

Ausoche

MLOps

Kernkonzepte im Überblick



Mark Treveil und das Dataiku-Team

Übersetzung von Marcus Fraaß



Zu diesem Buch – sowie zu vielen weiteren O'Reilly-Büchern – können Sie auch das entsprechende E-Book im PDF-Format herunterladen. Werden Sie dazu einfach Mitglied bei oreilly.plus +:

MLOps – Kernkonzepte im Überblick

Machine-Learning-Prozesse im Unternehmen nachhaltig automatisieren und skalieren

Mark Treveil und das Dataiku-Team

Deutsche Übersetzung von Marcus Fraaß



Mark Treveil und das Dataiku-Team

Lektorat: Alexandra Follenius Übersetzung: Marcus Fraaß

Korrektorat: Sibylle Feldmann, www.richtiger-text.de

Satz: III-satz, www.drei-satz.de Herstellung: Stefanie Weidner

Umschlaggestaltung: Karen Montgomery, Michael Oréal, www.oreal.de Druck und Bindung: mediaprint solutions GmbH, 33100 Paderborn

Bibliografische Information der Deutschen Nationalbibliothek

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

ISBN:

Print 978-3-96009-172-1 PDF 978-3-96010-580-0 ePub 978-3-96010-581-7 mobi 978-3-96010-582-4

1. Auflage

Translation Copyright für die deutschsprachige Ausgabe © 2021 dpunkt.verlag GmbH Wieblinger Weg 17 69123 Heidelberg

Authorized German translation of the English edition of *Introducing MLOps: How to Scale Machine Learning in the Enterprise*, ISBN 9781492083290 © 2020 Dataiku. This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.

Dieses Buch erscheint in Kooperation mit O'Reilly Media, Inc. unter dem Imprint »O'REILLY«. O'REILLY ist ein Markenzeichen und eine eingetragene Marke von O'Reilly Media, Inc. und wird mit Einwilligung des Eigentümers verwendet.

Hinweis:

Dieses Buch wurde auf PEFC-zertifiziertem Papier aus nachhaltiger Waldwirtschaft gedruckt. Der Umwelt zuliebe verzichten wir zusätzlich auf die Einschweißfolie.



Schreiben Sie uns:

Falls Sie Anregungen, Wünsche und Kommentare haben, lassen Sie es uns wissen: kommentar@oreilly.de.

Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten. Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Verlags urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.

Es wird darauf hingewiesen, dass die im Buch verwendeten Soft- und Hardware-Bezeichnungen sowie Markennamen und Produktbezeichnungen der jeweiligen Firmen im Allgemeinen warenzeichen-, marken- oder patentrechtlichem Schutz unterliegen.

Alle Angaben und Programme in diesem Buch wurden mit größter Sorgfalt kontrolliert. Weder Autor noch Verlag noch Übersetzer können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buches stehen.

Inhalt

Te	il I Was ist MLOps, und warum wird es benötigt?	11
1	Warum jetzt, und was sind die Herausforderungen? MLOps – Definition und Herausforderungen. MLOps zum Reduzieren von Risiken Risikobeurteilung. Risikominderung Responsible AI durch MLOps MLOps zur Skalierung von Machine-Learning-Modellen Abschließende Überlegungen	15 16 20 20 21 22 23 24
2	An MLOps-Prozessen beteiligte Personen Fachexperten Data Scientists Data Engineers Software Engineers DevOps Modellrisikomanager/Auditor Machine Learning Architects Abschließende Überlegungen	25 27 29 31 32 33 34 34 35
3	Die Kernkomponenten von MLOps Eine Einführung in Machine Learning. Modellentwicklung. Festlegen von Geschäftszielen. Datenquellen und explorative Datenanalyse Feature Engineering und Feature Selection Training und Evaluierung. Reproduzierbarkeit Responsible AI	37 38 38 38 40 40 40

	Uberfuhrung in die Produktion und Deployment	42
	Arten und Elemente des Modell-Deployments	42
	Anforderungen beim Deployment von Modellen	44
	Monitoring	44
	Verantwortungsbereiche des DevOps-Teams	45
	Verantwortungsbereiche des Data-Science-Teams	45
	Verantwortungsbereiche der Managementebene	47
	Iteration und Lebenszyklus	47
	Iteration	48
	Die Feedback-Schleife	49
	Governance	50
	Daten-Governance	52
	Prozess-Governance	53
	Abschließende Überlegungen	54
	00	
Te	il II MLOps einsetzen	
	·	
4	Modellentwicklung	57
	Was genau sind Machine-Learning-Modelle?	58
	Theoretischer Hintergrund	58
	Einsatz in der Praxis	59
	Erforderliche Komponenten	60
	Unterschiedliche ML-Algorithmen – unterschiedliche	
	MLOps-Herausforderungen	61
	Explorative Datenanalyse	63
	Feature Engineering und Feature Selection	64
	Feature-Engineering-Techniken	64
	Wie die Auswahl der Features die MLOps-Strategie beeinflusst	65
	Experimente	67
	Modelle evaluieren und vergleichen	68
	Ein geeignetes Qualitätsmaß auswählen	69
	Gegenprüfen des Modellverhaltens (Cross-Checking)	71
	Auswirkungen von Responsible AI auf die Modellentwicklung	72
	Versionsverwaltung und Reproduzierbarkeit	75
	Abschließende Überlegungen	77
5	Vorbereitung für die Produktion	79
_	Laufzeitumgebungen	80
	Modelle aus der Entwicklungs- in die Produktivumgebung	00
	überführen	80
	uberrumen	00

	Datenzugriff vor Validierung und Inbetriebnahme in der	
	Produktion	82
	Abschließende Überlegungen zu Laufzeitumgebungen	83
	Risikobeurteilung von Modellen	83
	Der Zweck der Modellvalidierung	83
	Die Risikotreiber bei Machine-Learning-Modellen	84
	Qualitätssicherung im Rahmen der Verwendung von	
	Machine Learning	85
	Wichtige Überlegungen zum Testen	86
	Reproduzierbarkeit und Überprüfbarkeit	87
	Potenzielle Sicherheitsrisiken im Zusammenhang mit	
	Machine Learning	89
	Adversarial Attacks	89
	Weitere Sicherheitsrisiken	90
	Das Modellrisiko eindämmen	91
	Änderungen in der Umgebung	91
	Wechselwirkungen zwischen Modellen	92
	Fehlverhalten von Modellen	93
	Abschließende Überlegungen	94
6	Deployment in die Produktivumgebung	95
-	CI/CD-Pipelines.	95
	ML-Artefakte bauen	97
	Was beinhaltet ein ML-Artefakt?	97
	Die Testpipeline	98
	Deployment-Strategien	99
	Varianten des Modell-Deployments	100
	Überlegungen beim Überführen von Modellen in die	
	Produktivumgebung	100
	Wartung von Modellen im Produktivbetrieb	102
	Containerisierung	102
	Deployments skalieren	104
	Anforderungen und Herausforderungen	106
	Abschließende Überlegungen	107
7	Monitoring und Feedback-Schleife	109
	Wie häufig sollten Modelle neu trainiert werden?	110
	Leistungsabfall von Modellen überwachen	113
	Bewertung auf Basis der Ground Truth	113
	Abweichungen in den Eingabedaten erkennen	
	(Input-Drift-Detection)	116

	Drift-Erkennung in der Praxis	11
	Mögliche Ursachen für systematische Abweichungen in den	
	Daten	11
	Methoden zur Erkennung systematischer Abweichungen in den	
	Eingabedaten	11
	Die Feedback-Schleife	12
	Logging-System	12
	Modelle evaluieren	12
	Evaluierung während des Produktivbetriebs	12
	Abschließende Überlegungen	13
8	Modell-Governance	13
	Wer entscheidet, wie die Governance des Unternehmens aussieht?	13
	Anpassung der Governance an das Risikoniveau	13
	Aktuelle Regulierungen als Treiber der MLOps-Governance	13
	Gesetzliche Richtlinien für die US-Pharmaindustrie: GxP	13
	Regulierung des Modellrisikomanagements in der	
	Finanzbranche	13
	Datenschutzbestimmungen gemäß DSGVO und CCPA	13
	Die nächste Welle an KI-spezifischen Regulierungen	13
	Die Entstehung einer verantwortungsvollen KI (Responsible AI)	13
	Schlüsselelemente von Responsible AI	14
	1. Element: Daten	14
	2. Element: Bias	14
	3. Element: Inklusivität	14
	4. Element: Modellmanagement im großen Maßstab	14
	5. Element: Governance	14
	Eine Vorlage für MLOps-Governance	14
	1. Schritt: Verstehen und Kategorisieren der Analytics-	
	Anwendungsfälle	14
	2. Schritt: Eine ethische Grundhaltung einnehmen	14
	3. Schritt: Verantwortlichkeiten festlegen	14
	4. Schritt: Richtlinien für die Governance aufstellen	14
	5. Schritt: Einbinden von Richtlinien in den MLOps-Prozess	14
	6. Schritt: Werkzeuge für das zentrale Governance-Management	
	auswählen	15
	7. Schritt: Einbinden und Schulen	15
	8. Schritt: Überwachen und Optimieren	15
	Abschließende Überlegungen	1 4

Teil III MLOps-Anwendungsfälle aus der Praxis

	Verbraucherkrediten Hintergründe des geschäftlichen Anwendungsfalls Modellentwicklung Überlegungen zu Bias in Modellen	157 158
	Modellentwicklung	158
		159
	Produktionsvorbereitung	160
	Deployment in die Produktivumgebung	161
	Abschließende Überlegungen	161
10	MLOps in der Praxis: Empfehlungssysteme im Marketing	163
	Empfehlungssysteme im Wandel der Zeit	163
	Die Rolle von Machine Learning	164
	Push- oder Pull-Empfehlungen?	164
	Datenaufbereitung	165
	Experimente konzipieren und verwalten	166
	Training und Deployment von Modellen	167
	Skalierbarkeit und Anpassungsmöglichkeiten	168
	Monitoring- und Retraining-Strategie	168
	Auswertung der Anfragen in Echtzeit (Real-Time-Scoring)	169
	Möglichkeit, das Empfehlungssystem ein- oder auszuschalten	169
	Aufbau der Pipeline und Deployment-Strategie	169
	Monitoring und Feedback	171
	Modelle neu trainieren (Retraining)	171
	Modelle aktualisieren	171
	Über Nacht laufen und tagsüber ruhen lassen	172
	Möglichkeiten zur manuellen Anpassung von Modellen	172
	Möglichkeit der automatischen Verwaltung von	
	Modellversionen	173
	Die Qualität des Modells überwachen	173
	Abschließende Überlegungen	174
11	MLOps in der Praxis: die Verbrauchsprognose am Beispiel der	
	Lastprognose	177
	Stromversorgungssysteme	177
	Datenerhebung	179
	Vom Anwendungsfall abhängig: Machine Learning verwenden	
	oder nicht?	181
	Räumliche und zeitliche Differenzierung	182

Umsetzung		
Modellentwicklung		
Deployment		
Monitoring		
Abschließende Überlegungen	1	

Vorwort

Wir haben einen Wendepunkt in der Geschichte des maschinellen Lernens erreicht, an dem die Technologie aus dem theoretischen und dem akademischen Umfeld in die »reale Welt« vorgedrungen ist – d. h. in Unternehmen, die alle möglichen Arten von Dienstleistungen und Produkten für Menschen auf der ganzen Welt anbieten. Dieser Wandel ist nicht nur äußerst spannend, sondern stellt auch eine Herausforderung dar, da er die Komplexität von Machine-Learning-Modellen mit der Komplexität moderner Unternehmen zusammenbringt.

Eine Schwierigkeit, wenn Unternehmen von der experimentellen Nutzung des Machine Learning zur Skalierung in Produktivumgebungen übergehen, stellt die Wartung dar. Wie können Unternehmen von der Verwaltung eines einzigen Modells zum Managen von Dutzenden, Hunderten oder sogar Tausenden übergehen? Hier kommt nicht nur MLOps ins Spiel, sondern auch die bereits erwähnten Komplexitäten, und zwar sowohl auf der technischen als auch auf der geschäftlichen Ebene. Dieses Buch führt Leserinnen und Leser in diese Herausforderungen ein und bietet gleichzeitig praktische Einblicke und Lösungen für die Entwicklung von Fähigkeiten im Bereich MLOps.

An wen sich dieses Buch richtet

Wir haben dieses Buch speziell für Managerinnen und Manager von Analytics- und IT-Operations-Teams geschrieben, also für die Personen, die direkt mit der Aufgabe betraut sind, Machine Learning (ML) in einer Produktivumgebung zu skalieren. Da MLOps ein neues Feld ist, haben wir dieses Buch als Leitfaden für das Erstellen einer erfolgreichen MLOps-Umgebung konzipiert, beginnend bei den organisatorischen bis hin zu den technischen Herausforderungen.

Aufbau des Buchs

Dieses Buch ist in drei Hauptteile gegliedert. Teil I, *Was ist MLOps, und warum wird es benötigt?*, stellt das Thema MLOps grundlegend vor und geht darauf ein, wie (und warum) es sich als Disziplin entwickelt hat, wer beteiligt sein sollte, um MLOps erfolgreich durchzuführen, und welche Bausteine erforderlich sind.

Teil II, *MLOps einsetzen*, orientiert sich im Wesentlichen an dem Lebenszyklus von Machine-Learning-Modellen und umfasst mehrere Kapitel, die sich mit der Entwicklung von Modellen, der Produktionsvorbereitung, dem Deployment in die Produktivumgebung, dem Monitoring und der Governance befassen. Diese Kapitel behandeln nicht nur allgemeine Aspekte, sondern auch solche, die sich speziell auf den Einsatz von MLOps in jeder Phase des Lebenszyklus beziehen, wodurch die in Kapitel 3, *Die Kernkomponenten von MLOps*, behandelten Themen noch genauer ausgeführt werden.

Teil III, *MLOps-Anwendungsfälle aus der Praxis*, enthält konkrete Beispiele, die Ihnen zeigen sollen, wie MLOps heute in Unternehmen aussieht, wie es konzeptioniert ist und mit welchen Implikationen zu rechnen ist. Obwohl die Firmennamen fiktiv gewählt wurden, basieren die Beispiele auf Erfahrungen, die reale Unternehmen im Zusammenhang mit MLOps und einem im großen Maßstab angelegten Modellmanagement gemacht haben.

Danksagungen

Wir möchten dem gesamten Dataiku-Team für seine Unterstützung bei der Entstehung dieses Buchs danken, angefangen von der Konzeption bis hin zur Fertigstellung. Es war eine echte Teamleistung und ist, wie die meisten Dinge, die wir bei Dataiku machen, das Ergebnis einer intensiven Zusammenarbeit zwischen unzähligen Menschen und Teams.

Danke an alle, die unsere Idee, dieses Buch mit O'Reilly herauszubringen, von Anfang an unterstützt haben. Danke an alle, die beim Schreiben und Herausgeben mitgeholfen haben. Ebenfalls danke an alle, die uns ehrliches Feedback gegeben haben (auch wenn es zur Folge hatte, noch mehr zu schreiben oder umzuschreiben). Danke an alle, die uns intern stets ermutigt haben, und natürlich an alle, die uns geholfen haben, das finale Produkt der weltweiten Öffentlichkeit vorstellen zu können.

Was ist MLOps, und warum wird es benötigt?

Warum jetzt, und was sind die Herausforderungen?

Machine Learning Operations (MLOps) entwickelt sich zusehends zu einer unverzichtbaren Komponente, um Data-Science-Projekte im Unternehmen erfolgreich in den Einsatz zu bringen (siehe Abbildung 1-1). Dabei handelt es sich um Prozesse, die dem Unternehmen und den Verantwortlichen dabei helfen, im Zusammenhang mit Data Science, Machine Learning und KI-Projekten langfristigen Wert zu generieren und Risiken zu reduzieren. Dennoch stellt MLOps ein relativ neues Konzept dar. Warum hat es also scheinbar über Nacht Einzug in das Data-Science-Lexikon erhalten? In diesem einführenden Kapitel wird erläutert, was MLOps auf einer übergeordneten Ebene ist, welche Herausforderungen es mit sich bringt, warum es für eine erfolgreiche Data-Science-Strategie im Unternehmen unverzichtbar geworden ist und, was besonders wichtig ist, warum es gerade jetzt in den Vordergrund rückt.

MLOps im Vergleich zu ModelOps und AlOps

MLOps (oder ModelOps) ist eine relativ neue Fachdisziplin, die seit Ende des Jahres 2018 unter diesen Namen in Erscheinung trat. Die beiden Termini – MLOps und ModelOps – werden zum Zeitpunkt der Erstellung dieses Buchs weitgehend synonym verwendet. Einige argumentieren jedoch, dass ModelOps umfassender als MLOps ist, da es nicht nur um Machine-Learning-(ML)-Modelle geht, sondern um jede Art von Modellen (z.B. auch regelbasierte Modelle). Im Rahmen dieses Buchs werden wir uns speziell mit dem Lebenszyklus von ML-Modellen befassen und daher den Begriff *MLOps* verwenden.

Auch wenn es manchmal mit MLOps verwechselt wird, bezieht sich AIOps hingegen auf ein ganz anderes Thema und bezeichnet den Prozess der Lösung operativer Herausforderungen im Rahmen des Einsatzes von künstlicher Intelligenz (d.h. KI für DevOps). Ein Beispiel wäre eine Form der vorausschauenden Wartung im Zusammenhang mit Netzwerkausfällen, bei der DevOps-Teams auf mögliche Probleme aufmerksam gemacht werden, bevor sie auftreten. Obwohl AIOps für sich genommen wichtig und interessant ist, liegt es außerhalb des Rahmens dieses Buchs.

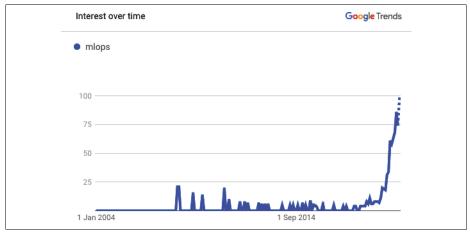


Abbildung 1-1: Darstellung des exponentiell verlaufenden Suchtrends des Begriffs »MLOps« (ohne gleichzeitige Berücksichtigung des Terminus »ModelOps«)

MLOps – Definition und Herausforderungen

Im Kern ist MLOps die Standardisierung und Straffung des Lebenszyklusmanagements von ML-Modellen (siehe Abbildung 1-2). Doch weshalb muss der ML-Lebenszyklus überhaupt gestrafft werden? Oberflächlich betrachtet, könnte man annehmen, dass die Arbeitsschritte, die vom Geschäftsproblem zu einem ML-Modell führen, sehr einfach sind.

Für die meisten traditionellen Unternehmen ist die Entwicklung mehrerer Machine-Learning-Modelle und deren Einsatz in einer Produktivumgebung relativ neu. Bis vor Kurzem war die Anzahl der Modelle vielleicht noch überschaubar, oder es bestand einfach weniger Interesse daran, diese Modelle und ihre Abhängigkeiten auf unternehmensweiter Ebene zu verstehen. Mit der fortschreitenden Automatisierung von Entscheidungsprozessen (d.h. mit einer zunehmenden Verbreitung von Entscheidungen, die ohne menschliches Zutun getroffen werden) rücken Modelle immer stärker in den Fokus, und parallel dazu wird auch das Management von Modellrisiken auf höchster Ebene immer wichtiger.

Insbesondere in Bezug auf die Anforderungen und die genutzten Tools erweist sich das Lebenszyklusmanagement von Machine-Learning-Modellen in einem Unternehmen tatsächlich als durchaus komplex (siehe Abbildung 1-3).

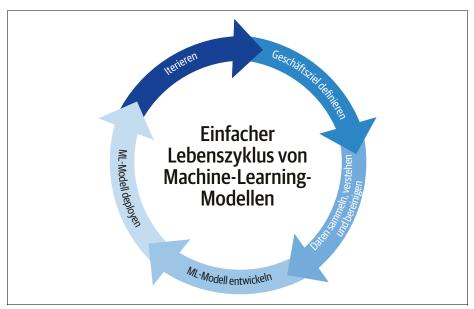


Abbildung 1-2: Eine vereinfachte Darstellung des Lebenszyklus von ML-Modellen, die die Notwendigkeit von MLOps nur unzureichend abbildet, speziell im Vergleich zu Abbildung 1-3

Es gibt drei Hauptgründe dafür, dass das Lebenszyklusmanagement skalierbarer ML-Modelle eine Herausforderung darstellt:

- Es gibt zahlreiche Abhängigkeiten. Nicht nur die Daten ändern sich ständig, sondern auch die geschäftlichen Anforderungen. Neue Informationen müssen kontinuierlich an das Unternehmen zurückgegeben werden, um sicherzustellen, dass der Produktivbetrieb des Modells, auch in Bezug auf die Akkuranz der Produktionsdaten, mit den Erwartungen übereinstimmt und was von entscheidender Bedeutung ist dass das ursprüngliche Problem gelöst bzw. die ursprüngliche Zielsetzung erreicht wird.
- Nicht alle sprechen die gleiche Sprache. Auch wenn am ML-Lebenszyklus Mitarbeiter aus Business-, Data-Science- und IT-Teams beteiligt sind, ist es nicht zwingend gegeben, dass diese Teams die gleichen Tools oder in vielen Fällen sogar die gleichen grundlegenden Fähigkeiten, die als Kommunikationsbasis dienen, teilen.
- Data Scientists sind keine Softwareentwickler. Die meisten sind auf die Entwicklung und Evaluierung von Modellen spezialisiert, und ihr Know-how liegt nicht zwingend in der Entwicklung von Anwendungen. Obwohl sich dies im Laufe der Zeit ändern könnte, da sich einige Data Scientists auf die Bereitstellung bzw. den Betrieb von Modellen spezialisieren werden, müssen derzeit viele Data Scientists mit verschiedenen Rollen gleichzeitig jonglieren, was es schwierig macht, eine davon vollständig auszufüllen. Die Überforde-

rung von Data Scientists wird insbesondere im Rahmen der Skalierung – wenn es immer mehr Modelle zu verwalten gibt – problematisch. Noch komplexer wird es, wenn man zusätzlich die Fluktuation der Mitarbeitenden in den Datenteams berücksichtigt: Schließlich gibt es nicht wenige Data Scientists, die sich plötzlich dazu gezwungen sehen, Modelle zu verwalten, die sie nicht selbst entwickelt haben.

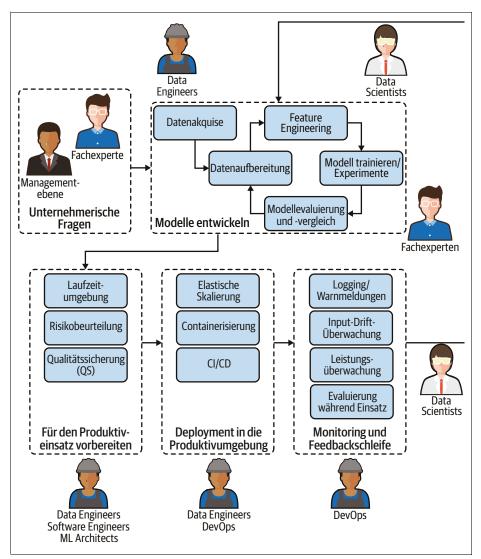


Abbildung 1-3: Ein realistischeres Bild des Lebenszyklus eines ML-Modells in einem modernen Unternehmen, in den viele verschiedene Personen mit völlig unterschiedlichen Fähigkeiten involviert sind, die oft völlig unterschiedliche Tools verwenden

Wenn Ihnen die Definition (oder lediglich die Bezeichnung MLOps) bekannt vorkommt, liegt das vor allem daran, dass sie sich stark an das Konzept, das hinter Dev-Ops steht, anlehnt: DevOps dient dazu, die Prozesse im Rahmen von Software-änderungen und -aktualisierungen zu straffen. In der Tat haben beide Konzepte ziemlich viel gemeinsam. Zum Beispiel geht es bei beiden darum,

- eine robuste Automatisierung und vertrauensvolle Zusammenarbeit zwischen den Teams zu gewährleisten,
- den Leitgedanken einer kooperativen Zusammenarbeit und einer verbesserten Kommunikation zwischen den Teams zu fördern,
- den Lebenszyklus des Diensts ganzheitlich (Build, Test, Release) zu berücksichtigen und
- den Schwerpunkt auf eine kontinuierliche Auslieferung (*Continuous Delivery*) und hohe Qualitätsanforderungen zu setzen.

Es gibt jedoch einen entscheidenden Unterschied zwischen MLOps und DevOps, der dafür sorgt, dass letzteres Konzept nicht sofort auf Data-Science-Teams übertragbar ist: In der Produktion unterscheidet sich das Deployment von Softwareprogrammen grundlegend vom Deployment von ML-Modellen. Während Softwareprogramme relativ statisch sind (»relativ«, da viele moderne Software-as-a-Service-(SaaS-)Unternehmen bereits über DevOps-Teams verfügen, die recht schnell iterieren und in der Produktion mehrmals am Tag deployen können), ändern sich Daten hingegen ständig, was bedeutet, dass ML-Modelle ständig neu (hinzu-)lernen und sich an neue Eingabedaten anpassen – oder eben nicht. Die dieser Umgebung zugrunde liegende Komplexität – einschließlich der Tatsache, dass ML-Modelle sowohl aus Programmcode als auch aus Daten bestehen – ist der Grund dafür, dass MLOps zu einer neuen und einzigartigen Disziplin heranwächst.

Und was hat es mit DataOps auf sich?

Zusätzlich zur komplexen Gegenüberstellung von MLOps und DevOps müssen wir noch den Begriff DataOps abgrenzen, der im Jahr 2014 von IBM eingeführt wurde. DataOps zielt darauf ab, geschäftsfähige Daten bereitzustellen, die schnell für die Nutzung verfügbar sind, wobei der Datenqualität und der Metadatenverwaltung ein besonderer Stellenwert beigemessen wird. Wenn es beispielsweise eine plötzliche Änderung in den Daten gibt, auf denen ein Modell beruht, würde ein Data-Ops-System das Businessteam alarmieren, damit es sich sorgfältig mit den neuesten Erkenntnissen befasst, und das Datenteam würde ebenfalls informiert werden, damit es die Änderung untersuchen oder ein Upgrade einer Bibliothek rückgängig machen und die entsprechende Partition neu erstellen kann.

Die Entwicklung von MLOps überschneidet sich daher auf einer gewissen Ebene mit DataOps, obwohl MLOps einen Schritt weitergeht und durch zusätzliche Kernfunktionen (die in Kapitel 3 ausführlicher erläutert werden) eine noch stärkere Robustheit bietet.

Wie bei DevOps und später auch bei DataOps konnten sich Teams bis vor Kurzem ohne vordefinierte und zentralisierte Prozesse behelfen, vor allem weil sie maschinelle Lernmodelle – auf Unternehmensebene – nicht in so großem Maßstab angelegt in die Produktion brachten. Jetzt wendet sich das Blatt, und die Teams suchen zunehmend nach Möglichkeiten, einen mehrstufigen, multidisziplinären und mehrphasigen Prozess mit einer heterogenen Umgebung und einem Rahmen für MLOps-Best-Practices zu formalisieren, was keine kleine Aufgabe darstellt. Teil II des Buchs, *MLOps einsetzen*, wird Ihnen hierzu einen Leitfaden bieten.

MLOps zum Reduzieren von Risiken

MLOps ist wichtig für jedes Team, das auch nur ein Modell im Produktivbetrieb hat, da je nach Modell eine kontinuierliche Leistungsüberwachung und -anpassung erforderlich ist. Indem es einen sicheren und zuverlässigen Betrieb ermöglicht, ist MLOps der Schlüssel zur Eindämmung der Risiken, die durch den Einsatz von ML-Modellen entstehen. Allerdings sind mit dem Einsatz von MLOps auch Kosten verbunden, für jeden Anwendungsfall sollte daher eine angemessene Kosten-Nutzen-Bewertung durchgeführt werden.

Risikobeurteilung

In Bezug auf ML-Modelle gibt es sehr unterschiedliche Risiken. Zum Beispiel sind die Risiken bei der Nutzung eines Empfehlungssystems, das einmal im Monat verwendet wird, um zu entscheiden, welches Marketingangebot an einen Kunden geschickt werden soll, viel geringer als bei einer Reiseplattform, deren Preissetzung und Umsatz von einem ML-Modell abhängen. Daher sollte sich die Analyse bei der Betrachtung von MLOps als Möglichkeit zur Risikominimierung auf folgende Risiken erstrecken:

- Das Risiko, dass das Modell für eine bestimmte Zeitspanne nicht verfügbar ist.
- Das Risiko, dass das Modell f
 ür eine bestimmte Beobachtung eine unzutreffende Vorhersage liefert.
- Das Risiko, dass die Genauigkeit oder die Fairness des Modells mit der Zeit abnimmt.
- Das Risiko, dass die zur Wartung des Modells erforderlichen Kompetenzen (d.h. die Fähigkeiten der jeweiligen Data Scientists) nicht mehr zur Verfügung stehen.

Bei Modellen, die weit verbreitet sind und außerhalb des eigenen Unternehmens eingesetzt werden, sind die Risiken in der Regel größer. Wie in Abbildung 1-4 gezeigt, basiert die Risikobeurteilung im Allgemeinen auf zwei Größen: der Eintrittswahrscheinlichkeit und dem Schadensausmaß des unerwünschten Ereignisses. Maßnahmen zur Risikominderung basieren in der Regel auf einer Kombination aus beidem, dem sogenannten Risikograd bzw. -ausmaß des Modells. Die Risikobeur-

teilung sollte zu Beginn eines jeden Projekts durchgeführt und in regelmäßigen Abständen neu bewertet werden, da Modelle auf eine ursprünglich nicht vorgesehene Weise verwendet werden können.

			5 x 5 Ris	iko-Matrix	(
1	Sehr wahrscheinlich	5 Mittel	10 Hoch	15 Hoch	20 Schwerwiegend	25 Schwerwiegend
רוועפונ	Wahrscheinlich	4 Mittel	8 Mittel	12 Hoch	16 Hoch	20 Schwerwiegend
	Möglich	3 Gering	6 Mittel	9 Mittel	12 Hoch	15 Hoch
S MAGILIS	Unwahr- scheinlich	2 Gering	4 Mittel	6 Mittel	8 Mittel	10 Hoch
Eintrittswahrscheinlichkeit	Selten	1 Gering	2 Gering	3 Gering	5 Mittel	6 Mittel
		Sehr niedrig	Niedrig	Mittel- mäßig	Groß	Sehr groß
		Sc	chadensau	smaß		

Abbildung 1-4: Eine Tabelle, die Entscheidungsträgern bei der quantitativen Risikobeurteilung hilft und auf Eintrittswahrscheinlichkeit und Schadensausmaß des Ereignisses basiert.

Risikominderung

MLOps trägt vor allem dann entscheidend zur Risikominderung bei, wenn ein zentrales Team (mit einer klaren Berichterstattung über seine Aktivitäten – was nicht bedeutet, dass es in einem Unternehmen nicht mehrere solcher Teams geben kann) mehr als eine Handvoll Modelle im operativen Einsatz hat. An diesem Punkt wird es schwierig, den Gesamtüberblick über die Zustände dieser Modelle ohne eine Form der Standardisierung zu behalten, die es ermöglicht, für jedes dieser Modelle die entsprechenden Maßnahmen zur Risikominderung ergreifen zu können (siehe den Abschnitt »Anpassung der Governance an das Risikoniveau« auf Seite 133).

Es ist aus vielen Gründen riskant, ML-Modelle in die Produktivumgebung zu überführen, ohne dass eine entsprechende MLOps-Infrastruktur vorhanden ist, zumal eine vollständige Bewertung der Leistung bzw. der Güte eines ML-Modells oft nur in der Produktivumgebung erfolgen kann. Warum? Weil Prognosemodelle nur so gut sind wie die Daten, auf denen sie trainiert wurden. Das bedeutet, dass die Trainingsdaten ein gutes Abbild der Daten sein müssen, die in der Produktivumgebung anfallen. Wenn sich die Rahmenbedingungen in der Produktion ändern, wird infolgedessen wahrscheinlich relativ schnell auch die Güte des Modells darunter leiden (siehe Kapitel 5 für Einzelheiten).