Tobias Bär

Algorithmic Bias: Verzerrungen durch Algorithmen verstehen und verhindern

Ein Leitfaden für Entscheider und Data Scientists

Springer Vieweg

ALGORITHMIC BIAS: VERZERRUNGEN DURCH ALGORITHMEN VERSTEHEN UND VERHINDERN

EIN LEITFADEN FÜR ENTSCHEIDER UND DATA SCIENTISTS

Tobias Bär

Springer Vieweg

Algorithmic Bias: Verzerrungen durch Algorithmen verstehen und verhindern: Ein Leitfaden für Entscheider und Data Scientists

Tobias Bär
Taipei, Taiwan

ISBN-13 (pbk): 978-3-662-66314-1 ISBN-13 (electronic): 978-3-662-66315-8
https://doi.org/10.1007/978-3-662-66315-8

Geschäftsführender Direktor, Apress Media LLC: Welmoed Spahr
Editor für Akquisitionen: Shiva Ramachandran
Entwicklungsredakteurin: Laura Berendson
Koordinierender Herausgeber: Rita Fernando

Umschlag gestaltet von eStudioCalamar

Springer Vieweg ist ein Imprint der eingetragenen Gesellschaft APress Media, LLC und ist ein Teil von Springer Nature.
Die Anschrift der Gesellschaft ist: 1 New York Plaza, New York, NY 10004, U.S.A.

booktranslations@springernature.com; bookpermissions@springernature.com.

Apress-Titel können in großen Mengen für akademische Zwecke, Unternehmen oder Werbezwecke erworben werden. Für die meisten Titel sind auch eBook-Versionen und -Lizenzen erhältlich. Weitere Informationen finden Sie auf unserer Webseite für Print- und eBook-Massenverkäufe unter www.apress.com/bulk-sales.

Jeglicher Quellcode oder anderes ergänzendes Material, auf das der Autor in diesem Buch verweist, steht den Lesern auf GitHub über die Produktseite des Buches zur Verfügung, die sich unter www.apress.com/9783662663141 befindet. Für weitere Informationen besuchen Sie bitte www.apress.com/source-code.

Gedruckt auf säurefreiem Papier

Für den Liebesalgorithmus im Kopf meines Partners –
ich weiß immer noch nicht, welche kognitive
Verzerrung ihn veranlasst hat, sich für mich zu
entscheiden, aber ich denke, es ist der beste Fehler, den
er in seinem Leben gemacht hat!

Inhaltsverzeichnis

Über den Autor

Tobias Bär ist Data Scientist, Psychologe und Top-Management-Berater mit über 20 Jahren Erfahrung in der Risikoanalyse. Bis Juni 2018 war er Master Expert und Partner bei McKinsey & Co. und baute in Indien ab 2004 McKinsey's Risk Advanced Analytics Center of Competence auf, leitete die Credit Risk Advanced Analytics Service Line weltweit und beriet Klienten in über 50 Ländern zu Themen wie der Entwicklung analytischer Entscheidungsmodelle für die Kreditvergabe, die Preisgestaltung von Versicherungen und das Steuerinkasso sowie die Bekämpfung von Vorurteilen in qualitativen Entscheidungen. Dr. Bär hat eine Forschungsagenda rund um Analytik und Entscheidungsfindung verfolgt, sowohl bei McKinsey (z. B. zur Objektivierung subjektiver Entscheidungen und zur Nutzung von maschinellem Lernen zur Entwicklung hochtransparenter Vorhersagemodelle) als auch an der University of Cambridge, UK (z. B. die Auswirkung mentaler Ermüdung auf Entscheidungsvoreingenommenheit).

Dr. Bär promovierte in Finanzwissenschaften an der Universität Frankfurt, besitzt einen MPhil in Psychologie von der Universität Cambridge und einen MA in Volkswirtschaft von der UWM und hat ein Grundstudium in Betriebswirtschaft und Recht an der Universität Gießen absolviert. Er begann bereits als Teenager, in einem deutschen Software-Magazin über Programmiertricks für den Commodore C64 Heimcomputer zu schreiben, und bloggt nun regelmäßig auf seiner LinkedIn-Seite, www.linkedin.com/in/tobiasbaer/.

Danksagungen

Zuallererst möchte ich meiner Verlegerin, Shiva Ramachandran, danken, der das alleinige Lob für die brillante Idee gebührt, dieses Buch zu schreiben, und meiner Lektorin, Rita Fernando, die durch ihre unermüdliche Ermutigung nicht nur eine schreibende Bestie in mir entfesselt hat, sondern auch die rote Tinte von meinem schrulligen Humor fernhielt. Sie ist an allem schuld – ich hatte viel mehr Aufsicht durch Erwachsene erwartet!

Ich möchte auch dem (inzwischen emeritierten) Professor Paul Shaman vom Fachbereich Statistik der Wharton School der University of Pennsylvania danken, bei dem ich 1999 zwei wertvolle Monate als Gastwissenschaftler verbringen durfte. Er öffnete mir die Augen für den Unterschied zwischen dem Ausführen eines Skripts zur Schätzung eines Modells und dem Verständnis von Daten – ein Großteil meiner kritischen Haltung gegenüber Daten geht auf seine Lehren zurück.

Schließlich möchte ich mich bei Clemens Baader bedanken, der freundlicherweise den Manuskriptentwurf gelesen hat und immer ein fabelhafter Gesprächspartner für meine Ideen war.

Vorwort

Warum habe ich dieses Buch geschrieben? Über algorithmische Verzerrungen ist bereits viel geschrieben worden; beunruhigende Beispiele für algorithmische Verzerrungen gibt es zuhauf. Über die tatsächlichen Ursachen algorithmischer Verzerrungen ist jedoch viel weniger geschrieben worden, und es scheint sehr wenig darüber bekannt zu sein, wie man das Problem lösen und algorithmische Verzerrungen entweder ganz verhindern oder so handhaben kann, dass sie keinen Schaden anrichten. Genau darum geht es in diesem Buch.

Dieses Buch ist praktisch. Es schlägt Lösungen vor, mit deren Umsetzung Sie schon morgen beginnen können. Einige der Maßnahmen können einige Zeit in Anspruch nehmen, bis sie abgeschlossen sind oder Früchte tragen – aber in diesem Buch geht es nicht um ausgefallene Theorie. Es gibt Schritt-für-Schritt-Anleitungen und Checklisten sowie unzählige Beispiele aus dem wirklichen Leben, um meine Argumente zu veranschaulichen. Vor allem aber regt dieses Buch zum kritischen Denken an, indem es vorschlägt, welche spezifischen Fragen man stellen sollte.

Je mehr ich bei meiner eigenen Modellierungs- und Beratungstätigkeit über algorithmische Verzerrungen herausfand, desto mehr wurde mir klar, dass es sich um weit mehr als ein technisches Problem handelt. Ja, die Statistik liefert sowohl einige der Ursachen für algorithmische Verzerrungen als auch einige der Lösungen. Das Problem ist jedoch tief in der menschlichen Psychologie verwurzelt, und wir können algorithmische Voreingenommenheit nicht angehen, ohne zu verstehen, wie menschliche Voreingenommenheit und die Vorurteile von Nutzern, Datenwissenschaftlern und der Gesellschaft im Allgemeinen Entscheidungsvoreingenommenheit schaffen und verbreiten.

Deshalb stürze ich mich nicht gleich in technische Lösungen, sondern nehme mir die Zeit, zu erklären, woher algorithmische Verzerrungen kommen – und was das für ihre Bekämpfung bedeutet.

Und die (nicht-technischen) Nutzer von Algorithmen – wie Manager und Beamte – haben viel mehr Möglichkeiten, algorithmische Verzerrungen zu bekämpfen und zu verhindern, als sie vielleicht glauben. Dieses Buch möchte alle dazu befähigen, besser mit algorithmischen Verzerrungen umzugehen und sie gemeinsam zu verhindern.

Für wen dieses Buch bestimmt ist

Wir leben in einer Welt, in der wir alle von Algorithmen betroffen sind, und viele von uns nutzen sie, vielleicht sogar ohne zu wissen, dass es sich um einen Algorithmus handelt. Deshalb habe ich dieses Buch für uns alle geschrieben.

Datenwissenschaftler sind die wenigen Experten, die Algorithmen entwickeln, und spielen daher eine wichtige Rolle im Umgang mit und bei der Vermeidung von algorithmischer Voreingenommenheit. Ich hoffe daher, dass viele, vielleicht sogar alle, dieses Buch lesen werden – und der letzte, technischste Teil dieses Buches ist sogar ihnen gewidmet.

Die meisten Menschen sind jedoch keine Datenwissenschaftler, und viele hassen Statistik geradezu. Das Buch ist daher mit Blick auf den Laien geschrieben. Es verwendet eine nicht-technische Sprache, anschauliche Analogien und versucht, den Spaßfaktor durch übermäßigen Einsatz von Humor hochzuhalten. Warnung! Am Ende könnte man Statistik sogar mögen, zumindest das verzerrte Bild, das dieses Buch von ihr vermittelt …

Da das Thema der algorithmischen Verzerrung zunehmend Aufmerksamkeit erfahren hat, haben natürlich auch Compliance-Teams und Aufsichtsbehörden begonnen, sich damit zu befassen und nach Wegen zu suchen, um Schaden durch Algorithmen zu verhindern. Daher richtet sich dieses Buch nicht nur an die eigentlichen Entwickler und Nutzer von Algorithmen – und an die Unternehmensleiter und Beamten, die entscheiden, wo und wie sie eingesetzt werden –, sondern auch an Compliance-Teams und Regulierungsbehörden, deren Aufgabe es ist, die Entscheidungsprozesse zu kontrollieren.

Und ich werde argumentieren, dass viele algorithmische Vorurteile ein Spiegel tief verwurzelter gesellschaftlicher Vorurteile sind. Daher ist das Problem der algorithmischen Voreingenommenheit ein viel größeres, und ich habe dieses Buch auch für Politiker, Journalisten und Philosophen geschrieben, die wissen müssen, dass Algorithmen sowohl eine Lösung für die Bekämpfung gesellschaftlicher Vorurteile als auch ein Problem sein können, wenn sie solche Voreingenommenheit aufrechterhalten und verstärken.

Nicht zuletzt ist das Buch für Marsmenschen und Zeta-Reticulaner gedacht. Sie werden bald herausfinden, warum!

Was dieses Buch nicht ist

Dieses Buch ist kein Statistik-Lehrbuch. Es verweist auf zahllose statistische Techniken für die Datenwissenschaftler (und interessierten Laien) unter den Lesern – aber es wird sie nicht erklären. Datenwissenschaftler kennen die meisten dieser Techniken bereits oder wissen zumindest, wo sie sie nachschlagen können.

Dieses Buch ist auch kein juristisches Lehrbuch. Es behandelt zwar rechtliche und ethische Fragen auf philosophischer Ebene – einschließlich der Frage, wie die Europäische Datenschutzgrundverordnung einige zentrale Erkenntnisse über Algorithmen sowohl anerkennt als auch verfehlt –, aber es zielt nicht darauf ab, alle Gesetze zu katalogisieren, die für den Umgang mit algorithmischer Voreingenommenheit irgendwie relevant sind, oder Anleitungen zu geben, wie man bestimmte rechtliche Anforderungen erfüllt. Dafür braucht man Juristen – am besten solche, die dieses Buch auch gelesen haben.

Schließlich ist dieses Buch kein Patentrezept. Vorurteile zu bekämpfen ist schwer. In gewissem Sinne sind Vorurteile eine Form der Konformität – Konformität mit dem, „wie es ist", was Ihr Chef sagt, was die Daten sagen, was Ihr fauler Verstand sagt (weil Sie es immer so gemacht haben). Es besteht keine Chance, dass Sie aus der Lektüre dieses Buches irgendeinen Nutzen ziehen, wenn Sie nicht einige der Dinge, die Sie tun, ändern. Ich lade Sie ein, immer wieder darüber nachzudenken, was die Erkenntnisse aus diesem Buch für Sie bedeuten und was Sie aufgrund des Gelesenen anders machen können. Ich würde mich freuen, wenn Sie einen Kommentar in meinem Blog unter `www.linkedin.com/in/tobiasbaer/` hinterlassen. Und wenn Sie verhindern wollen, dass Ihnen all Ihre guten Ideen entgleiten und Sie wieder in alte Gewohnheiten zurückfallen, nachdem Sie dieses Buch gelesen und einen guten Platz dafür in Ihrem Bücherregal gefunden haben, wo es reichlich allergenen Staub ansammeln kann, sollten Sie sich vielleicht sogar gleich eine Erinnerung in Ihren Kalender eintragen!

Wie dieses Buch aufgebaut ist

Das Buch besteht aus vier Teilen. Der erste Teil ist eine Einführung – er behandelt die Psychologie menschlicher Voreingenommenheit sowie die Verwendung und Entwicklung von Algorithmen. In den Kapiteln des ersten Teils werden die Terminologie und die Rahmenbedingungen erläutert, auf die ich im weiteren Verlauf des Buches immer wieder Bezug nehmen werde.

Im zweiten Teil werden sechs verschiedene Quellen von algorithmischen Verzerrungen vorgestellt. Das Verständnis dieser Quellen ist die Grundlage für den Umgang mit und die Vermeidung von algorithmischen Verzerrungen und wird daher im weiteren Verlauf des Buches immer wieder erwähnt.

Im dritten Teil wird erörtert, wie Nutzer von Algorithmen (im weitesten Sinne definiert als jeder, der kein Datenwissenschaftler ist) mit algorithmischer Voreingenommenheit umgehen können und welche wirksamen Möglichkeiten sie haben, diese zu verhindern.

Und der vierte und letzte Teil bietet Data Scientists umfassende, praktische Anleitungen zur Vermeidung algorithmischer Verzerrungen durch spezifische Techniken für die Entwicklung und Implementierung. Dieser Teil des Buches

ist daher der technischste – aber ich habe ihn dennoch so geschrieben, dass jeder, vom Studenten bis zum erfahrenen Leiter der Abteilung Analytik, ihm folgen und wertvolle Erkenntnisse gewinnen kann.

Jeglicher Quellcode oder anderes ergänzendes Material, auf das der Autor in diesem Buch verweist, ist für die Leser auf GitHub über die Produktseite des Buches verfügbar, die sich unter www.apress.com/9783662663141 befindet. Für weitere Informationen besuchen Sie bitte www.apress.com/source-code.

TEIL

I

Eine Einführung in Verzerrungen und Algorithmen

KAPITEL

1

Einführung

Was ist eine Verzerrung? Eine viel zitierte Quelle[1] definiert sie wie folgt:

> *Neigung oder Vorurteil für oder gegen eine Person oder Gruppe, insbesondere in einer Weise, die als ungerecht empfunden wird.*

Vorurteile sind zweischneidige Schwerter. Wie Sie im nächsten Kapitel sehen werden, sind Vorurteile in der Regel kein Charakterfehler oder eine seltene Abweichung, sondern vielmehr der notwendige Preis dafür, dass der menschliche Verstand jeden Tag Tausende von Entscheidungen scheinbar mühelos und blitzschnell treffen kann. Haben Sie sich schon einmal darüber gewundert, wie Sie einem sich schnell bewegenden Objekt, z. B. einem Auto, das Sie zu überfahren drohte, in Sekundenbruchteilen entkommen konnten? Neurowissenschaftler und Psychologen haben begonnen, die Geheimnisse des Verstandes zu entschlüsseln, und haben herausgefunden, dass das Gehirn diese Geschwindigkeit nur durch zahlreiche Abkürzungen erreichen kann.

Eine Abkürzung bedeutet, dass der Verstand vorschnell eine Schlussfolgerung zieht (z. B. ein Gericht für ungenießbar oder einen Fremden für gefährlich hält), ohne alle Fakten gebührend zu berücksichtigen. Mit anderen Worten: Der Verstand nutzt Vorurteile, um schneller zu sein.

Die Verwendung von Vorurteilen bei der Entscheidungsfindung ist daher insofern ungerecht, als sie bestimmte Fakten, die für eine andere Entscheidung

[1] David Marshall, „Erkennen Sie Ihre unbewussten Vorurteile“, *Business Matters*, www.bmmagazine.co.uk/in-business/recognising-unconscious-bias/, 22. Oktober 2013.

T. Bär, *Algorithmic Bias: Verzerrungen durch Algorithmen verstehen und verhindern*, https://doi.org/10.1007/978-3-662-66315-8_1

sprechen könnten, (absichtlich) außer Acht lässt. Wenn Ihr Partner beispielsweise einmal eine Bouillabaisse-Fischsuppe gegessen hat und ihm danach furchtbar übel wurde, wird er oder sie bestimmt nie wieder Bouillabaisse essen und sich vielleicht sogar weigern, die schöne Bouillabaisse zu probieren, die Sie gerade gekocht haben, wobei er oder sie geflissentlich die Tatsache ignoriert, dass Sie die Kochschule mit Auszeichnung absolviert und die besten und frischesten Zutaten gekauft haben, die im ganzen Land erhältlich sind.

Algorithmen sind mathematische Gleichungen oder andere logische Regeln zur Lösung eines bestimmten Problems, z. B. zur Entscheidung über eine binäre Frage (ja/nein) oder zur Schätzung einer unbekannten Zahl. Ähnlich wie das Gehirn in Sekundenbruchteilen Entscheidungen trifft, versprechen Algorithmen eine sofortige Antwort (in den meisten Fällen kann der Ergebniswert der Gleichung des Algorithmus in einem Bruchteil einer Sekunde berechnet werden), und sie sind auch eine Abkürzung, weil sie nur eine begrenzte Anzahl von Faktoren in einer vorgegebenen Weise berücksichtigen.

Auf einer Ebene sind Algorithmen eine Möglichkeit für Maschinen, menschliche Entscheidungsträger nachzuahmen oder zu ersetzen. So kann eine Bank, die jeden Monat Tausende von Kreditanträgen genehmigen muss, anstelle menschlicher Kreditsachbearbeiter einen von einem Computer angewandten Algorithmus einsetzen, um diese Kredite zu prüfen; der Grund dafür ist oft, dass ein Algorithmus sowohl schneller als auch billiger ist als ein Mensch.

Auf einer anderen Ebene können Algorithmen jedoch auch eine Möglichkeit sein, Verzerrungen zu verringern oder sogar zu beseitigen. Statistiker haben Techniken entwickelt, um Algorithmen speziell unter der Bedingung der Unvoreingenommenheit zu entwickeln. So ist beispielsweise die gewöhnliche Kleinstquadratregression (OLS) eine statistische Technik, die als BLUE definiert ist, die beste lineare unvoreingenommene Schätzung. Leider musste ich schreiben, dass Algorithmen Verzerrungen reduzieren oder beseitigen „*können*" – Algorithmen können genauso voreingenommen oder sogar schlechter sein als menschliche Entscheidungen. Mehrere Kapitel dieses Buches sind der Erläuterung der vielen Möglichkeiten gewidmet, wie ein Algorithmus voreingenommen sein kann.

Im Zusammenhang mit Algorithmen sollte die Definition von *Voreingenommenheit* jedoch spezifischer sein. Für Probleme, die von Algorithmen gelöst werden, gibt es zumindest theoretisch eine richtige Antwort. Wenn ich zum Beispiel die Anzahl der Haare auf dem Kopf eines bekannten Präsidenten schätze, hat sie vielleicht noch niemand gezählt, aber jeder, der unbegrenzt Zeit und Zugang zu dem Präsidenten hat, könnte meine Schätzung von 107.817 Haaren überprüfen.

In den meisten Situationen (auch bei Präsidentschaftswahlen) ist die richtige Antwort zumindest *a priori* (d. h. zum Zeitpunkt der Anwendung des Algorithmus) nicht bekannt. Algorithmen sind daher oft eine Möglichkeit, Vorhersagen

zu treffen. Durch Vorhersagen helfen Algorithmen, Unsicherheiten zu verringern und zu bewältigen. Wenn ich beispielsweise einen Kredit beantrage, weiß die Bank (noch) nicht, ob ich den Kredit zurückzahlen werde, aber wenn ein Algorithmus der Bank sagt, dass die Wahrscheinlichkeit, dass ich den Kredit nicht zurückzahle, 5 % beträgt, kann die Bank entscheiden, ob sie mit mir einen Gewinn macht, wenn sie mir den Kredit zu einem Zinssatz von 5,99 % gewährt, indem sie den erwarteten Verlust mit den Zinsen und anderen Kosten vergleicht, die der Bank entstehen. Dies ist ein typisches Beispiel für die Verwendung von Algorithmen: Algorithmen schätzen Wahrscheinlichkeiten für bestimmte Ereignisse (z. B. dass ein Kunde einen Kredit nicht zurückzahlt, ein Auto bei einem Unfall beschädigt wird oder eine Person erst am Ende der Laufzeit eines Lebensversicherungsvertrags stirbt), und diese Wahrscheinlichkeiten ermöglichen es einem Unternehmen, das Risiken übernimmt, auf der Grundlage eines objektiven Kriteriums für die erwartete risikoadjustierte Rendite eine Genehmigungs-/Ablehnungsentscheidung zu treffen.

Algorithmen werden in Situationen eingesetzt, in denen die Informationen unvollkommen sind (z. B. weiß der Algorithmus für die Kreditwürdigkeitsprüfung der Bank nichts über die Spielschulden, die ich gestern Abend gemacht habe, und er weiß auch nicht, ob mein Unternehmen mich nächsten Monat entlassen wird). Algorithmen *können* daher Fehler machen, sollten aber *im Durchschnitt* korrekt sein. Eine **Verzerrung** liegt vor, wenn der Durchschnitt aller Vorhersagen systematisch von der richtigen Antwort abweicht. Wenn der Algorithmus der Bank beispielsweise 10.000 verschiedenen Kunden eine Ausfallwahrscheinlichkeit von 5 % zuweist, würde man erwarten, dass 500 der 10.000 Kunden ausfallen werden (500/10.000 = 5 %). Wenn man die Situation untersucht und feststellt, dass in Wirklichkeit 10 % der Kunden säumig sind, der Algorithmus aber jedes Mal, wenn ein Antragsteller einen deutschen Pass hat, die wahre Schätzung um die Hälfte reduziert, ist der Algorithmus voreingenommen – in diesem Fall zugunsten der Deutschen. (Ist es ein Zufall, dass dieser Algorithmus von einem Deutschen entwickelt wurde?)

Systematische Fehler bei Vorhersagen – ob von Menschen oder Algorithmen gemacht – können schwerwiegende Folgen für Unternehmen haben, und leider kommen sie immer wieder vor. So wurden in einer Studie über Mega-Infrastrukturprojekte, in der 258 Projekte in 20 verschiedenen Ländern analysiert wurden, bei fast 9 von 10 Projekten Kostenüberschreitungen festgestellt, was auf eine systematische Unterschätzung der tatsächlichen Kosten hindeutet.[2] Während der globalen Finanzkrise gingen Banken wie Northern Rock, Lehman Brothers und Washington Mutual unter, weil sie Kredit-, Markt- und Liquiditätsrisiken systematisch unterschätzt hatten.

[2] B. Flyvbjerg, M.S. Holm, und S. Buhl, „Underestimating costs in public works projects: Error or lie?“, *Journal of the American Planning Association*, 68(3), 279–295, 2002.

Manchmal ist menschliche Voreingenommenheit daran schuld. Eine US-Bank verfügte beispielsweise über ein ökonomisches Kapitalmodell (ein ausgeklügeltes Modell zur Quantifizierung der „unerwarteten Verluste" eines bestimmten Portfolios, die zu einem Bank-Run oder Konkurs führen können), das vor der globalen Finanzkrise auf die übergroßen Risiken hinwies, die bei nachrangigen Hypotheken drohten, indem es unerwartete Verluste schätzte, die um ein Vielfaches höher waren als die erwarteten Verluste; tragischerweise glaubte die Geschäftsleitung diese Schätzungen nicht, weil sie unerwartete Verluste gewohnt war, die in ihrer Größenordnung viel näher an den erwarteten Verlusten lagen, und daher das Modell für fehlerhaft hielt.

In anderen Fällen sind jedoch die Algorithmen selbst fehlerhaft. So kaufte eine asiatische Bank ein Scoring-Modell für Verbraucherkreditkarten, das den Nutzungsgrad der Karte als einen Faktor zur Vorhersage eines Zahlungsausfall betrachtete. Die Algorithmen gingen davon aus, dass Kunden mit einer geringen Nutzung (z. B. nur 10 % des Kreditlimits) sicherer waren als Kunden mit einer hohen Nutzung; für sichere Kunden erhöhte der Algorithmus das Limit. Dies führte jedoch zu einem zirkulären Bezug: In dem Moment, in dem der Algorithmus das Kreditlimit erhöhte, sank die Auslastung (berechnet durch Division des aktuellen Kreditsaldos durch das Kreditlimit), was den Algorithmus dazu veranlasste, das Limit weiter zu erhöhen (bei einem Kreditsaldo von 10 und einem Limit von 100 lag die Auslastung also bei 10 %; erhöhte das System das Limit um 25 % von 100 auf 125, sank die Auslastung auf 8 % (= 10/125), was eine weitere Erhöhung des Limits auslöste, und so weiter). Dies geschah so lange, bis die Kreditlimits stratosphärische Höhen erreichten, die die Möglichkeiten der Kunden zur Rückzahlung an die Bank völlig überstiegen. Als immer mehr Kunden begannen, ihre sehr hohen Kreditlimits tatsächlich zu nutzen, fielen natürlich viele aus, und die Bank ging fast in Konkurs, nachdem sie mehr als eine Milliarde USD an uneinbringlichen Forderungen abgeschrieben hatte.

Algorithmische Verzerrungen gibt es in allen möglichen Formen und Farben. Im Jahr 2016 veröffentlichte ProPublica einen Forschungsbericht, aus dem hervorging, dass COMPAS, ein Algorithmus, der von US-Behörden verwendet wird, um die Wahrscheinlichkeit einer erneuten Straftat eines Kriminellen abzuschätzen, rassistisch gegen Schwarze eingestellt ist.[3] Das MIT berichtete, dass Algorithmen zur Verarbeitung natürlicher Sprache sexistisch sind, da sie Programmierer mit Männern und „home-maker", was im Englischen sowohl Hausmann als auch Hausfrau bedeuten kann, einseitig mit Frauen assoziieren.[4] Und Untersuchungen aus dem Jahr 2014 haben gezeigt, dass die Einstellung des Nutzerprofils auf „weiblich" in den

[3] J. Larson, S. Mattu, L. Kirchner und J. Angwin, „How we analyzed the COMPAS Rückfälligkeitsalgorithmus analysiert haben", *ProPublica*, 9, 2016.
[4] W. Knight, „How to Fix Silicon Valley's Sexist Algorithms", *MIT Technology Review*, November 23, 2016.

Anzeigeneinstellungen von Google dazu führen kann, dass weniger gut bezahlte Stellenangebote in den Anzeigen erscheinen.[5] Da immer mehr Entscheidungen von Algorithmen getroffen werden – mit Auswirkungen auf Verbraucher, Unternehmen, Mitarbeiter, Regierungen, die Umwelt und sogar auf Haustiere und unbelebte Gegenstände –, nehmen die Gefahren und Auswirkungen algorithmischer Voreingenommenheit von Tag zu Tag zu. Dies ist jedoch nicht zwangsläufig der Fall – Voreingenommenheit ist lediglich ein Nebeneffekt der Funktionsweise eines Algorithmus und somit ein Nebenprodukt bewusster und unbewusster Entscheidungen, die von den Entwicklern und Nutzern von Algorithmen getroffen werden. Diese Entscheidungen können überprüft und geändert werden, um algorithmische Verzerrungen zu verringern oder sogar zu beseitigen.

In diesem Buch geht es um algorithmische Verzerrungen. Zunächst einmal wollen wir besser verstehen, was sie sind, woher sie kommen und wie sie wichtige Entscheidungen beeinträchtigen können. Zweitens wollen wir den Schaden eindämmen, indem wir untersuchen, wie man mit algorithmischer Voreingenommenheit umgehen kann – sei es als Nutzer oder als Regulierer. Und drittens wollen wir untersuchen, wie sogenannte Datenwissenschaftler (Data Scientists), also die Entwickler von Algorithmen, algorithmische Voreingenommenheit verhindern können.

Der erste Teil, Kap. 2–5, führt in das Thema ein. Ich beginne mit einem kurzen Überblick über die Psychologie und die menschlichen Vorurteile bei Entscheidungen, da algorithmische Vorurteile diese mehr widerspiegeln, als man auf den ersten Blick sieht (Kap. 2), und erörtere, wie Algorithmen helfen können, solche Vorurteile bei Entscheidungen zu beseitigen (Kap. 3). Da viele Leser dieses Buches Laien und keine Datenwissenschaftler sind, werde ich anschließend erläutern, wie die Wurst gemacht wird, d. h. wie Algorithmen entwickelt werden (Kap. 4), und entmystifizieren, was hinter dem maschinellen Lernen („Machine Learning" auf Englisch) steckt (Kap. 5).

Im zweiten Teil des Buches, den Kap. 6–11, wird untersucht, woher algorithmische Verzerrungen kommen. In Kap. 6 wird untersucht, wie Vorurteile in der realen Welt durch Algorithmen widergespiegelt (anstatt korrigiert) werden können. Kap. 7 befasst sich mit der Person des Datenwissenschaftlers und der Frage, wie die eigenen (menschlichen) Voreingenommenheiten des Datenwissenschaftlers zu algorithmischen Verzerrungen führen können. Kap. 8 befasst sich eingehender mit der Rolle der Daten, und in Kap. 9 wird untersucht, wie die Natur der Algorithmen zur so genannten Stabilitäts-Verzerung (stability bias) führt. Kap. 10 befasst sich mit neuen Verzerrungen, die durch statistische Artefakte entstehen, und Kap. 11 taucht tief in die sozialen Medien ein, wo sich menschliches Verhalten und

[5] A. Datta, M.C. Tschantz, and A. Datta, „Automated experiments on ad privacy settings," *Proceedings on Privacy Enhancing Technologies*, 92–112, 2015.

algorithmische Verzerrungen auf besonders teuflische Weise gegenseitig verstärken können.

Der dritte Teil des Buches, die Kap. 12–17, befasst sich mit der algorithmischen Verzerrung aus der Sicht des Nutzers. Zunächst wird kurz erörtert, ob ein Algorithmus tatsächlich verwendet werden sollte oder nicht (Kap. 12) und wie die Schwere des Risikos der algorithmischen Verzerrung für ein bestimmtes Entscheidungsproblem zu bewerten ist (Kap. 13). Kap. 14 gibt einen Überblick über Techniken, mit denen man sich vor algorithmischer Verzerrung schützen kann. In Kap. 15 werden Techniken zur Diagnose von algorithmischer Voreingenommenheit beschrieben, und in Kap. 16 werden Managementstrategien zur Überwindung von Voreingenommenheit erörtert, die in einem Algorithmus (wenn auch nicht im wirklichen Leben) verankert ist. In Kap. 17 wird erörtert, wie Nutzer von Algorithmen einen entscheidenden Beitrag zur Entschärfung von algorithmischen Vorurteilen leisten können, indem sie unvoreingenommene Daten produzieren.

Der vierte Teil des Buches, Kap. 18–23, befasst sich mit Datenwissenschaftlern, die Algorithmen entwickeln. Kap. 18 gibt einen Überblick über die verschiedenen Möglichkeiten, wie sich Datenwissenschaftler vor algorithmischen Verzerrungen schützen können. Kap. 19 befasst sich eingehend mit spezifischen Techniken zur Identifizierung verzerrter Daten. In Kap. 20 wird erörtert, wie man bei der Entwicklung eines Algorithmus zwischen maschinellem Lernen und anderen statistischen Verfahren wählen kann, um algorithmische Verzerrungen zu minimieren, und Kap. 21 baut darauf auf, indem es hybride Ansätze vorschlägt, die das Beste aus beiden Welten kombinieren. In Kap. 22 wird erörtert, wie die in diesem Buch vorgestellten Debiasing-Techniken für selbstverbessernde maschinelle Lernmodelle angepasst werden können, die eine fortlaufende Validierung in Echtzeit erfordern. Und Kap. 23 nimmt die Perspektive einer großen Organisation ein, die zahlreiche Algorithmen entwickelt, und beschreibt, wie die besten Praktiken zur Vermeidung algorithmischer Verzerrungen in einen robusten Modellentwicklungs- und -einsatzprozess auf institutioneller Ebene eingebettet werden können.

KAPITEL

2

Voreingenommenheit in der menschlichen Entscheidungsfindung

Wie Sie in den folgenden Kapiteln sehen werden, haben algorithmische Verzerrungen ihren Ursprung in menschlichen kognitiven Verzerrungen oder spiegeln diese in vielerlei Hinsicht wider. Der beste Weg, algorithmische Voreingenommenheit zu verstehen, ist daher, menschliche Voreingenommenheit zu verstehen. Auch wenn „Voreingenommenheit" umgangssprachlich oft als etwas Schlechtes angesehen wird, das rücksichtsvolle, wohlmeinende Menschen meiden würden, so ist sie doch ein zentraler Bestandteil der Arbeit-

T. Bär, *Algorithmic Bias: Verzerrungen durch Algorithmen verstehen und verhindern*,
https://doi.org/10.1007/978-3-662-66315-8_2

sweise des menschlichen Gehirns. Der Grund dafür ist, dass die Natur drei konkurrierende Ziele gleichzeitig verfolgen muss: Genauigkeit, Geschwindigkeit und (Energie-)Effizienz.

Genauigkeit ist ein offensichtliches Ziel. Wenn Sie auf der Jagd nach Beute sind, aber ein schlecht funktionierendes kognitives System Sie dazu bringt, in jedem zweiten Baumstamm oder Felsen, auf den Sie stoßen, ein Tier zu sehen, wird es Ihnen offensichtlich schwer fallen, etwas Essbares zu erlegen.

Die Schnelligkeit hingegen wird oft übersehen. Das Überleben in der Wildnis ist oft eine Frage von Millisekunden. Wenn ein Tiger in Ihrem Blickfeld auftaucht, dauert es mindestens 200 Millisekunden, bis Ihr Frontallappen – der Ort des logischen Denkens im Gehirn – erkennt, dass Sie einen Tiger anstarren. Zu diesem Zeitpunkt kann es durchaus sein, dass sich der Tiger bereits auf Sie stürzt, und kurz darauf haben Sie Ihr Leben als Frühstück des Tigers beendet. Unser Überleben als Spezies hing wahrscheinlich davon ab, dass es der Natur gelungen ist, die Zeit, in der der Flucht-oder-Kampf-Reflex einsetzt, auf 30–40 Millisekunden zu verkürzen – nur 160 Millisekunden Unterschied zwischen dem Aussterben und dem Aufstieg zur Krone der Schöpfung (wie manche behaupten)! Wie John Coates in seinem Buch *The Hour Between Dog and Wolf (Die Stunde zwischen Hund und Wolf)* sehr detailliert beschreibt, musste die Natur ([1]) eine verblüffende Reihe von Verbesserungen und Tricks anwenden, um dies zu erreichen. Ein wichtiger Aspekt der Lösung: Im Zweifelsfall sollte man annehmen, dass man einen Tiger sieht. Wie Sie sehen werden, sind Vorurteile also ein wichtiger Bestandteil des Werkzeugkastens der Natur, um Entscheidungen zu beschleunigen.

Effizienz ist der am wenigsten bekannte Aspekt des Denkens und der Entscheidungsfindung in der Natur. Wahrscheinlich sind Sie in dem Glauben aufgewachsen, dass logisches, bewusstes Denken alles ist, was Ihr Gehirn leistet. Von wegen! Die meisten Denkvorgänge laufen in Wirklichkeit unbewusst ab. Selbst das, was sich wie bewusstes Denken anfühlt, ist oft ein Hin und Her zwischen bewusstem und unterbewusstem Denken. Stellen Sie sich zum Beispiel vor, Sie möchten heute Abend essen gehen. Für welches Restaurant würden Sie sich entscheiden? Bitte halten Sie hier inne und treffen Sie tatsächlich eine Entscheidung! Sind Sie bereit? Haben Sie Ihre Wahl getroffen? GUT. War es eine bewusste oder unbewusste Entscheidung? Wahrscheinlich sind Ihnen ein paar Optionen eingefallen und Sie haben dann bewusst eine Wahl getroffen. Aber wie ist diese kurze Liste von Optionen, die Sie in Betracht gezogen haben, zustande gekommen? Haben Sie eine Tabelle erstellt, um die Dutzenden oder Tausende von Restaurants in Ihrer Stadt akribisch durchzugehen, sie nach sorgfältig ausgewählten Kriterien zu bewerten und dann eine Entscheidung zu treffen? Oder ist Ihnen auf magische Weise eine recht kleine Auswahl an Restaurants eingefallen? Das ist ein

[1] John Coates, *The Hour Between Dog and Wolf*, New York: The Penguin Press, 2012.

Beispiel dafür, dass Ihr Unterbewusstsein Ihrem bewussten Denken auf die Sprünge hilft – es hat Ihnen die Entscheidung für ein Restaurant sehr erleichtert, indem es die Auswahl auf eine relativ kurze Liste reduziert hat.

Der Grund, warum die Natur so sehr auf Effizienz bedacht ist, liegt darin, dass Ihr logisches, bewusstes Denken erschreckend ineffizient ist. Das durchschnittliche Gehirn macht weniger als 2 % des Körpergewichts eines Menschen aus, verbraucht aber 20 % der Energie des Körpers.[2] Das bedeutet, dass 20 % der Nahrung, die Sie aufnehmen und verdauen, allein für die Versorgung Ihres Gehirns verwendet werden! Das ist eine Menge Energie für einen so kleinen Teil des Körpers. Und der größte Teil dieser Energie wird durch das logische Denken verbraucht (im Gegensatz zur fast mühelosen unterbewussten Mustererkennung). So wie moderne Flugzeuge und Schiffe über alle möglichen technischen Raffinessen verfügen, um den Energieverbrauch zu senken, hat auch Mutter Natur alle möglichen Mechanismen in das Gehirn eingebaut, um den Energieverbrauch durch logisches Denken zu minimieren (damit man nicht 20 Steaks pro Tag essen muss). Es überrascht nicht, dass sie dadurch auch allerhand Verzerrungen eingeführt hat.

Wenn man alle in der psychologischen Literatur beschriebenen Vorurteile sammelt, kommt man auf über 100 davon.[3] Viele von ihnen sind jedoch spezifische Anwendungen grundlegenderer Prinzipien der Funktionsweise des Gehirns, und daher haben mehrere Autoren die Literatur auf 4–5 Haupttypen von Vorurteilen reduziert. Mir persönlich gefällt die von Dan Lovallo und meinem ehemaligen Kollegen Olivier Sibony entwickelte Systematik:[4] Sie unterscheiden zwischen handlungsorientierten, stabilitätsbezogenen, mustererkennenden, interessensbezogenen und sozialen Effekten. Ich werde mich lose an diesen Rahmen halten, wenn ich im Folgenden einige der wichtigsten Vorurteile erörtere, die für das Verständnis der algorithmischen Verzerrung erforderlich sind.

Handlungsorientierte Voreingenommenheit

Handlungsorientierte Neigungen spiegeln die Einsicht der Natur wider, dass Schnelligkeit oft der König ist. Was glauben Sie, wer in der Wildnis eher überleben wird? Der umsichtige Planer, der eine 20-seitige Risikobewertung erstellt und mindestens fünf verschiedene Reaktionsmöglichkeiten durchdenkt, bevor er entscheidet, ob Kampf oder Flucht die bessere Reaktion auf den Tiger wäre, der gerade fünf Meter vor ihm aufgetaucht ist, oder der

[2] Daniel Drubach, *The Brain Explained*. New Jersey: Prentice-Hall, 2000.
[3] Buster Benson, „Cognitive Bias Cheat Sheet“, https://betterhumans.coach.me/cognitive-bias-cheat-sheet-55a472476b18, September 1, 2016.
[4] D. Lovallo und O. Sibony, „The case for behavioral strategy“, *McKinsey Quarterly*, 2(1), 30–43, 2010.

Draufgänger, der sich in einem Sekundenbruchteil entscheidet, gegen den Tiger zu kämpfen?

Eine Reihe von Vorurteilen veranschaulicht die Art der handlungsorientierten Vorurteile. Zunächst einmal lenken Vorurteile wie der von-Restorff-Effekt (Konzentration auf den einen Gegenstand, der sich von den anderen vor uns abhebt) und der Bizarrheitseffekt (Konzentration auf den Gegenstand, der sich am meisten von dem unterscheidet, was wir zu sehen erwarten haben) unsere Aufmerksamkeit auf den gelben Pelz unter all den Büschen und Bäumen um uns herum; übermäßiger Optimismus und übermäßiges Selbstvertrauen dämpfen dann die Selbstzweifel, die zu tödlichem Zögern führen könnten.

Der Bizarrheits-Effekt kann unsere Wahrnehmung verzerren, so wie Ausreißer und Leverage-Punkte einen übergroßen Effekt bei der Schätzung der Koeffizienten eines Algorithmus haben können. Der Grund dafür ist die Verfügbarkeitsverzerrung: Wenn wir uns an einen bestimmten Datenpunkt leichter erinnern als an andere Datenpunkte (z. B. weil er sich von den meisten anderen Datenpunkten abhebt), überschätzen wir die Repräsentativität des besonders einprägsamen Datenpunkts. Dies kann erklären, warum z. B. ein einziger Vorfall, bei dem ein Ausländer ein spektakuläres Verbrechen begeht, unsere Wahrnehmung von Menschen mit der Nationalität dieses Ausländers stark verzerren kann, was zu unverhältnismäßiger Feindseligkeit und Aggression gegen sie führt.

Übermäßiges Selbstvertrauen verdient unsere besondere Aufmerksamkeit, weil es auch erklärt, warum nicht genug gegen Vorurteile im Allgemeinen und algorithmische Verzerrungen im Besonderen getan wird. Viele Forscher haben Selbstüberschätzung nachgewiesen, indem sie Menschen gefragt haben, wie sie sich mit anderen vergleichen.[5] So glaubten beispielsweise 70 % der befragten Gymnasiasten, dass sie „überdurchschnittliche" Führungsqualitäten haben, aber nur 2 %, dass sie „unterdurchschnittlich" sind (wobei per Definition jeweils etwa 50 % unter- bzw. überdurchschnittlich sein müssten). Bei der Fähigkeit, mit anderen auszukommen, glaubten sogar 60 %, zu den besten 10 % und 25 %, zu den besten 1 % zu gehören. Ähnliche Ergebnisse wurden für technische Fähigkeiten wie Autofahren und Softwareprogrammierung gefunden. Überoptimismus ist im Wesentlichen die gleiche Voreingenommenheit, wird aber auf die Bewertung von Ergebnissen und Ereignissen angewandt, z. B. darauf, ob ein großes Bauprojekt sein Kostenbudget einhalten kann.

Was bedeutet das für die Bekämpfung von Voreingenommenheit? Selbst wenn Menschen die Tatsache akzeptieren, dass andere voreingenommen sein könnten, überschätzen sie ihre eigene Fähigkeit, bei Entscheidungen ihren Vorurteilen zu widerstehen – und widersetzen sich daher Bemühungen

[5] Die hier aufgeführten Beispiele stammen aus D. Dunning, C. Heath und J.M. Suls, „Flawed self-assessment: Implications for health, education, and the workplace", *Psychological science in the public interest*, 5(3), 69–106, 2010.

anderer, ihre eigenen Vorurteile zu entkräften. Da die meisten Menschen zu optimistisch sind, kann es leicht zu einer Situation kommen, in der die meisten Menschen akzeptieren, dass es Vorurteile gibt, sich aber trotzdem weigern, etwas dagegen zu unternehmen.

Ein weiterer faszinierender Aspekt der Forschung zum Überoptimismus: Dieser wurde nur in der westlichen Kultur gefunden, nicht aber im Fernen Osten.[6] Dies zeigt, dass sowohl die Persönlichkeit des Einzelnen als auch die allgemeine Kultur eines Landes (oder eines Unternehmens/einer Organisation) einen Einfluss auf die Art und Weise haben, wie wir Entscheidungen treffen, und somit auch auf Vorurteile. Eine Voreingenommenheit, die wir in einem bestimmten Kontext beobachten, tritt in einem anderen möglicherweise nicht auf – stattdessen können andere Voreingenommenheiten auftreten.

Hinweis Ein hervorragendes Exempel für Selbstüberschätzung ist die Tatsache, dass ich schreibe, dass die meisten Menschen aufgrund ihrer Selbstüberschätzung nichts gegen ihre Vorurteile unternehmen – aber ich trotzdem ein Buch verfasse, wie man Verzerrungen von Algorithmen bekämpft, weil ich irgendwie dennoch glaube, dass ich trotz aller Widrigkeiten in der Lage sein werde, die menschliche Voreingenommenheit meiner Leser zu überwinden und sie dafür zu gewinnen, meine Vorschläge umzusetzen. Ich weiß aber auch, dass *Sie*, liebe Leserin, lieber Leser, anders sind als der Durchschnittsleser und viel eher dazu neigen, tatsächlich Maßnahmen zu ergreifen als andere. Lassen Sie mich daher nur darauf hinweisen, dass Sie, um Ihrem wohlverdienten positiven Selbstbild gerecht zu werden, noch heute einen Aktionsplan aufstellen sollten, wie Sie die Erkenntnisse und Empfehlungen aus diesem Buch in Ihrer täglichen Arbeit anwenden werden, und damit aktiv dem verlockenden Glauben widerstehen, dass Sie gegen Voreingenommenheit immun sind, so dass Sie nicht die hohen Erwartungen von uns beiden in unsere jeweiligen Fähigkeiten enttäuschen. ☺

Stabilitätsverzerrungen

Stabilitätsvorurteile sind eine Möglichkeit für die Natur, effizient zu sein. Stellen Sie sich vor, Sie wären der einzige Besucher einer Matinee-Vorstellung eines Kunstfilms – Sie könnten sich also buchstäblich jeden der 200 Plätze aussuchen. Was würden Sie tun: alle 30 Sekunden aufspringen, um einen anderen

---

[6] Eine allgemeine Einschränkung der Sozialpsychologie besteht darin, dass die meisten empirischen Untersuchungen im Kontext der westlichen Kultur durchgeführt werden, wobei ein erheblicher Teil der Untersuchungen sogar noch enger mit nordamerikanischen College-Studenten durchgeführt wird. Die wenigen Studien, in denen westliche Theorien in asiatischen Kulturen wie Japan oder China getestet werden, stellen regelmäßig erhebliche kulturelle Unterschiede fest.

auszuprobieren, oder sich auf einem Sitz niederlassen und ihn höchstens ein- oder zweimal wechseln, um vielleicht mehr Beinfreiheit zu gewinnen oder der kalten Brise einer unangenehmen Klimaanlage zu entkommen? Aus Sicht der Natur haben Sie jedes Mal, wenn Sie nur daran denken, den Sitzplatz zu wechseln, bereits geistigen Treibstoff verbraucht, und wenn Sie tatsächlich aufstehen, um den Sitzplatz zu wechseln, verbrauchen Ihre Muskeln teure Energie, ganz zu schweigen davon, dass Sie vielleicht die beste Szene des Films verpassen. Eine Reihe von Vorurteilen versucht, die Verschwendung geistiger und körperlicher Ressourcen zu verhindern, indem sie Sie am Status quo „festhalten".

Beispiele für diese Voreingenommenheit sind die Status-quo-Voreingenommenheit und die Verlustaversion. Man mag den Platz, auf dem man sitzt, lieber als andere Plätze, einfach weil es der Status quo ist – und man hasst die Vorstellung, ihn zu verlieren. In Experimenten mit Kaffeebechern und Kugelschreibern von Universitäten wurde gezeigt, dass der Mindestpreis, zu dem Sie bereit sind, einen Gegenstand zu verkaufen, sobald er sich in Ihrem Besitz befindet (d. h., Sie mit dem Gegenstand „ausgestattet" sind), etwa *doppelt so hoch* ist wie der Höchstpreis, den Sie für den Gegenstand bei Neuanschaffung zu zahlen bereit wären.[7]

Während Wirtschaftswissenschaftler eine solche Situation als irrational und abnormal betrachten, erscheint sie aus der Sicht der Natur vollkommen vernünftig – die Natur möchte, dass Sie sich entweder ausruhen oder produktivere Dinge tun, als mit unbedeutenden Gegenständen zu handeln, die nur einen geringen persönlichen Gewinn bringen! Manchmal schießt diese Status-quo-Einstellung jedoch über das Ziel hinaus. Zum Beispiel weisen Unternehmensentscheidungen bei der jährlichen Budgetierung eine sehr starke Tendenz zum Status quo auf, wobei eine Analyse eine 90-prozentige Korrelation bei den Budgetzuweisungen Jahr für Jahr (für einzelne Abteilungen oder Referate) ergab. Dadurch konnte zwar eine erbitterte Debatte über die Streichung von Budgets aus einigen Abteilungen vermieden werden, aber diese Stabilität hat enorme wirtschaftliche Kosten zur Folge: Unternehmen mit einer dynamischeren Budgetzuweisung wachsen doppelt so schnell wie solche, die sich dem Status quo beugen.[8]

Eine weitere wichtige Stabilitätsverzerrung ist der Verankerungseffekt. Ökonometriker, die sich mit Zeitreihenmodellen befassen, sind oft überrascht, wie gut das so genannte naive Modell funktioniert[9] – für viele Zeitreihen ist der Wert der aktuellen Periode ein hervorragender Prädiktor für den Wert der nächsten Periode, und viele komplexe Zeitreihenmodelle übertreffen

[7] D. Kahneman, J.L. Knetsch, und R.H. Thaler, „Anomalies: The endowment effect, loss aversion, and status quo bias," *Journal of Economic Perspectives*, 5(1), 193–206, 1991.
[8] T. Bär, S. Heiligtag, und H. Samandari, *The business logic in debiasing*, McKinsey & Co, 2017.
[9] https://blogs.sas.com/content/forecasting/2014/04/30/a-naive-forecast-is-not-necessarily-bad/.

dieses naive Modell kaum. Die Natur muss dies bemerkt haben, denn wenn Menschen eine Schätzung vornehmen, stützen sie sich oft stark auf den ihnen vorliegenden Anfangswert und nehmen nur geringfügige Anpassungen vor, wenn sich im Laufe der Zeit neue Informationen ergeben. Manchmal führt diese Voreingenommenheit jedoch ernsthaft in die Irre – nämlich dann, wenn der Ausgangswert komplett falsch oder einfach zufällig ist. Eine beliebte Demonstration des Verankerungseffekts besteht darin, die Teilnehmer aufzufordern, die letzten beiden Ziffern ihrer Sozialversicherungs- oder Telefonnummer aufzuschreiben, bevor sie den Preis eines Gegenstands schätzen, z. B. einer Flasche Wein oder einer Schachtel Pralinen. Obwohl es offensichtlich keinerlei Zusammenhang zwischen diesen Zahlen und dem Preis des Artikels gibt, schätzen diejenigen, die hohe Zahlen aufschreiben, die Preise durchweg 60 bis 120 Prozent höher ein als diejenigen mit niedrigen Zahlen.[10]

Verzerrungen bei der Erkennung von Mustern

Die Verzerrungen bei der Mustererkennung haben mit einem sehr lästigen Problem unserer Wahrnehmung zu tun: Viele unserer Sinneswahrnehmungen sind unvollständig, und es gibt eine Menge Rauschen in dem, was wir wahrnehmen. Stellen Sie sich vor, als Sie das letzte Mal mit jemandem gesprochen hatten – wahrscheinlich ist das erst ein paar Minuten her, vielleicht haben Sie mit dem Zugführer oder dem Flugbegleiter gesprochen, wenn Sie dieses Buch unterwegs lesen. Denken Sie an einen gehaltvollen, informationsreichen Satz, den Ihr Gesprächspartner mitten im Gespräch gesagt hat. Es ist gut möglich, dass ein Teil des Satzes durch ein lautes Geräusch (z. B. das Niesen einer anderen Person) völlig übertönt wurde, dass mehrere Silben gemurmelt wurden, oder dass Sie einen Teil des Satzes verpasst haben, weil Sie auf Ihr Telefon geschaut haben. Haben Sie die Person gebeten, den Satz zu wiederholen? Oder haben Sie irgendwie trotzdem eine gute Vorstellung davon, was die Person gesagt hat? Sehr oft ist Letzteres der Fall – dank der erstaunlichen Fähigkeit unseres Gehirns, „Lücken zu füllen". Unsere Gehirne sind sehr gut im Raten – aber manchmal sind diese Vermutungen systematisch falsch, und das ist der Bereich der Mustererkennungsfehler.

Verzerrungen bei der Mustererkennung sind für dieses Buch besonders relevant, da die Mustererkennung im Wesentlichen die Aufgabe von Algorithmen ist.

Um das Problem zu lösen, aus verrauschten, unvollständigen Daten (seien es visuelle oder andere sinnliche Wahrnehmungen oder tatsächliche Daten wie ein Bericht eines Management-Informationssystems voller klein gedruckter Tabellen) sinnvolle Schlüsse zu ziehen, muss das Gehirn Regeln entwickeln.

[10] E. Teach, „Avoiding Decision Traps", *CFO*, 1. Juni 2004; abgerufen am 29. Oktober 2018.

Systematische Fehler (d. h. Verzerrungen) treten auf, wenn entweder die Regeln falsch sind oder eine Regel falsch angewendet wird.

Der „Texas Sharpshooter"-Trugschluss ist ein Beispiel für eine fehlerhafte Regel. Ihr Gehirn sieht Regeln (d. h. Muster) in den Daten, wo keine vorhanden sind. Dies könnte viele Aberglauben erklären. Wenn eine Verkäuferin dreimal hintereinander ein Geschäft abschließt, während sie ein rotes Halstuch trägt, das sie von ihrem Mann zum Geburtstag bekommen hat, könnte das Gehirn zu dem Schluss kommen, dass es sich um ein „Glückshalstuch" handelt. Interessanterweise hat das Gehirn vielleicht gar nicht so Unrecht – es ist möglich, dass die Farbe Rot eine psychologische Wirkung auf Käufer hat, die die Wahrscheinlichkeit eines Geschäftsabschlusses erhöht –, aber drei abgeschlossene Geschäfte sind eine statistisch unbedeutende Stichprobe und viel zu wenig Daten, um daraus eine zuverlässige Schlussfolgerung zu ziehen. Dies veranschaulicht, dass die Natur bei der Mustererkennung stark von einer „lieber auf Nummer sicher gehen"-Mentalität geprägt ist – wie oft muss der Nachbarshund Sie beißen, damit Sie zu dem Schluss kommen, dass Sie diesem süßen Hündchen besser nicht zu nahe kommen? Umgekehrt ist das Gehirn so verdrahtet, dass es denkt: Selbst wenn die Wahrscheinlichkeit gering ist, dass das rote Halstuch hilft, warum sollte man ein großes Risiko eingehen, wenn man es nicht trägt?

Confirmation Bias kann ein Komplize des Texas Sharpshooter Trugschlusses sein und ist die Art und Weise, wie die Natur bei der Erkennung von Mustern effizient ist. Der Bestätigungsfehler kann als ein „hypothesengesteuerter" Ansatz zum Sammeln von Datenpunkten betrachtet werden. Das bedeutet, dass Sie, wenn Sie eine Hypothese haben (z. B. dass Sie bereits davon überzeugt sind, dass es eine gute Idee war, dieses Buch zu kaufen), dazu neigen, neue Daten auszuwählen, die Ihre Überzeugung bestätigen (z. B. loben Sie die Fünf-Sterne-Rezension dieses Buches für ihre brillanten Einsichten), und widersprüchliche Daten zurückweisen (z. B. bezeichnen Sie den Autor der Ein-Stern-Rezension als Idioten – natürlich zu Recht, wie ich anmerken möchte!). Hinter dem Confirmation Bias scheint der Wunsch der Natur zu stehen, schnell zu einer Entscheidung zu kommen und den kognitiven Aufwand zu verringern. Laborexperimente haben gezeigt, dass die Teilnehmer mit größerer Wahrscheinlichkeit Nachrichtenartikel lesen, die ihre Ansichten unterstützen, als solche, die ihnen widersprechen. Sie werden daher in Kap. 11 über algorithmische Verzerrungen in sozialen Medien auf den Confirmation Bias als zentralen Feind stoßen.

Bestätigungsvorurteile können auch die Art und Weise beeinflussen, wie wir „verrauschte" Informationen verarbeiten. Stellen Sie sich die oben erwähnte Interaktion mit einer Flugbegleiterin oder einem Zugbegleiter vor. Sie fragt Sie nach dem Buch, das Sie gerade lesen, und Sie zeigen ihr stolz das Cover dieses Buches. Gerade als sie antwortete, übertönte ein lautes Geräusch einen Teil ihres Satzes. Es gibt keine Möglichkeit festzustellen, ob sie gesagt hat: „Ich

habe das Buch geliebt!“ oder „Ich habe das Buch verabscheut!“ Außer, dass Sie wahrscheinlich „gehört“ haben, dass sie das Buch geliebt hat. Das liegt daran, dass Ihr Gehirn natürlich erwartet hätte, dass sie das sagt, und ein nicht eindeutiger Laut würde automatisch und unbewusst durch den erwarteten Inhalt ersetzt werden.

Die Stereotypisierung ist eine Erweiterung des Bestätigungsfehlers und ein Beispiel für eine Voreingenommenheit, bei der eine Regel übermäßig starr angewendet wird. Stellen Sie sich zunächst vor, dass Sie in einem schicken Restaurant sitzen. Der Kellner hat gerade die Rechnung an den Tisch neben Ihnen gebracht, als ein stattlicher, älterer, weißer Mann einen schwarzen Gegenstand aus seiner Hose zieht. Was denken Sie, was es ist? Sie haben wahrscheinlich an eine Brieftasche gedacht. Nun stellen Sie sich vor, dass ein Polizeiauto an einer sichtlich verstörten Frau vorbeifährt, die am Straßenrand liegt. Als das Polizeiauto vorbeifährt, ruft die Frau: „Meine Geldbörse, meine Geldbörse!“ und winkt in die Luft. In diesem Moment werden die Polizeibeamten auf einen jungen schwarzen Mann aufmerksam, der in der Nähe in Richtung einer U-Bahn-Station läuft. Sie rennen sofort hinter dem Mann her, rufen „Stopp! Polizei!“ und zielen mit ihren Gewehren auf den Mann. Als der Mann die Stufen zum Eingang der U-Bahn-Station erreicht, zieht er einen schwarzen Gegenstand aus seiner Tasche. Was ist das? Wenn Sie an eine Pistole gedacht haben (und nicht an die Brieftasche mit dem U-Bahn-Pass, den der Mann schnell hervorholen muss, wenn er seinen Zug nicht verpassen und somit zu spät zu seiner Klavierstunde kommen will), dann sind Sie Opfer einer Stereotypisierung geworden. Aufgrund des Kontextes der Situation hat Ihr Gehirn bereits einige Erwartungen, was als Nächstes passieren könnte. Eine Person in einem Restaurant, die gerade eine Rechnung erhalten hat, wird wahrscheinlich eine Brieftasche, eine Kreditkarte oder ein Bündel Geldscheine aus der Tasche ziehen; eine Person, die offenbar einen Raubüberfall begangen hat, wird wahrscheinlich ein Messer, eine Pistole oder eine Handgranate aus der Tasche ziehen, wenn sie versucht, vor der Polizei zu fliehen. Wenn das Gehirn nur weiß, dass ein „schwarzer Gegenstand“ aus der Tasche gezogen wird, „füllt es die Lücken“ auf der Grundlage dieser stereotypen Ansichten darüber, was eine Person in einem solchen Kontext am wahrscheinlichsten in ihrer Tasche hat. Das Dilemma ist, dass diese Vermutung falsch sein könnte. Es liegt auf der Hand, dass ein Polizeibeamter, der einen Verdächtigen in dem Moment erschießt, in dem ein schwarzer Gegenstand aus der Tasche gezogen wird, mit geringerer Wahrscheinlichkeit erschossen wird und somit eher überlebt als ein vorsichtigerer und bedächtigerer Beamter, der erst dann abdrückt, wenn der Verdächtige ohne jeden Zweifel eine Waffe auf ihn gerichtet hat, so dass die Evolution nicht gerade der größte Fan eines ordnungsgemäßen Ermittlungsverfahrens war. Wenn der Beamte jedoch einen Unschuldigen erschießt, weil dieser in den Augen des Beamten wie ein „stereotypischer“ Räuber aussieht, haben die List und die Voreingenommenheit der Natur auf tragische Weise ein Leben gefordert.