloomfield, und Mark W. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27(8):775–810.
- Loughran, Tim, und Bill McDonald. 2016. Textual analysis in accounting and finance: a survey. *Journal of Accounting Research* 54(4):1187–1230.
- Marten, Kai Uwe, Reiner Quick, und Klaus Ruhnke. 2015. *Wirtschaftsprüfung: Grundlagen des betriebswirtschaftlichen Prüfungswesens nach nationalen und internationalen Normen*, 5. Aufl., Stuttgart: Schäffer-Poeschel.
- McAfee, Andrew, und Erik Brynjolfsson. 2017. *Machine, platform, crowd: Harnessing our digital future*. New York: WW Norton & Company.

- Mehdiyev, Nijat, und Peter Fettke. 2020. *Prescriptive process analytics with deep learning and explainable artificial intelligence*. 28<sup>th</sup> European Conference on Information Systems (ECIS2020).
- Menacher, Julia, Harm H. Schütt, und Thorsten Sellhorn. 2017. Informationsverarbeitung an Finanzmärkten im Umbruch – zwischen Informationsflut und automatisierter Datenanalyse. *Kapitalmarktorientierte Rechnungslegung* 17:217–224.
- Mullainathan, Sendhil, und Jann Spiess. 2017. Machine learning: an applied econometric approach. *Journal of Economic Perspectives* 31(2):87–106.
- No, Won Gyun, Kyungha Kari Lee, Feiqi Huang, und Qiao Li. 2019. Multidimensional audit data selection (MADS): a framework for using data analytics in the audit data selection process. *Accounting Horizons* 33(3):127–140.
- Rehse, Jana R., Nijat Mehdiyev, und Peter Fettke. 2019. Towards explainable process predictions for industry 4.0 in the DFKI-smart-Lego-factory. *KI-Künstliche Intelligenz* 33(2):181–187.
- Rudin, Cynthia und Joanna Radin. 2019. Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.5a8a3a3d>.
- Schwab, Klaus. 2017. *The fourth industrial revolution*. München: Currency.
- Sellhorn, Thorsten, Gereon Hillert, und Julia Menacher. 2019. Automatisierte Textanalyse für Bewertungszwecke und Portfolioentscheidungen. In *Digitalisierung und Unternehmensbewertung*, Hrsg. W. Ballwieser, D. Hachmeister, 119–148. Stuttgart: Schäffer-Poeschel.
- Shmueli, Galit. 2010. To explain or to predict? *Statistical Science* 25(3):289–310.
- Sun, Ting. 2019. Applying deep learning to audit procedures: An illustrative framework. *Accounting Horizons* 33(3):89–109.
- Warning, Anja, Thorsten Sellhorn, und Jan-Pelle Kummer. 2020. Digitalisierung und Beschäftigung: Empirische Befunde für die Rechts- und Steuerberatung sowie Wirtschaftsprüfung. *Betriebswirtschaftliche Forschung und Praxis*. Im Erscheinen.
- Watts, Ross L., und Jerold L. Zimmerman. 1986. *Positive accounting theory*. Edgewood Cliff: Prentice-Hall.
- Yan, Zhaokai, und Kevin C. Moffitt. 2019. Contract analytics in auditing. *Accounting Horizons* 33(3): 111–126.