Maull (Hrsg.)

Managementleitfaden Data Science

Aufbereitung datengetriebener Fachbegriffe und deren Zusammenhänge

Finanz Colloquium Heidelberg, 2021

Zitiervorschlag:
Autor, in: Maull (Hrsg.): Managementleitfaden Data Science, S. XX.

ISBN: 978-3-95725-968-4

Im Bosseldorn 30, 69126 Heidelberg
www.FCH-Gruppe.de
Info@FCH-Gruppe.de
Satz: MetaLexis GbR, Niedernhausen
Druck: koronamedien, Dudenhofen

Maull (Hrsg.)

Managementleitfaden Data Science

Aufbereitung datengetriebener Fachbegriffe und deren Zusammenhänge

Kristin Benedikt
Richterin am Bayerischen Verwaltungsgericht
Regensburg

Dr. Markus Held
Referatsleiter Informationssicherheit in der IT-Konsolidierung
Bundesamt für Sicherheit in der Informationstechnik (BSI)
Bonn

Christian Koch
Data Architect
TeamBank AG Nürnberg
Nürnberg

Christian Maull (Hrsg.)
Datenschutzbeauftragter
TeamBank AG Nürnberg
Nürnberg

Denise Primus
Rechtsanwältin,
Schlatter Rechtsanwälte Steuerberater PartG mbB

Rüdiger Vicari
Leiter Data Architecture & Governance
TeamBank AG Nürnberg
Nürnberg

Xu Zhu
Doktorand am Lehrstuhl für Betriebswirtschaftslehre
Universität Bremen

Finanz Colloquium Heidelberg, 2021

A Vorwort

Big Data, Machine Learning, Smart Data, Künstliche Intelligenz.

Begriffe, die vor wenigen Jahren eher dem Themengebiet der Science-Fiction zugeordnet worden wäre, sind heute mehr und mehr in aller Munde. Firmen wie Google oder Facebook zeigen eindrucksvoll auf, welche Chancen, Potenziale und Ertragsmöglichkeiten in der Nutzung derartiger Datenbestände liegen. So verwundert es nicht, dass sich auch die, zumeist bodenständige, deutsche Wirtschaft für diese Themen zu interessieren beginnt, obwohl das Internet für viele zunächst doch Neuland zu sein scheint.

Dem geneigten Leser werden an dieser Stelle in diesem Buch zentrale Begrifflichkeiten erläutert und Praxisbezüge zu diesem weiten Themenfeld aufgezeigt. Wie so oft sollte bei der Betrachtung von Chancen und Möglichkeiten immer eine Würdigung limitierender oder möglicher (Risiko-)Faktoren die Ausführungen abrunden. Aus diesem Grund werden eng mit dem Themenfeld Data Science verknüpfte Fachgebiete zunächst in ihren Grundzügen dargestellt, um anschließend die Brücke zu dem skizzierten Praxisfall zu schlagen.

Im Zuge dieser Betrachtung gehen u. a. Würdigungen aus bankaufsichtsrechtlicher und datenschutzaufsichtsrechtlicher Sicht in diesem Werk auf. Daneben werden Aspekte im Kontext der Informationssicherheit und der IT-Governance, IT-Rechtsgesichtspunkten, Blickwinkel aus dem Risikocontrolling und dem Datenschutz näher beleuchtet.

B Aufsichtliche Würdigung zum Datenschutz

Bei der Entwicklung und Nutzung von KI-Systemen sind diverse rechtliche Anforderungen zu beachten. Darunter fallen nicht nur Anforderungen des Haftungs- und Patentrechts, sondern auch datenschutzrechtliche Anforderungen. Insbesondere wenn bei den unterschiedlichen Anwendungsszenarien personenbezogene Daten verarbeitet werden, sind die Anforderungen der Datenschutz-Grundverordnung (DSGVO) zu beachten. Die Datenschutz Grundverordnung schützt die Grundrechte und Grundfreiheiten natürlicher Personen, insbesondere das Recht auf Schutz personenbezogener Daten. Die Datenschutz-Grundverordnung enthält eine Vielzahl an Anforderungen, die bei der Entwicklung und beim Einsatz von KI-Systemen berücksichtigt werden müssen. Vor allem bei technischen Innovationen werden Verantwortliche vor die Herausforderung gestellt, die abstrakten Regelungen der DSGVO auf ihre konkreten technischen Entwicklungen zu übertragen. In diesem Zusammenhang kommt den Datenschutzaufsichtsbehörden eine besondere Rolle zu. Diese Rolle der Aufsichtsbehörden sowie ausgewählte Fragestellungen des Datenschutzrechts werden im Folgenden erläutert.

I. Rolle der Aufsichtsbehörden

Die Datenschutz-Grundverordnung regelt die Aufgaben und Befugnisse der europäischen Datenschutzaufsichtsbehörden, um einen einheitlichen Vollzug der datenschutzrechtlichen Vorgaben in Europa sicherzustellen. Zu den bedeutendsten Aufgaben der Aufsichtsbehörden im Hinblick auf Maschine-Learning, KI und Big Data gehören die Zusammenarbeit sowie der Informationsaustausch zwischen den Behörden, das Verfolgen von maßgeblichen Entwicklungen sowie die Beratung gem. Art. 57 Abs.1 lit. g), i) und l) DSGVO.

Vor allem die deutschen Aufsichtsbehörden verfolgen das Ziel, Verantwortliche bei der technischen Entwicklung durch Beratungsangebote und Öffentlichkeitsarbeit zu unterstützen. Aufsichtliche Maßnahmen, wie die Anordnung konkreter technische Maßnahmen oder gar das Verbot eine Datenverarbeitung durchzuführen, erst recht Sanktionen mittels Bußgeld waren in der Vergangenheit bei den deutschen Aufsichtsbehörden das letzte Mittel. Im Vordergrund der deutschen Aufsichtsbehörden steht nach wie vor, Verstöße gegen die Datenschutz-Grundverordnung zu verhindern.

1. Beratung und Öffentlichkeitsarbeit

Die Aufsichtsbehörden dürfen die Öffentlichkeit über Risiken, Vorschriften, Garantien und Rechte im Zusammenhang mit der Verarbeitung personenbezogener Daten sensibilisieren und darüber aufklären, vgl. Art. 57 Abs. 1 lit. b) DSGVO. Diese Aufgabe ist vor allem im Hinblick auf technologische Entwicklungen und Innovationen von großer Bedeutung. Die Wirkung von Warnungen der Aufsichtsbehörden darf von Verantwortlichen keinesfalls unterschätzt werden. Äußerungen der Aufsichtsbehörden in der

Öffentlichkeit haben eine zunehmend mediale Wirkung und können das Vertrauen und das Konsumverhalten der Verbraucher maßgeblich beeinflussen. Vor allem seit Geltung der Datenschutz-Grundverordnung ist das Thema Datenschutz im Bewusstsein der breiten Öffentlichkeit. Dies führt in vielen Fällen dazu, dass Aufsichtsbehörden von der Presse kontaktiert werden und als Experten im Datenschutz zur aktuellen Geschehnissen oder Entwicklungen interviewt werden. Nicht selten äußerten Aufsichtsbehörden Bedenken zur datenschutzrechtlichen Zulässigkeit eines Produkts oder kündigten gar aufsichtliche Maßnahmen an. Bereits in der Vergangenheit haben Warnungen der Aufsichtsbehörden zu Big Data oder KI-Anwendungen die mediale Aufmerksamkeit gewonnen beispielsweise im Zusammenhang mit Gesichtserkennung[1], Sprachanalysen bei Bewerbungsverfahren[2] oder Videokonferenzsystemen[3].

Verantwortliche sollten daher insbesondere bei Produkteinführungen im Consumer-Bereich in Erwägung ziehen, aktiv auf die zuständige Aufsichtsbehörde zuzugehen und gegebenenfalls vorab über KI-Systeme informieren. Eine solche „Vorabinformation" kann auch dazu dienen, noch offene Fragen zu diskutieren, sowie die Fachkunde der Aufsichtsbehörde in Anspruch zu nehmen. Im Rahmen einer Beratung durch die Aufsichtsbehörde kann auch erörtert werden, ob der Verantwortliche alle Risiken für die Rechte und Freiheiten bei der Anwendung von KI-Systemen erfasst und diese bei der Umsetzung der technischen und organisatorischen Maßnahmen berücksichtigt hat.

Darüber hinaus sollten Verantwortliche die Publikationen der Aufsichtsbehörden auf nationaler sowie europäischer Ebene verfolgen. Die Datenschutzkonferenz, die Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder, haben sich bereits mehrfach zu den Anforderungen Datenschutz konformer KI-Systeme geäußert. So gehören unter anderem das „Positionspapier der DSK zu empfohlenen technischen und organisatorischen Maßnahmen bei der Entwicklung und dem Betrieb von KI-Systemen"[4] sowie die „Hambacher Erklärung zur künstlichen Intelligenz" [5] zur Pflichtlektüre eines jeden Verantwortlichen. Die deutschen Datenschutzaufsichtsbehörden klären in den Veröffentlichungen über die wesentlichen datenschutzrechtlichen Anforderungen auf, die bei der Entwicklung bis hin zum Einsatz der KI-Systeme zu beachten sind.

Ebenso gehört das Gutachten der Datenethikkommission zur Standardliteratur für Verantwortliche. [6] Die Datenethikkommission, ein

[1] Heike Anger, Gesichtsscan im Supermarkt ist unbedenklich, 12.6.2017, abrufbar unter: https://www.handelsblatt.com/politik/deutschland/datenschuetzer-zu-real-werbebildschirm-gesichtsscan-im-supermarkt-ist-unbedenklich/19921456.html?ticket=ST-3662636-fA3eG0eTv9spsida6ABL-ap3

[2] Silke Gronwald, „Bewerbungsgespräch: Was die Stimme über unsere Persönlichkeit verrät", 03.09.2016, abufbar unter: https://www.stern.de/wirtschaft/job/bewerbungsgespraech--was-die-stimme-ueber-unsere-persoenlichkeit-verraet-7038038.html

[3] Berliner Beauftrage für Datenschutz und Informationsfreiheit zu rechtswidrigen Videokonferenzsystemen, abrufbar unter: https://www.datenschutz-berlin.de/fileadmin/user_upload/pdf/orientierungshilfen/2020-BlnBDI-Empfehlungen_Videokonferenzsysteme.pdf

[4] DSK, Positionspapier der DSK zu empfohlenen technischen und organisatorischen Maßnahmen bei der Entwicklung und dem Betrieb von KI-Systemen, 6. November 2019, abrufbar unter: https://www.datenschutzkonferenz-online.de/media/en/20191106_positionspapier_kuenstliche_intelligenz.pdf

[5] DSK, Hambacher Erklärung zur künstlichen Intelligenz, 3. April 2019, abrufbar unter: https://www.datenschutzkonferenz-online.de/media/en/20190405_hambacher_erklaerung.pdf

[6] Abrufbar unter: https://www.bmi.bund.de/DE/themen/it-und-digitalpolitik/datenethikkommission/arbeitsergebnisse-der-dek/arbeitsergebnisse-der-dek-node.html

Expertengremium, dem unter anderem Experten des Datenschutzes angehörten, befasste sich mit ethischen Maßstäben und Leitlinien im Zusammenhang mit datenbasierten Technologien. Ein wesentlicher Arbeitsauftrag war es, konkrete Handlungsempfehlungen, insbesondere zum Datenschutz, zu entwickeln.

2. Vollzug und Sanktionen

Viele Verantwortliche fürchten die hohen Bußgelder von bis zu 20 Millionen EUR oder im Falle eines Unternehmens von bis zu 4 % des weltweit erzielten Jahresumsatzes, die die Datenschutz-Grundverordnung regelt. Aber auch aufsichtliche Maßnahmen, wie ein vorübergehendes oder endgültiges Verbot von Verarbeitungstätigkeiten kann zu erheblichen wirtschaftlichen Einbußen führen. Darüber hinaus sollten Verantwortliche berücksichtigen, dass ein Verstoß, der in der Öffentlichkeit bekannt wird, stets ein Reputationsverlust mit sich bringt und gegebenenfalls weiteren wirtschaftlichen Schaden verursacht.

Nichtsdestotrotz sollten die Sanktionsbefugnisse der Aufsichtsbehörden nicht der alleinige Antrieb für eine datenschutzkonforme Entwicklung sein. Die datenschutzrechtlichen Anforderungen, wie die Pflicht zur Transparenz oder die Sicherheit der Verarbeitung dienen nicht ausschließlich dem Schutz personenbezogener Daten, sondern zugleich dem Geschäfts- und Geheimnisschutz und somit auch dem Verantwortlichen selbst.

Für den Fall, dass die Aufsichtsbehörden ein aufsichtliches Verfahren oder ein Bußgeldverfahren öffnen, kommt der Rechenschaftspflicht gemäß Art. 5 Abs. 2 DSGVO eine besondere Bedeutung zu. Der Verantwortliche ist verpflichtet, die Einhaltung der datenschutzrechtlichen Anforderungen nachzuweisen. Zwar gilt im Verwaltungsverfahren als auch im Ordnungswidrigkeitenverfahren der Amtsermittlungsgrundsatz, d. h. die zuständige Aufsichtsbehörde muss alle relevanten Tatsachen ermitteln, um einen Verstoß festzustellen. Allerdings stoßen die Aufsichtsbehörden im Rahmen ihrer Ermittlungen an ihre Grenzen, da sie auf die Mitwirkung des Verantwortlichen angewiesen sind, um einen Sachverhalt vollständig zu erfassen und datenschutzrechtlich zu bewerten.

Ob personenbezogene Daten vollständig gelöscht wurden, ob es sich bei Trainingsdaten der KI-Komponenten um anonymisierte Daten handelt, die tatsächlich keinen Rückschluss auf eine identifizierte oder identifizierbare Person ermöglichen und ob die ergriffenen technischen Maßnahmen geeignet sind, um ein dem Risiko angemessenes Schutzniveau zu gewährleisten, lässt sich nur überprüfen, wenn der Verantwortliche dies anhand einer umfassenden Dokumentation nachweisen und auf Verlangen auch demonstrieren kann.

Die Rechenschaftspflicht kann dem Verantwortlichen auch dazu dienen, sich in einem aufsichtliche Verfahren oder Bußgeldverfahren zu entlasten. Nicht immer mag die Rechtsauffassung des Verantwortlichen mit der der Aufsichtsbehörden übereinstimmen.

Kann der Verantwortliche jedoch substantiiert nachweisen, dass er sich mit den datenschutzrechtlichen Anforderungen bei KI-Systemen vertieft auseinandergesetzt hat, die Publikationen der Aufsichtsbehörden berücksichtigt hat, Gutachten eingeholt oder Audits durchgeführt hat, so dürfte dies in vielen Fällen die Vorwerfbarkeit Bußgeldverfahren entfallen lassen.

II. Transparenz

Eine besondere Herausforderung bei der Entwicklung und beim Einsatz von KI Systemen ist die Transparenz. Die Datenschutz Grundverordnung verlangt, dass personenbezogene Daten in einer für die betroffene Person nachvollziehbaren Weise verarbeitet werden müssen, Art. 5 Abs. 1 a) DSGVO. Der Grundsatz der Transparenz wird in der Datenschutz-Grundverordnung durch eine Vielzahl an Regelungen konkretisiert. So muss der Verantwortliche die betroffene Person in präziser, transparenter, verständlicher und leicht zugänglicher Form in einer klaren und einfachen Sprache informieren. Inhaltlich umfasst diese Informationspflicht unter anderem die Angabe

- der Zwecke, für die personenbezogene Daten verarbeitet werden,
- die Empfänger oder Kategorien von Empfängern der personenbezogenen Daten sowie
- die Speicherdauer.

Weitere Mindestanforderungen der Informationspflicht sind in den Artikeln 13 und 14 DSGVO geregelt.

Die Transparenz spielt auch eine entscheidende Rolle, wenn für die Verarbeitung die Einwilligung der betroffenen Person eingeholt werden soll. Eine Einwilligung ist nur wirksam, wenn die Einwilligung von der betroffenen Person vorab, in informierter Weise und freiwillig für den bestimmten Fall abgegeben wurde. Informiert ist die Einwilligung, wenn die betroffene Person alle Informationen erhalten hat, um die Tragweite und Folgen einer Datenverarbeitung abzuschätzen.

In Beratungsgesprächen mit den Aufsichtsbehörden tragen Verantwortliche häufig vor, dass die Transparenzanforderungen der Datenschutz-Grundverordnung bei KI-Systemen nicht oder nicht vollständig erfüllt werden können.[7] Insbesondere bei nicht-deterministischen Systemen, deren Entscheidungsweise nicht vollständig vorhergesagt werden kann, könne die betroffene Person nicht hinreichend informiert werden. Diese pauschale Behauptung, KI könne nicht transparent gemacht werden, weil das Ergebnis der KI-Anwendung nicht vorhersehbar sei, überzeugt nicht.

Der Verantwortliche muss vor und während der Entwicklung sowie bei der Anwendung der KI-Systeme sicherstellen, dass er die Anwendung beherrscht. Dies gelingt ihm nur,

- wenn er vorab die Zwecke der Verarbeitung bestimmt,
- sowie im Falle einer Zweckänderung, dies dokumentiert und überprüft, ob der neue Zweck mit dem ursprünglich festgelegten Zweck vereinbar ist und
- die Risiken für die Freiheiten und Rechte natürlicher Personen bei der Verarbeitung ermittelt die Ergebnisse kontinuierlich überprüft, insbesondere im Hinblick auf eine Diskriminierung der betroffenen Person.

Kann der Verantwortliche diese Mindestanforderungen erfüllen, so wird es ihm auch gelingen die betroffene Person transparent über die KI-Anwendung zu informieren, über die involvierte Logik hinreichend aufzuklären und im Falle einer Einwilligung der betroffenen Person die Reichweite ihrer Entscheidung verdeutlichen.

[7] Vgl. u. a. Stefan Krempel, DSGVO und KI: Unverträglichkeiten beim Datenschutz, abrufbar unter: https://www.heise.de/newsticker/meldung/DSGVO-und-KI-Unvertraeglichkeiten-beim-Datenschutz-4049785.html

III. Rolle der Beteiligten

Bisher wenig beleuchtet ist die Rolle der einzelnen Akteure, die an der Entwicklung und dem Einsatz von KI-Systemen beteiligt sind. KI-Systeme werden häufig unter der Federführung mehrerer beteiligter Unternehmen, Forschungseinrichtungen, Behörden oder sonstigen Stellen entwickelt und eingesetzt. Immer dann, wenn mehrere Akteure an einer Datenverarbeitung beteiligt sind, ist im Vorfeld zu klären, wem welche datenschutzrechtliche Rolle zukommt. Ein an der Datenverarbeitung Beteiligter ist entweder Verantwortlicher, Auftragsverarbeiter oder Dritter. Bei datenbasierten Technologien verarbeitet im Regelfall nicht nur ein Verantwortlicher personenbezogene Daten, sondern die Verarbeitung wird gemeinsam mit einem oder mehreren anderen Verantwortlichen durchführt. Im Falle einer solchen gemeinsamen Verantwortlichkeit gilt es weitere Anforderungen gemäß Art. 26 des GVO zu beachten.[8] Bestimmen mehrere Verantwortliche gemeinsam die Zwecke und (wesentlichen) Mittel der Verarbeitung, müssen sie

- eine Vereinbarung schließen, in der sie verbindlich festlegen, wer und in welchem Umfang die datenschutzrechtlichen Anforderungen bei der gemeinsamen Verarbeitung umsetzt. Dies gilt insbesondere für die Frage, wer die Betroffenen Personen informiert und Betroffenenrechte gewährleistet.
- Das Wesentliche dieser Vereinbarung muss den betroffenen Personen zur Verfügung gestellt werden, zum Beispiel als Ergänzung in den Datenschutzbestimmungen.
- Schließlich müssen die Akteure sicherstellen, dass sich der Betroffene an jeden der gemeinsamen Verantwortlichen wenden kann, ungeachtet dessen, welche konkreten Vereinbarungen die gemeinsam Verantwortlichen im Hinblick auf die Betroffenen Rechte vereinbart haben.

Zusätzlich ist zu berücksichtigen, dass Dienstleister, die den Verantwortlichen beispielsweise bei der technischen Umsetzung unterstützen, häufig in der datenschutzrechtlichen Rolle als Auftragsverarbeiter tätig sind. Auch in diesem Fall gilt es zusätzliche datenschutzrechtliche Anforderungen zu beachten, die sich aus Art. 28 DSGVO ergeben.[9]

Verantwortliche sollten bereits während der Entwicklung von KI-Systemen die Rolle der Beteiligten verbindlich klären. Nur so lassen sich während der laufenden Entwicklung Fragen zu Berechtigungen, Rechtsgrundlagen, sowie technisch organisatorischen Maßnahmen klären und in laufende Prozesse umsetzen. Dieses vor allem dann relevant, wenn personenbezogene Daten an Dritte übermittelt werden. In diesem Fall muss vorab geprüft werden, auf welcher Rechtsgrundlage dies erfolgt und ob gegebenenfalls eine Einwilligung der betroffenen Person eingeholt werden muss.

Ein praktisches Anwendungsszenario, welches die Rolle der unterschiedlichen Akteure aufzeigt, ist das gemeinsame Pilotprojekt „Sicherheitsbahnhof Berlin Südkreuz" des Bundesministeriums des Innern, für Bau und Heimat sowie der Bundespolizei und der Deutsche Bahn AG. In

[8] Muster zur gemeinsamen Verantwortlichkeit des Landesbeauftragten für Datenschutz und Informationsfreiheit Baden-Württemberg, abrufbar unter: https://www.baden-wuerttemberg.datenschutz.de/mehr-licht-gemeinsame-verantwortlichkeit-sinnvoll-gestalten/

[9] Muster eines Vertrags zur Auftragsverarbeitung des Bayerischen Landesamts für Datenschutzaufsicht, abrufbar unter: https://www.lda.bayern.de/de/thema_auftragsverarbeitung.html

dem Pilotprojekt wurde der Nutzen von geometrischer Gesichtserkennungstechnik in Live-Videoströmen der Überwachungskameras der Deutschen Bahn AG für polizeiliche Zwecke getestet. Die an dem Pilotprojekt beteiligten Akteure erarbeiteten im Vorfeld ein Datenschutzkonzept, in welchem unter anderem Fragen zur Transparenz, Rechtmäßigkeit sowie zur datenschutzrechtliche Rolle der Beteiligten bewertet wurden. So konnten im Vorfeld Verantwortlichkeiten festgelegt und erforderliche Verträge zur Auftragsverarbeitung mit Dienstleistern geschlossen werden.[10]

IV. Betroffenenrechte

Betroffene Personen sind nicht nur transparent über die Datenverarbeitung zu informieren. Der Verantwortliche muss auch weitere Betroffenenrechte gewährleisten können. Dazu zählen:

- das Recht auf Widerruf einer Einwilligung, Art. 7 DSGVO
- das Recht auf Auskunft, Art. 15 DSGV
- das Recht auf Berichtigung, Art. 16 DSGV
- das Recht auf Löschung, Art. 17 DSGV
- das Recht auf Einschränkung der Verarbeitung, Art. 18
- das Recht auf Datenübertragbarkeit, Art. 20 DSGVO und
- das Recht auf Widerspruch, Art. 21 des GVO

Verantwortliche müssen innerhalb eines Monats die Anfrage des Betroffenen beantworten. Bei umfangreichen Datenverarbeitung wie Big Data oder KI- Anwendungen muss der Verantwortliche im Vorfeld prüfen, welche Daten Betroffenenrechten unterliegen. Darüber hinaus muss er einen Prozess entwickeln, um fristgerecht die Betroffenenanfrage zu beantworten und gegebenenfalls die Daten zu löschen, zu berichtigen, zu beauskunften oder die Verarbeitung künftig zu unterlassen, wenn der Betroffene seine Einwilligung widerrufen oder der Datenverarbeitung widersprochen hat. Diese Pflichten kann der Verantwortliche nur erfüllen, wenn er dies bei der Entwicklung der KI-Systeme berücksichtigt hat. Ist das KI-System bereits im Einsatz wird es häufig nur mit enormem finanziellen, technischen und zeitlichen Aufwand möglich sein, fehlende Prozesse zur Gewährleistung der Betroffenenrechte nachträglich zu implementieren. In einem solchen Fall kann sich der Verantwortliche nicht darauf berufen, die Umsetzung der Betroffenenrechte sei unverhältnismäßig und könne daher nicht erfolgen.

Verantwortliche, die Betroffenenrechte bereits bei der Entwicklung unberücksichtigt lassen, setzen sich nicht nur einem Bußgeldverfahren aus, sondern riskieren, dass die gesamte Verarbeitung rechtswidrig ist.

Aus diesen Gründen ist es besonders empfehlenswert, möglichst zeitnah mit der zuständigen Aufsichtsbehörde in Kontakt zu treten und zu prüfen, wie Betroffenenrechte bei der Entwicklung von KI-Systemen sichergestellt werden können und welche Fallstricke es zu vermeiden gilt.

V. Rechtmäßigkeit

Die Datenschutz-Grundverordnung enthält in Art. 6 DSGVO Voraussetzung für die rechtmä-

[10] Abschlussbericht des Bundespolizeipräsidiums zur biometrischen Gesichtserkennung (Erprobung von Systemen zur intelligenten Videoanalyse), abrufbar unter: https://www.bundespolizei.de/Web/DE/04Aktuelles/01Meldungen/2018/10/181011_abschlussbericht_gesichtserkennung_down.pdf;jsessionid=AF93954429C2014462278C3075D483DA.2_cid324?__blob=publicationFile&v=1

ßige Verarbeitung. Darüber hinaus ergeben sich in bestimmten Fällen aus spezialgesetzlichen Regelungen weitere Ermächtigungen, wie zum Beispiel aus dem Sozialrecht, dem Versicherungsrecht oder dem Steuerrecht.

Die Datenverarbeitung darf erst beginnen, wenn der Verantwortliche dies auf eine Rechtsgrundlage oder die Einwilligung der betroffenen Person stützen kann. Dies ist nicht erst dann relevant, wenn KI-Systeme angewendet werden, sondern bereits bei der Entwicklung, wenn beispielsweise Rohdaten zu Trainingsdaten verarbeitet werden.[11] unzulässig ist es, mit der Verarbeitung zu beginnen und nachträglich die Einwilligung der betroffenen Person einzuholen.

Die Datenschutz-Grundverordnung kennt keine „Heilung" einer datenschutzwidrigen Verarbeitung. Dies gilt auch, wenn bei der Anwendung von KI-Systemen eine Änderung der Verarbeitungszwecke eintritt. Kommt der Verantwortliche zu dem Ergebnis, dass der neue Zweck nicht mit dem ursprünglichen Verarbeitungszweck vereinbar ist im Sinne des Art. 6 Abs. 4 DSGVO, so ist die Verarbeitung rechtswidrig. Um dies zu vermeiden, muss der Verantwortliche von Beginn an sämtliche Risiken für die Rechte und Freiheiten der Betroffenen Personen ermitteln und während des gesamten Lebenszyklus von KI-Systemen die Einhaltung der datenschutzrechtlichen Anforderungen validieren.

VI. Fazit

Die Anforderungen der Datenschutz-Grundverordnung dürfen nicht vorschnell als Hindernis oder Hemmschwelle für technische Innovationen verurteilt werden. Bei der Entwicklung und Anwendung von KI-Systemen ist der Datenschutz nur ein Rechtsbereich, der sicherstellen soll, dass KI-Systeme in einer freiheitlich demokratischen Gesellschaft die Rechte und Freiheiten der Bürger gewährleisten. Ziel des Datenschutzes ist es, Risiken, die zweifelsohne immer mit Innovationen einhergehen auf ein Mindestmaß zu reduzieren. Die Datenschutzaufsichtsbehörden tragen hierzu einen wesentlichen Teil bei – und zwar nicht als Bußgeld- oder Vollzugsbehörde, sondern in ihrer beratenden Funktion. Verantwortliche sollten sich nicht scheuen, aktiv auf die Aufsichtsbehörden zuzugehen und die von der Datenschutz-Grundverordnung festgelegte Aufgabe der Beratung einfordern. Hiervon profitieren am Ende nicht nur die Verantwortlichen, sondern auch die Aufsichtsbehörden, denn ein fachkundiger Austausch ist unabdingbar, um abstraktes Recht mit technischen Innovationen in Einklang zu bringen.

[11] O. o. A. DSK, Positionspapier der DSK zu empfohlenen technischen und organisatorischen Maßnahmen bei der Entwicklung und dem Betrieb von KI-Systemen, S. 7.

C Data Science und Machine Learning in Unternehmen am Beispiel der TeamBank AG

I. Zielsetzung

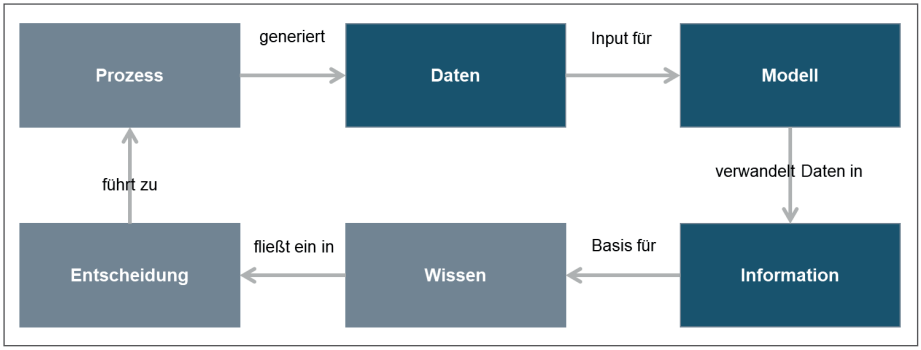
Dieses Kapitel führt in die Nutzung von Data Science und Machine Learning in Unternehmen ein. Neben theoretischen Begriffsdefinitionen verdeutlichen wir die wichtigsten Konzepte anhand eines realen Anwendungsfalls bei der TeamBank AG. Der Abschnitt bildet die Grundlage für die weiteren Kapitel dieses Buchs.

II. Grundbegriffe

Um uns dem Thema Data Science zu nähern, müssen wir uns mit den Begriffen *Prozess* und *Daten* beschäftigen. In ihrem Buch „Doing Data Science" erklären Cathy O'Neil und Rachel Schutt, dass jedes Phänomen in eine dieser beiden Kategorien eingeordnet werden kann.[12] Ein Prozess ist ein Vorgang, der die Welt verändert. Steigen wir in unser Auto und fahren zur Arbeit, bestellen wir ein Produkt im Internet oder schließen wir einen Kreditvertrag ab, hat die Welt nach dem Vorgang einen anderen Zustand als vorher. Prozesse hinterlassen dabei Spuren, sogenannte Daten. Nach einer Autofahrt verfügt der Anbieter unseres Navigationssystems über unsere GPS-Daten, der Händler nach Bestellung über einen Auftrag und eine Bank nach Vertragsschluss über Kunden- und Finanzdaten.

Daten haben für sich alleine genommen keinen intrinsischen Wert. Dieser entsteht erst durch

Abbildung 1: Prozesse, Daten und Information[13]



[12] Vgl. (O'Neil & Schutt, 2013, S. 18–19).

[13] Eigene Darstellung in Anlehnung an (O'Neil & Schutt, 2013) und (Provost & Fawcett, 2013).

ihre Umwandlung in Information und die anschließende Nutzung. Eine Information hat eine für einen Menschen oder eine Maschine interpretierbare Form. Es kann sich um eine Kennzahl handeln, doch auch Visualisierungen erfüllen diesen Zweck. Für eine Autofahrt lässt sich beispielsweise die Durchschnittsgeschwindigkeit berechnen oder eine Route zeichnen. Händler orientieren sich an ihrer Konversionsrate und Banken an ihrer Aufwand-Ertrag-Relation und der Zinsentwicklung.

Die Transformation von Daten zu Information wird mit Hilfe von *Modellen* realisiert. Ein Modell kann verschiedene Formen annehmen. Von einfachen mathematischen Formeln bis zu komplexen Algorithmen finden sich verschiedene Varianten, von denen wir im Folgenden einige vorstellen.

Sobald ein Mensch eine Information aufgenommen hat – sie sich in seinem „Kopf" befindet – sprechen wir von Wissen. Dieses ist Grundlage für zu treffende Entscheidungen, die wiederum in Prozessen resultieren. Durch das Zusammenspiel entsteht ein Zyklus aus Prozessausführung, Informationsverarbeitung und Entscheidung. Dieser ist in Abbildung 1 dargestellt. Die Datenanalyse umfasst dabei die Sammlung von Daten und deren Umwandlung in Information.

1. Business Intelligence & Data Science

In Unternehmen finden wir zwei grundsätzliche Formen der Datenanalyse:

- Business Intelligence
- Data Science

Wie der Begriff nahelegt, beschäftigt sich *Business Intelligence* (kurz: BI) mit der Sammlung und Auswertung von Daten in Unternehmen.[14] Hauptfokus ist nach unserer Definition die direkte Nutzung durch den Menschen. Im Rahmen des sogenannten Reporting werden Informationen für Entscheider in Berichten zusammengefasst und aufbereitet. Ziel von Business Intelligence ist es, Menschen in Unternehmen mit Informationen zu versorgen und ihnen zu helfen fundierte Entscheidungen zu treffen.

Data Science hat ebenfalls zum Ziel, Daten in Informationen zu verwandeln. Die hierbei eingesetzten Methoden sind jedoch technischer, als dies bei Business Intelligence der Fall ist. Ein Data Scientist setzt bei seiner Arbeit fortgeschrittene wissenschaftliche Methoden aus den Bereichen Mathematik, Statistik und Informatik ein. Fokus liegt auf der Modellgenerierung, Vorhersagen und maschinellem Lernen. Neben Entscheidern gehören Maschinen ebenfalls zur Zielgruppe eines Data Scientists. Die angewendeten Techniken kommen auch außerhalb der Wirtschaft zum Einsatz, beispielsweise in der Medizin oder der Politik.[15]

Eine bekannte Darstellung des Profils eines Data Scientists bietet das Venn-Diagramm von Conway in Abbildung 2.[16] Neben Fähigkeiten aus den Bereichen Mathematik und Statistik verfügt er demnach über Hacking Skills (das heißt Techniken aus der Informatik) sowie Fachexpertise der jeweiligen Branche. Branchenwissen grenzt nach Conway die Bereiche Data Science und reines Machine Learning (kurz: ML) voneinander ab. Ein Data Scientist ist durch seine

[14] Eine Einführung in das Thema bietet (Howson, 2013).

[15] Zum Thema Data Science in Unternehmen siehe (Provost & Fawcett, 2013).

[16] Quelle: (Conway, 2010).