VDSG
VIENNA DATA SCIENCE GROUP

Katherine **MUNRO**
Stefan **PAPP**
Zoltan **TOTH**
Wolfgang **WEIDINGER**
Danko **NIKOLIĆ**

3. Auflage

HANDBUCH DATA SCIENCE UND KI

Mit **Machine Learning** und **Datenanalyse** Wert aus Daten generieren

HANSER

Munro / Papp / Toth / Weidinger / Nikolić
Antosova Vesela / Bruckmüller / Cadonna / Eder / Gorzala / Hahn / Langs /
Licandro / Mata / McIntyre / Meir-Huber / Móra / Pasieska / Rugli / Wazir / Zauner

Handbuch Data Science und KI

Katherine Munro, Stefan Papp, Zoltan C. Toth,
Wolfgang Weidinger, Danko Nikolić,
Barbora Antosova Vesela, Karin Bruckmüller,
Annalisa Cadonna, Jana Eder, Jeannette Gorzala,
Gerald A. Hahn, Georg Langs, Roxane Licandro,
Christian Mata, Sean McIntyre, Mario Meir-Huber,
György Móra, Manuel Pasieska, Victoria Rugli,
Rania Wazir, Günther Zauner

Handbuch Data Science und KI

Mit Machine Learning und Datenanalyse Wert aus Daten generieren

3., aktualisierte und erweiterte Auflage

HANSER

Print-ISBN: 978-3-446-47937-1
E-Book-ISBN: 978-3-446-48072-8
Epub-ISBN: 978-3-446-48357-6

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter http://dnb.d-nb.de abrufbar.

Kolbergerstraße 22 | 81679 München | info@hanser.de
www.hanser-fachbuch.de
Lektorat: Sylvia Hasselbach
Copy editing: Jürgen Dubau, Freiburg/Elbe
Coverkonzept: Marc Müller-Bremer, www.rebranding.de, München
Covergestaltung: Tom West
Titelmotiv: gettyimages/ValeryBrozhinsky
Herstellung: le-tex publishing services GmbH, Leipzig
Satz: Eberl & Koesel Studio, Kempten
Druck: CPI books GmbH, Leck
Printed in Germany

Inhalt

Vorwort

Dieses Vorwort wurde NICHT von ChatGPT (oder Ähnlichem) geschrieben.

Während ich diese Aussage treffe, frage ich mich, wie oft sie in Zukunft für Texte oder andere Medienformen gelten wird. In den letzten zwei Jahren hat dieses KI-gestützte Tool enorme Popularität erlangt und Data Science und KI einen unglaublichen Bekanntheitsgrad verschafft. Infolgedessen sind die Erwartungen an Künstliche Intelligenz exponentiell gestiegen und haben solche Höhen erreicht, dass man sich fragen könnte, ob sie jemals erreicht werden können.

Das Thema KI folgt dem bekannten Hype-Zyklus. Einige dieser hohen Erwartungen sind wohlverdient: Diese leistungsstarke Technologie wird die Art und Weise, wie wir leben und arbeiten, in vielerlei Hinsicht verändern. Um ein Beispiel zu nennen: Einige Universitäten erwägen, von ihren Studenten keine Seminararbeiten mehr zu verlangen, da es nicht möglich ist zu überprüfen, ob sie von einem KI-Tool geschrieben wurden.

Aber wir müssen uns auch auf einige Enttäuschungen in der Zukunft gefasst machen, da die KI unweigerlich die überzogenen Erwartungen mancher Leute nicht erfüllen kann. Selbst wenn die Vorstellungen vernünftig sind, ist der Zeitrahmen, den diese Menschen und Organisationen für die Umsetzung von KI-Projekten im Sinn haben, oft nicht realistisch. Dies führt zu weiteren Enttäuschungen, wenn die erhoffte Wirkung und der erhoffte Wert nicht innerhalb des gewünschten Zeitrahmens erreicht werden können.

Die ersten Anzeichen dafür sind bereits zu erkennen, denn ChatGPT und ähnliche Tools liefern eine Fülle von wortgewandten und kohärenten – aber nicht korrekten – Informationen. Die neue Welle von „KI-Experten“, die immer haarsträubendere Versprechungen über die von ihnen oder ihren Unternehmen erfundenen Tools machen, die nur schwer zu halten sein werden, trägt nicht zu mehr Vertrauen bei. Im Grunde genommen verkaufen sie digitales „Schlangenöl“.

All dies erhöht den Druck auf die Data Scientists, mit diesen Erwartungen umzugehen und gleichzeitig weiterhin das gleiche Ziel zu erreichen, das sie seit Jahrzehnten verfolgen:

verständliche Antworten auf Fragen anhand von Daten zu geben.

Aus diesem Grund sind neutrale Organisationen wie die Vienna Data Science Group (VDSG, *www.vdsg.at*), die den interdisziplinären und internationalen Wissensaustausch zwischen Datenexperten fördert, so notwendig und wichtig. Wir engagieren uns nach wie vor stark für die Entwicklung des gesamten Data-Science- und KI-Ökosystems (Ausbildung, Zertifizierung, Standardisierung, Studien zu den gesellschaftlichen Auswirkungen usw.) in Europa und darüber hinaus. Dieses Buch ist nur eine unserer Bemühungen, um dieses Ziel zu erreichen. Denn trotz all des Hypes und der Übertreibungen in der KI- und Datenlandschaft bleibt Data Science dasselbe: eine interdisziplinäre Wissenschaft, die eine sehr heterogene Gruppe von Spezialisten versammelt. Sie setzt sich aus drei großen Strömungen zusammen, und wir sind stolz darauf, dass wir in jeder von ihnen Experten haben:

- Informatik und IT
- Mathematik und Statistik
- Fachwissen in der Branche oder dem Bereich, in dem Data Science und künstliche Intelligenz angewendet werden.

Die VDSG (*www.vdsg.at*) hat schon immer einen ganzheitlichen Ansatz für Data Science verfolgt, und das ist auch in diesem Buch nicht anders: Ab Kapitel 1 stellen wir ein fiktives Unternehmen vor, das datengetriebener werden möchte, und begleiten es im Laufe des Buches bis zum Ende seiner Datentransformation in Kapitel 28. Auf dem Weg dorthin gehen wir auf viele Herausforderungen ein und bieten Ihnen so praktische Einblicke, die nur dank des regen Austauschs in unserer großen Data-Science- und KI-Community möglich waren.

Das Ergebnis ist eine stark erweiterte Ausgabe unseres Data Science & AI-Handbuchs mit zehn neuen Kapiteln zu Themen wie Aufbau von KI-Lösungen (Kapitel 13), Foundation Models (Kapitel 15), Large Language Models und generative KI (Kapitel 16) sowie Klimawandel und KI (Kapitel 25). Ergänzend dazu werden auch die grundlegenden Themen Datenarchitektur, Engineering und Governance (Kapitel 4, 5 und 6) behandelt und mit Machine Learning Operations (MLOps, Kapitel 7) abgerundet, das sich zu einer eigenen, sehr wichtigen Disziplin entwickelt hat.

Um eine solide Grundlage zu schaffen, die Ihnen hilft, all dies zu verstehen, haben wir wieder eine Einführung in die zugrunde liegende Mathematik (Kapitel 9) und Statistik (Kapitel 10), die in Data Science verwendet werden, sowie Kapitel über die Theorie hinter Machine Learning, der Signalverarbeitung und der Computer Vision (Kapitel 12, 14 und 18) aufgenommen. Wir haben auch Themen behandelt, die mit der Wertschöpfung aus Daten zu tun haben, wie z. B. Business Intelligence (Kapitel 11) und Data Driven Enterprises (Kapitel 21), sowie wichtige Informationen, die Ihnen helfen,

Daten sicher zu nutzen, einschließlich Kapiteln über das neue EU-KI-Gesetz (Kapitel 23) und vertrauenswürdige KI (Kapitel 27).

Diese umfangreiche Erweiterung des Opus Magnum der VDSG dient vor allem einem Zweck:

ein realistisches und ganzheitliches Bild von Data Science und KI zu vermitteln.

Data Science und KI entwickeln sich derzeit in einem unglaublich schnellen Tempo, und das gilt auch für ihre Auswirkungen auf die Gesellschaft. Das bedeutet, dass die Verantwortung, die auf den Schultern der Data Scientists lastet, ebenfalls gewachsen ist, und damit auch die Notwendigkeit für Organisationen wie die VDSG, sich zu engagieren und diese Herausforderungen zu bewältigen.

Packen wir's an!

Sommer 2024

Wolfgang Weidinger

Danksagungen

Wir, die Autoren, möchten diese Gelegenheit nutzen, um unseren Familien und Freunden, die uns geholfen haben, unsere Gedanken und Einsichten in diesem Buch auszudrücken, unseren aufrichtigen Dank auszusprechen. Ohne ihre Unterstützung und Geduld wäre diese Arbeit nicht möglich gewesen.

Ein besonderer Dank aller Autoren geht an Katherine Munro, die viel zu diesem Buch beigetragen und viel Zeit und Mühe in die Bearbeitung unserer Manuskripte investiert hat.

Für meine Eltern, die immer gesagt haben, dass ich alles schaffen kann. Wir hätten nie erwartet, dass es so etwas sein würde.

Katherine Munro

Ich möchte mich bei meiner Frau und der Vienna Data Science Group für ihre kontinuierliche Unterstützung auf meinem beruflichen Weg bedanken.

Zoltan C. Toth

Wenn ich an die Menschen denke, die mich am meisten unterstützt haben, möchte ich mich bei meinen Eltern bedanken, die immer an mich geglaubt haben, egal, was passiert ist, und bei meiner Partnerin Verena, die in den letzten Monaten wieder sehr geduldig war, während ich an diesem Buch gearbeitet habe.

Darüber hinaus bin ich sehr dankbar für die Unterstützung und Motivation, die ich von den Menschen erhalten habe, die ich durch die Vienna Data Science Group kennengelernt habe.

Wolfgang Weidinger

1 Einführung

Stefan Papp

„Ich möchte CDO werden anstelle des CDOs.“

Iznogoud (angepasst)

Fragen, die in diesem Kapitel beantwortet werden:

- Wie könnte ein fiktives Unternehmen aussehen, das vor seiner Transformation in ein datengesteuertes Unternehmen steht?
- Welche Herausforderungen muss ein Unternehmen bewältigen, um datengesteuert zu werden?
- Wie werden die Kapitel in diesem Buch Ihnen, dem Leser, helfen, solche Herausforderungen in Ihrer eigenen Organisation zu erkennen und zu bewältigen?

1.1 Über dieses Buch

Dieses Buch bietet einen praktischen, erfahrungsbasierten Einblick in verschiedene Aspekte von Data Science und künstlicher Intelligenz. In dieser, unserer dritten, Auflage tauchen die Autoren auch tief in einige der aufregendsten und sich schnell entwickelnden Themen unserer Zeit ein, darunter Large Language Models und generative KI.

Das vorrangige Ziel der Autoren ist es, dem Leser einen ganzheitlichen Zugang zu diesem Bereich zu vermitteln. Aus diesem Grund ist dieses Buch nicht rein technisch: Die Reife von Data Science und KI hängt ebenso sehr von der Arbeitskultur ab, insbesondere vom kritischen Denken und evidenzbasierter Entscheidungsfindung, wie von den Kenntnissen in Mathematik, neuronalen Netzen, KI-Frameworks und Datenplattformen.

In den letzten Jahren sind sich die meisten Experten einig geworden, dass künstliche Intelligenz unsere Arbeits- und Lebensweise verändern wird. Für eine ganzheitliche Betrachtung müssen wir auch den Status quo betrachten, wenn wir verstehen wollen, was getan werden muss, um unsere vielfältigen Ambitionen mithilfe von KI zu erfüllen. Ein nützlicher Rahmen dafür ist es zu untersuchen, wie Menschen mit den Herausforderungen der Datentransformation aus einer organisatorischen Perspektive umgehen. Aus diesem Grund werden wir dem Leser ein fiktives Unternehmen vorstellen, das am Anfang seiner Reise steht, evidenzbasierte Entscheidungsfindung in seine Unternehmensidentität zu integrieren. Wir werden dieses fiktive Unternehmen, in dem vieles datenorientierter sein könnte, als Modell verwenden, um mögliche Herausforderungen zu skizzieren, denen Organisationen begegnen können, wenn sie datenorientierter werden wollen. Am Ende dieses Buches wird unser hypothetisches Unternehmen auch als Modell dafür dienen, wie ein datengesteuertes Unternehmen aussehen könnte. In den dazwischen liegenden Kapiteln gehen wir auf viele dieser Herausforderungen ein und geben praktische Ratschläge, wie man sie bewältigen kann.

Falls Sie als Leser lieber keine Prosa über ein fiktives Unternehmen lesen möchten, um etwas über solche typischen organisatorischen Herausforderungen zu erfahren, empfehlen wir Ihnen, dieses Kapitel zu überspringen und mit einem Kapitel zu beginnen, das Ihren Interessen entspricht. Als ganzheitliches Buch über dieses Gebiet behandeln die Autoren künstliche Intelligenz, maschinelles Lernen, generative KI, Modellierung, Verarbeitung natürlicher Sprache, Computer Vision und andere relevante Bereiche. Wir behandeln Engineer-Themen wie Datenarchitektur und Datenpipelines, die für die Umsetzung datengetriebener Projekte in die Produktion unerlässlich sind. Schließlich gehen wir auch auf kritische soziale und rechtliche Fragen im Zusammenhang mit der Nutzung von Daten ein. Jeder Autor geht sehr detailliert auf sein Fachgebiet ein, damit Sie einen großen Nutzen daraus ziehen können.

Wir bitten die Leser, sich direkt mit uns in Verbindung zu setzen und uns mitzuteilen, wie wir unser ehrgeiziges Ziel erreichen können, die Standardliteratur für einen ganzheitlichen Ansatz in diesem Bereich zu werden. Wenn Sie der Meinung sind, dass einige neue Inhalte in einer der nächsten Ausgaben behandelt werden sollten, können Sie die Autoren über berufliche Netzwerke wie LinkedIn finden.

In diesem Sinne, fangen wir an.

1.2 Die Halford Group

Bob betrat das Bürogebäude der Halford Group, einem Hersteller von Konsumgütern, zu denen auch die meistverkaufte Gummiente gehörte. Nachdem er die Bürotüren passiert hatte, fühlte er sich in die Achtzigerjahre zurückversetzt. Die Besucher mussten sich am Eingang anmelden, Formulare ausfüllen, um im Falle eines Unfalls haft-