



# Lernende Entscheidungsbäume

## Überholtes Verfahren oder vielseitige KI-Methode?

Tobias Knuth<sup>1</sup>

Angenommen: 24. Juni 2021 / Online publiziert: 16. September 2021  
© Der/die Autor(en) 2021

### Zusammenfassung

Entscheidungsbäume bereichern moderne künstliche Intelligenz durch vielseitige Weiterentwicklungen. Lernende Entscheidungsbäume sind seit Jahrzehnten ein tradiertes Werkzeug zur Erstellung von erklärbaren Modellen zu Klassifikation und Regression. Während moderne Trends wie Deep Learning die Aufmerksamkeit der KI-Entwicklung auf sich ziehen, werden Entscheidungsbäume kontinuierlich weiterentwickelt, in neuen Bereichen eingesetzt und mit scheinbar im Wettstreit stehenden Methoden verschmolzen. Der Artikel stellt die Funktionsweise von Entscheidungsbäumen dar und zeigt Entwicklungen der aktuellen Forschung auf.

Entscheidungsbäume bereichern die explosionsartige Entwicklung künstlicher Intelligenz im 21. Jahrhundert auf vielfältige Weise durch Erklärbarkeit, leistungsstarke Ensembles und kreative Verwendungsmöglichkeiten wie zur Anomaliedetektion.

### Einleitung

Bäume als Datenstruktur bilden heute ein fundamentales Konzept der Informatik, wo sie in vielen Bereichen effiziente Verfahren ermöglichen, beispielsweise bei Such- und Sortierverfahren. Erstmals nutzte der englische Mathematiker Arthur Cayley Bäume zur Strukturierung chemischer Verbindungen im Jahre 1857 [22].

Im Bereich der künstlichen Intelligenz (KI), genauer des maschinellen Lernens, stellen Entscheidungsbäume eine tradierte Methode dar, Muster zu erlernen und vorherzusagen. Die ersten Lernverfahren für Entscheidungsbäume wurden zwischen 1960 und 1980 publiziert [16] und seitdem sind viele Varianten entstanden. Besonders das explosionsartige Interesse an künstlicher Intelligenz im 21. Jahr-

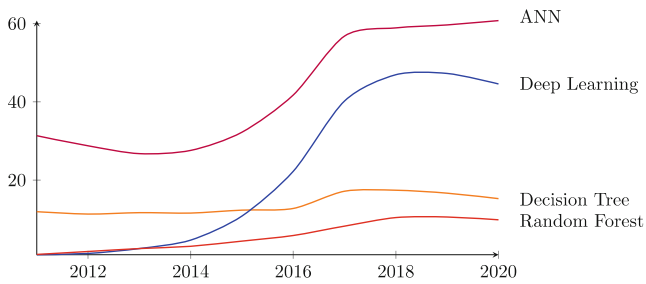
hundert, das mit zahlreichen Prognosen über den Einfluss von KI auf Wirtschaft und Gesellschaft einhergeht [4, 25], bringt ständig neue Verfahren hervor und Trends wie das *Deep Learning*, eine Spezialform neuronaler Netze – welche ebenfalls eine jahrzehntelange Tradition pflegen –, ziehen die Aufmerksamkeit auf sich.

Das Interesse an Entscheidungsbäumen und ihren Weiterentwicklungen wie Random Forests stellt Abb. 1 im Vergleich zu neuronalen Netzen und Deep Learning am Beispiel der Suchmaschine Google für die letzten 10 Jahre dar. Deutlich zu erkennen ist ein konstantes, aber niedrigeres Suchvolumen in Bezug auf Entscheidungsbäume. Daher stellt sich die Frage, welche Rolle Entscheidungsbäume heutzutage im Vergleich zu anderen Verfahren spielen und welche Antworten sie auf die aktuellen methodischen Anforderungen der künstlichen Intelligenz geben.

Zu diesen Anforderungen gehört neben dem Streben nach hoher Güte von Klassifikatoren, die anhand eines Merkmals, wie beispielsweise der Quote richtiger Entscheidungen (Accuracy), gemessen wird, auch eine Reihe von komplexen Themen, die KI-Systeme berücksichtigen müssen, um für die Gesellschaft nützlich zu sein. Sie beinhalten Fragen unter anderem in Bezug auf Ethik, Sicherheit und Korrektheit [23]. Ein interessanter Forschungszeitpunkt ist *Explainable AI (XAI)*, denn der Interpretierbarkeit und Erklärbarkeit von Modellen kommt eine wachsende Bedeutung zu [20]. So schafft beispielsweise die neue Datenschutzgrundverordnung ein *Recht zur Erklärung* von algorithmischen Entscheidungen [8], aber bei Weitem nicht

✉ Tobias Knuth  
contact@tobiasknuth.de

<sup>1</sup> FOM Hochschule für Ökonomie & Management gGmbH,  
Herkulesstraße 32, 45127 Essen, Deutschland



**Abb. 1** Relatives Suchinteresse nach Themengebieten (ohne Einheit), aggregiert pro Jahr. Die Daten stammen von Google Trends ([google.com/trends](https://google.com/trends))

alle KI-Verfahren lassen dies zu. Neuronale Netze gelten beispielsweise als Blackbox-Modelle.

Entscheidungsbäume werden in verschiedensten Bereichen eingesetzt, wie zur Betrugsprävention im Gesundheitswesen [14] und im Finanzbereich [17], zur (erklärenden) Untersuchung von Autounfällen [9] oder im Bereich Recommender Systems zur Generierung von Empfehlungen für Kunden [13]. In vergleichenden Studien teilen sich Entscheidungsbäume und neuronale Netze teilweise die Spitzenposition in Bezug auf die Popularität [14]. Entscheidungsbäume eignen sich durch ihre natürliche Transparenz für alle Anwendungen, in denen ein Verständnis der Entscheidungen unverzichtbar ist, beispielsweise in medizinischen Anwendungen oder bei potenziell diskriminierenden Fragestellungen, zu denen Einschätzungen zur Kreditwürdigkeit oder dem Betrugsrisiko gehören, die mit persönlichen bzw. demografischen Daten beantwortet werden sollen. Darüber hinaus können Entscheidungsbäume helfen, Wissen aus Daten zu extrahieren und explizit darzustellen. Entscheidungsbäume lassen sich leicht visualisieren, ähneln den Entscheidungsprozessen von Menschen und können in Regeln übersetzt werden [12, 16].

Der Artikel leistet 3 Beiträge: Erstens wird ein konzeptioneller Rahmen zur Beschreibung der Komponenten maschinellen Lernens vorgestellt, in den Entscheidungsbäume eingeordnet werden. Zweitens werden die charakteristischen Ideen von Entscheidungsbäumen anschaulich dargestellt. Drittens werden Entscheidungsbäume aus verschiedenen Perspektiven beleuchtet, um die vielseitigen Einsatzgebiete dieser Technik aufzuzeigen, darunter leistungsstarke Ensemblemethoden wie Gradient Boosted Trees und Random Forests sowie ein kreativer Ansatz zur Anomalieerkennung mittels Isolation Forests.

## Komponenten maschinellen Lernens

Der Begriff der künstlichen Intelligenz umfasst Disziplinen wie die Entwicklung intelligenter Agenten, Robotik und auch das maschinelle Lernen (ML). Eine klassische De-

finition maschinellen Lernens besagt, dass ein Programm lernt, wenn die an einer Metrik gemessene Leistung des Programms bei der Bewältigung einer definierten Aufgabe mit zunehmender Erfahrung steigt [16]. Die Verfahren werden in überwachtes, semiüberwachtes und unüberwachtes Lernen unterteilt. Ein Modell lernt überwacht, wenn neben den Eingabedaten (unabhängigen Variablen, Features, Prädiktoren) auch die Daten der Zielvariablen bereitstehen, also Referenzwerte für richtige Entscheidungen existieren. Entscheidungsbäume fallen in diese Kategorie. Im Gegensatz hierzu erfordern unüberwachte Lernverfahren keine Referenzwerte und semiüberwachtes Lernen verbindet beide Ansätze miteinander.

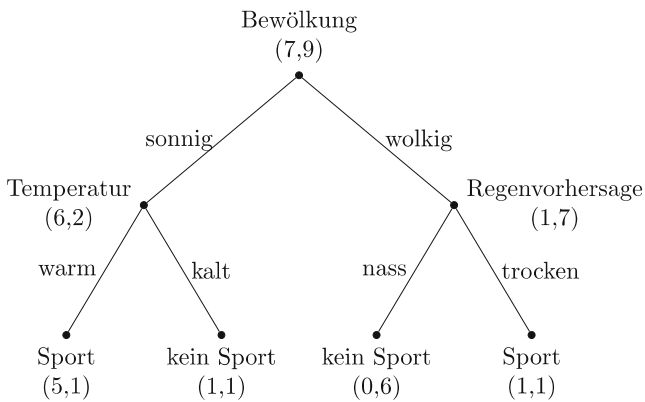
Eine andere Form der Strukturierung orientiert sich an 3 Komponenten, in die sich jedes ML-Verfahren zerlegen lässt: Repräsentation, Evaluation und Optimierung. Die Repräsentation beschreibt die Form des Modells, wie die Geradengleichung einer linearen Regression oder die Struktur und die Verbindungsgewichte eines neuronalen Netzes. Die Evaluation bezieht sich auf die Frage, wie die Güte des Modells gemessen werden kann; dies geschieht beispielsweise mit der Summe der Fehlerquadrate oder der Entropie. Die Optimierung stellt das Werkzeug, um von einem Gütezustand in einen besseren zu gelangen. Hierfür stehen je nach Problemstellung Methoden wie der Gradientenabstieg oder lineare Programmierung zur Verfügung [5, 16].

Entscheidungsbäume werden durch die Baumstruktur mit ihrer Wurzel, den Verzweigungen, inneren Knoten und Blättern repräsentiert. Ein traditionelles Optimierungsverfahren stellt an jeder Verzweigung die Frage: „Welche Variable sollte für die nächste Verzweigung verwendet werden?“ Die Evaluation erfolgt abhängig von Klassifikation oder Regression beispielsweise mit der Quote richtiger Entscheidungen (Accuracy). Diese Komponenten werden im folgenden Abschnitt kurz vorgestellt.

## Grundlagen von Entscheidungsbäumen

Seit den ersten Publikationen sind zahlreiche algorithmische Varianten entwickelt worden. Im Folgenden werden die zentralen Gedanken des bekannten ID3-Algorithmus vorgestellt [16, 19, 24], neben Varianten wie C4.5 und CART [11]. ID3 ist ein Greedy-Algorithmus, d. h. das Lernproblem wird in aufeinanderfolgende Teilschritte zerlegt, die einzeln optimiert werden; mittlerweile existieren auch Ansätze zur globalen Optimierung [18].

Abb. 2 zeigt einen Entscheidungsbaum mit diskreten Variablen an den Knoten. Die 3 Variablen *Bewölkung*, *Temperatur* und *Regenvorhersage* werden herangezogen, um zu einer Entscheidung zu gelangen. Zur Prognose eines Werts folgt man dem jeweiligen Pfad des Baums und liest das Ergebnis ab. Ist es beispielsweise sonnig und warm, wird



**Abb. 2** Exemplarischer, manuell konstruierter Entscheidungsbaum zur Frage, ob heute Sport gemacht werden sollte

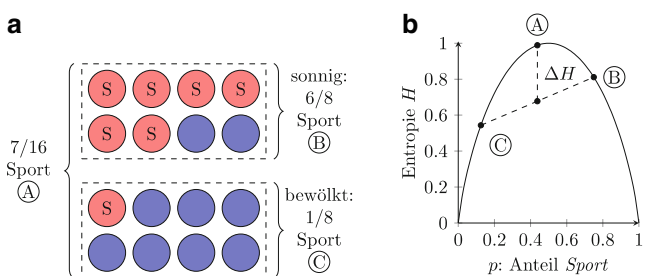
Sport gemacht. Ist es wolzig und nass, wird kein Sport gemacht. Die Zahlen in Klammern geben an, wie viele Instanzen des Datensatzes auf den jeweiligen Pfad zutreffen (Sport, kein Sport), d. h. im Datensatz wurde insgesamt 7-mal Sport gemacht und 9-mal nicht; an sonnigen und warmen Tagen wurde 5-mal Sport gemacht und 1-mal nicht. Man beachte, dass in diesem Beispiel anhand der verfügbaren Daten keine fehlerfreie Klassifikation möglich ist.

Entscheidungsbäume haben darüber hinaus die nützliche Eigenschaft, dass sie sich in Mengen von alternativen Regeln übersetzen lassen. Der Baum in der Abbildung kann beispielsweise ebenfalls durch die folgenden Regeln ausgedrückt werden, welche zu denselben Klassifikationsergebnissen führen. Jeder Pfad des Baums entspricht hierbei einer *und*-verknüpften Bedingung. Alle Pfade desselben Ergebnisses sind *oder*-verknüpft.

$$(sonnig \wedge warm) \vee (wolzig \wedge trocken) \Rightarrow Sport$$

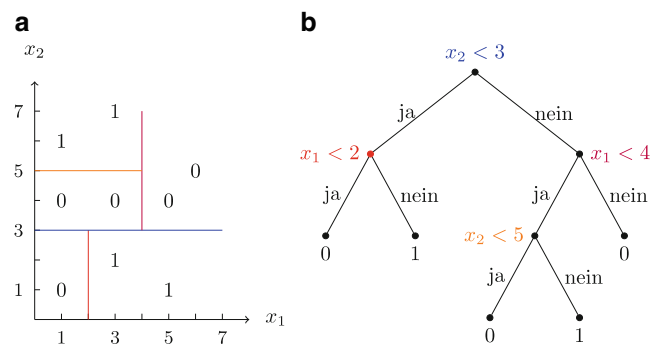
$$(sonnig \wedge kalt) \vee (wolzig \wedge nass) \Rightarrow kein Sport$$

Wie aber wird ein Baum wie oben dargestellt konstruiert? Abb. 3 stellt den iterativen Teilungsprozess, der dem

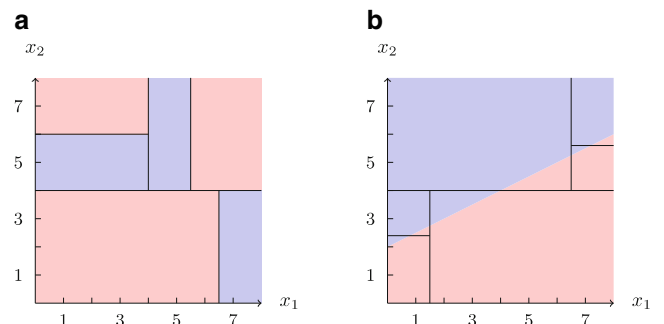


**Abb. 3** Die Entropie ( $H = \sum_{i=1}^2 -p_i \log_2 p_i$ ) gibt das Maß der Unsicherheit an, wobei  $p$  den prozentualen Anteil der Variablen von Interesse darstellt. Je besser die Daten (a) sich anhand der Ausprägungen einer Variablen in *Sport* und *kein Sport* unterteilen lassen, desto kleiner die Entropie und desto größer der Informationsgewinn durch die Verzweigung (b)

ID3-Algorithmus (Iterative Dichotomiser) seinen Namen gibt, anhand der ersten Variablen dar. Der Baum in Abb. 2 beginnt mit der Variable *Bewölkung*, weil diese den größten *Information Gain* bietet. Information Gain bezeichnet die Differenz zwischen 2 Entropiewerten, einem Maß für die Entscheidungsunsicherheit. In der Abbildung besteht der gesamte Datensatz A – wie oben – aus 7 Instanzen, in denen Sport gemacht wird und 9, in denen kein Sport gemacht wird. Die Entropie ist in der Abbildung (rechts) entsprechend hoch, da beide Ereignisse fast mit gleicher Wahrscheinlichkeit auftreten. Wird der Datensatz nun anhand der *Bewölkung* in sonnig und bewölkt unterteilt, sinkt das gewichtete Mittel der Teildatensätze B (enthält nun überwiegend Sportinstanzen) und C (enthält nun fast keine Sportinstanzen) und liegt nun unter A, d. h. die Klassifikation der Teildatensätze gelingt besser als die Klassifikation des Gesamtdatensatzes. Dieses Greedy-Verfahren wird an den Knoten wiederholt, bis ein Abbruchkriterium erreicht ist, spätestens aber, wenn jeder Pfad des Baums zu einer eindeutigen Klassifikation führt oder keine Variablen zum Teilen mehr verfügbar sind.



**Abb. 4** Entscheidungsbaum (b) mit eingezeichneten Entscheidungsgrenzen (a). Die Ungleichungen (Tests) in b entsprechen den Geraden in a, welche die Daten in die zu prognostizierenden Kategorien unterteilen, in diesem Beispiel in Nullen und Einsen



**Abb. 5** Die exemplarischen, rechtwinkligen und nichtlinearen Entscheidungsgrenzen (a) lassen sich durch den Entscheidungsbaum perfekt abbilden. Die linearen Entscheidungsgrenzen (b) können nur approximiert werden, welche eine lineare Regression problemlos modellieren könnte

Numerische Attribute lassen sich mit Ungleichungen integrieren, wie Abb. 4b mithilfe der numerischen Variablen  $x_1$  und  $x_2$  zeigt. Die durch die Ungleichungen entstehenden Regionen lassen sich visualisieren (Abb. 4a). Jede Ungleichung einer Variablen unterteilt die entsprechende Region im Diagramm in 2 Subregionen. Beispielsweise definiert die Ungleichung  $x_1 < 2$  im Baum links in Verbindung mit der Wurzel  $x_2 < 3$  das Rechteck unten links in der Abbildung von den Koordinaten (0,0) bis (2,3). Da Entscheidungsbäume den Raum rechtwinklig partitionieren, sind sie strukturell benachteiligt, wenn Entscheidungsgrenzen beispielsweise diagonal verlaufen. Eine Diagonale kann in diesem Falle nur durch häufige Unterteilung angenähert werden, ähnlich wie eine Treppe eine Gerade annähert, je kleiner die Stufen sind. Diesen Sachverhalt stellt Abb. 5 dar. Dass die Stärke von Entscheidungsbäumen in der Verarbeitung nichtlinearer Probleme mit diskreten Variablen liegt, bestätigt eine vergleichende Studie zur Diagnose von Hautläsionen, in der die Bäume aufgrund des hohen Anteils numerischer Variablen schlechtere Ergebnisse erzielten als alternative Verfahren [6].

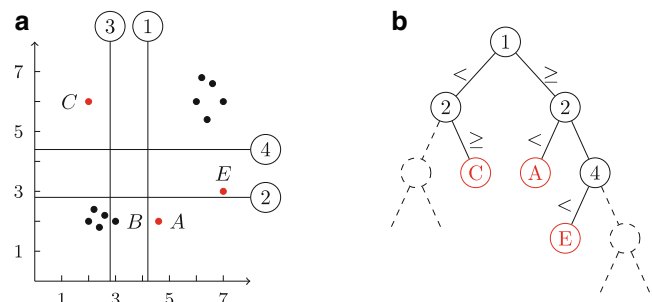
Entscheidungsbäume tendieren dazu, lang und komplex zu werden, um die Trainingsdaten bestmöglich abzubilden. Dies führt zu *Overfitting* [12], d.h. das Modell bildet die Trainingsdaten besser ab als neue, unbekannte Daten, was falsche Sicherheit vermittelt und zu vermeiden ist. Kürzere, einfachere Bäume sind daher zu bevorzugen und diverse, teilweise nachträgliche Kürzungsverfahren (Pruning) von Bäumen existieren. Der Preis für die Verständlichkeit und Einfachheit von Entscheidungsbäumen ist ihre im Vergleich zu anderen Verfahren geringere Genauigkeit und Robustheit [10, 12, 27], d.h. Klassifikationen sind relativ betrachtet häufiger falsch und kleine Änderungen der Daten können grundlegende Änderungen der Baumstruktur bewirken. Dies ist allerdings kein spezieller Nachteil von Entscheidungsbäumen, sondern Verständlichkeit und Leistungsfähigkeit von ML-Modellen scheinen in einem inversen Verhältnis zueinanderzustehen [10]. Innerhalb der Familie der Entscheidungsbäume existieren unter anderem Random Forests und Gradient Boosted Trees, welche diese Nachteile ausgleichen und in den folgenden Abschnitten vorgestellt werden. Daneben wird ein kreatives Verfahren von Entscheidungsbäumen zur Anomaliedetektion vorgestellt, die Isolation Forests.

### Fortgeschrittene Methoden

Eine typische Strategie zur Stärkung von Modellen im maschinellen Lernen ist der Einsatz von Ensembles. Random Forests bestehen aus einer Vielzahl demokratisch abstim-mender einzelner Bäume, deren Variablenauswahl zufällig variiert wird. Das Ergebnis sind leistungsstarke Modelle. In-

tensive Debatten mit den Titeln „Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?“ zugunsten der Random Forests und „Are Random Forests Truly the Best Classifiers?“ als Antwort darauf werden zu der Frage geführt, wie leistungsstark Random Forests sein können [7, 26]. Eine weitere Alternative stellen Gradient Boosted Trees dar [1], welche nicht wie bei Random Forests gleichzeitig eine Vielzahl von Bäumen erstellen, sondern nacheinander, sodass jeder zusätzliche Baum eine Verbesserung darstellt. Man kann sich dies wie ein Team vorstellen, in dem jedes neue Mitglied eine Funktion übernimmt, die dem Team noch fehlte. Beide Varianten verschaffen Entscheidungsbäumen den Sprung in die Oberklasse der überwachten Lernalgorithmen, büßen dabei aber durch die Ensemblemethode die besondere Erklärbarkeit der Bäume ein.

Eine besonders kreative Anwendung von Entscheidungsbäumen sind die Isolation Forests zur Anomaliedetektion [15]. Anomalien sind Auffälligkeiten in Daten. Eine Form sind räumliche Anomalien, d.h. solche Datenpunkte sind ungewöhnlich, die von anderen weit entfernt liegen. Der Gedanke hinter Isolation Forests lautet: Wenn Entscheidungsgrenzen so lange *zufällig* gesetzt werden, bis jeder Datenpunkt allein in einem eigenen abgegrenzten Kästchen liegt (d.h. isoliert wurde), ist es wahrscheinlicher, dass Anomalien schneller isoliert werden. Das bedeutet, dass Anomalien im Baum näher an der Wurzel sind. Mittels eines Isolation Forests lässt sich die durchschnittliche Höhe eines Punkts im Baum als Anomaliewert interpretieren. Das Konzept eines einzelnen Isolation Trees ist in Abb. 6 dargestellt. Je weiter ein Punkt von anderen entfernt ist, desto schneller lässt er sich isolieren. Die Abbildung rechts zeigt, dass die Verwendung der Linien 1, 2 und 4 bereits ausreicht, um alle 3 Anomalien zu finden. In der Praxis funktioniert die Methode erstaunlich gut, obwohl Entscheidungsbäume konzeptionell weit von klassischen Verfahren der Anomaliedetektion entfernt sind. Darüber hinaus zeichnen sich Isolation Forests durch lineare Laufzeit und hohe Skalierbarkeit aus [15].



**Abb. 6** Isolation Tree. Die zufällig gesetzten Linien 1–4 (a) unterteilen den Raum in 9 Rechtecke. Nach 4 Schritten sind bereits alle 3 Anomalien isoliert und im Baum vertreten

## Schluss

Entscheidungsbäume zählen mittlerweile zu den klassischen Methoden maschinellen Lernens und haben sich im Laufe der Jahrzehnte in viele Richtungen weiterentwickelt. Einzelne Entscheidungsbäume zeichnen sich durch ihre Erklärbarkeit aus, Random Forests und Gradient Boosted Trees erreichen beeindruckende Ergebnisse im Vergleich zu führenden Alternativen wie neuronalen Netzen. Kreative Anwendungen wie Isolation Forests führen Entscheidungsbäume in untypische Bereiche wie die Anomaliedetektion.

Zuletzt lassen sich einzelne, erklärbare Entscheidungsbäume mit Blackbox-Modellen auf verschiedene Arten verschmelzen [28] oder ergänzen: Entscheidungsbäume werden nicht anhand eines Datensatzes trainiert, sondern mithilfe eines vortrainierten Blackbox-Modells, wie eines neuronalen Netzes. Die Nutzung von sogenannten Post-Hoc-Modellen besteht daher aus 2 Schritten: Zuerst wird mithilfe eines Datensatzes ein Blackbox-Modell trainiert und im Anschluss wird der Output dieses Modells verwendet, um einen Entscheidungsbaum zu konstruieren. Faszinierend ist dabei, dass Post-Hoc-Modelle, die dadurch lernen, das Blackbox-Modell zu imitieren, bessere Ergebnisse als Entscheidungsbäume erzielen, die direkt mit den ursprünglichen Daten trainiert wurden [2, 3, 21]; die Blackbox-Modelle gehen mit den Entscheidungsbäumen eine symbiotische Beziehung ein, ein spannendes Forschungsfeld mit Potenzial für erklärbare und leistungsstarke Modelle. Die traditionellen Entscheidungsbäume werden ständig weiterentwickelt und stellen ein vielseitiges Werkzeug für moderne KI-Anwendungen dar.

**Funding** Open Access funding enabled and organized 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>.

## Literatur

1. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. arXiv. <https://doi.org/10.1145/2939672.2939785>
2. Confalonieri R, Weyde T, Besold TR, Moscoso del Prado Martín F (2021) Using ontologies to enhance human understandability of global post-hoc explanations of black-box models. *Artif Intell* 296:103471
3. Craven MW, Shavlik JW (1995) Extracting tree-structured representations of trained networks. Proceedings of the 8th International Conference on Neural Information Processing Systems. NIPS'95. MIT Press, Cambridge, S 24–30
4. Davenport T, Guha A, Grewal D, Bressgott T (2020) How artificial intelligence will change the future of marketing. *J of the Acad Mark Sci* 48(1):24–42
5. Domingos P (2012) A few useful things to know about machine learning. *Commun ACM* 55(10):78–87
6. Dreiseitl S, Ohno-Machado L, Kittler H, Vinterbo S, Billhardt H, Binder M (2001) A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *J Biomed Inform* 34(1):28–36
7. Fernández-Delgado M, Cernadas E, Barro S, Amorim D (2014) Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res* 15(1):3133–3181
8. Goodman B, Flaxman S (2016) European Union regulations on algorithmic decision-making and a “right to explanation”. *AI Mag* 38(3):50–57 (vorgestellt auf dem 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY)
9. Griselda L, Juan DO, Joaquín A (2012) Using decision trees to extract decision rules from police reports on road accidents. *Procedia Soc Behav Sci* 53:106–114
10. Gunning D, Aha D (2019) DARPA’s explainable artificial intelligence (XAI) program. *AI Mag* 40(2):44–58
11. Hastie T, Tibshirani R, Friedman J (2013) The elements of statistical learning. Springer, New York
12. James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. Springer texts in statistics, Bd. 103. Springer, New York
13. Karimi R, Nanopoulos A, Schmidt-Thieme L (2015) A supervised active learning framework for recommender systems based on decision trees. *User Model User-Adapt Interact* 25(1):39–64
14. Li J, Huang KY, Jin J, Shi J (2008) A survey on statistical methods for health care fraud detection. *Health Care Manage Sci* 11(3):275–287
15. Liu FT, Ting KM, Zhou ZH (2008) Isolation forest. 2008 Eighth IEEE International Conference on Data Mining. IEEE, Pisa, S 413–422
16. Mitchell T (1997) Machine learning. McGraw-Hill series in computer science. McGraw-Hill, Singapur
17. Ngai EWT, Hu Y, Wong YH, Chen Y, Sun X (2011) The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. *Decis Support Syst* 50(3):559–569
18. Norouzi M, Collins MD, Johnson M, Fleet DJ, Kohli P (2015) Efficient non-greedy optimization of decision trees. NIPS’15 Proceedings of the 28th International Conference on Neural Information Processing Systems. MIT Press, Cambridge, S 1729–1737
19. Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1(1):81–106
20. Rai A (2020) Explainable AI: from black box to glass box. *J of the Acad Mark Sci* 48(1):137–141
21. Renard X, Woloszko N, Aigrain J, Detyniecki M (2019) Concept tree: high-level representation of variables for more interpretable surrogate decision trees. ICML Workshop on Human in the Loop Learning (HILL 2019), Long Beach, USA. arXiv:1906.01297

22. Rosen KH (2012) Discrete mathematics and its applications, 7. Aufl. McGraw-Hill, New York
23. Russell S, Dewey D, Tegmark M (2015) Research priorities for robust and beneficial artificial intelligence. *AI Mag* 36(4):105
24. Russell SJ, Norvig P, Davis E (2010) Artificial intelligence: a modern approach, 3. Aufl. Prentice Hall series in artificial intelligence. Prentice Hall, Upper Saddle River
25. Shankar V (2018) How artificial intelligence (AI) is reshaping retailing. *J Retail* 94(4):vi–xi
26. Wainberg M, Alipanahi B, Frey BJ (2016) Are random forests truly the best classifiers? *J Mach Learn Res* 17:1–5
27. Xu F, Uszkoreit H, Du Y, Fan W, Zhao D, Zhu J (2019) Explainable AI: a brief survey on history, research areas, approaches and challenges. In: Tang J, Kan MY, Zhao D, Li S, Zan H (Hrsg) Natural language processing and Chinese computing, Bd. 11839. Springer, Cham, S 563–574
28. Yang Y, Morillo IG, Hospedales TM (2018) Deep neural decision trees. 2018 ICML Workshop on Human Interpretability in Machine Learning (WHI 2018), Stockholm, Sweden. arXiv:1806.06988