



Felix Weber

Künstliche Intelligenz für Business Analytics

Algorithmen, Plattformen
und Anwendungsszenarien

EBOOK INSIDE

 Springer Vieweg

Künstliche Intelligenz für Business Analytics

Felix Weber

Künstliche Intelligenz für Business Analytics

Algorithmen, Plattformen und
Anwendungsszenarien

Felix Weber
Lehrstuhl für Wirtschaftsinformatik
Universität Duisburg-Essen
Essen, Deutschland

ISBN 978-3-658-29772-5 ISBN 978-3-658-29773-2 (eBook)
<https://doi.org/10.1007/978-3-658-29773-2>

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

Springer Vieweg

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2020

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung: Sybille Thelen

Springer Vieweg ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Vorwort

Mit einem rein akademischen Hintergrund überrascht es doch viele Neueinsteiger, was die Leute in der Praxis Business Intelligence (BI) und Business Analytics (BA) nennen. Der erste naive Gedanke von Hochschulabsolventen ist sicherlich, dass wirklich komplexe Künstliche Intelligenz (KI) und fortschrittliche Modelle des Maschinellen Lernens in den international aufgestellten deutschen Unternehmen angewandt werden. Wie sonst sollte man auch konkurrenzfähig bleiben, wenn man nicht schon vor dem Konsumenten weiß, welche Farbe des Neuwagens oder Handymodells in der Zukunft gewünscht wird und dies dann situationsabhängig über alle denkmöglichen Werbekanäle ausspielt.

Aber nach kurzer Untersuchung war ich sehr überrascht, als ich feststellte, dass das meiste davon nur einfache deskriptive Statistiken sind.¹ Einer der größten deutschen Einzelhändler erfasst die Umsatzberichte nur wöchentlich und aggregiert diese nicht einmal aus dem eingesetzten ERP-System, sondern veranlasst vielmehr die einzelnen Händler, die Zahlen manuell zusammenzustellen und ins hauseigene Informationssystem einzutragen. Ein anderer beschäftigt sich gerade jetzt erst mit grundlegenden Analysen des Geschäft's basierend auf einfachen Kennzahlen wie beispielsweise der Zusammensetzung der Kassenbons und des Werbeanteils, also der Anzahl der beworbenen Produkte in Relation zu nicht beworbenen, je Bon. Wahrscheinlich ist dieses plötzliche Interesse auf die aufkommende Konkurrenz durch Onlinehändler wie Amazon, der erst 2017 mit Amazon Fresh in den deutschen Lebensmittelhandel eingestiegen ist, zurückzuführen.²

Ein Großteil der Analysen im Geschäftsumfeld sind deskriptive Analysen. Sie berechnen deskriptive Statistiken (d. h. Zähler, Summen, Durchschnitte, Prozentsätze, Minimum,

¹ Diese Feststellung gilt allerdings für viele der aus der idealisierten Welt der Forschung stammenden Ideen, Konzepte und Handlungsempfehlungen. Die Systemarchitekturen sind monolithischen Mainframe-Architekturen ähnlicher als dem State of the Art von verteilten Service-orientierten Architekturen – oder die Planung und Durchführung von Projekten ignoriert die letzten Jahrzehnte der Forschung im Bereich Projektmanagement und nutzt stattdessen, wenn überhaupt, eine Microsoft Excel-basierte „Planung“.

² Ehrlicherweise muss man auch sagen, dass der Erfolg der deutschen Einzelhändler in der Vergangenheit bisher auch keine tiefere Auseinandersetzung mit diesen Themen erforderlich gemacht hätte.

Maximum und einfache Arithmetik), welche bestimmte Gruppierungen oder gefilterte Versionen der Daten zusammenfassen, die typischerweise einfache Zählungen einiger Ereignisse sind. Diese Analysen basieren meist auf Standard-Aggregatfunktionen in Datenbanken, die nichts anderes erfordern als Mathematik in der Grundschule. Selbst grundlegende Statistiken (z. B. Standardabweichungen, Varianz, p-Wert etc.) sind recht selten. Der Zweck der deskriptiven Analytik ist es, einfach zusammenzufassen und Ihnen zu sagen, was passiert ist: Zum Beispiel den Umsatz, die Anzahl der Kunden, der Anteil der beworbenen Produkte vom Gesamtumsatz, Seitenaufrufe etc. Es gibt buchstäblich Tausende dieser Metriken – es ist sinnlos, sie aufzulisten –, aber sie sind alle nur einfache Ereigniszähler. Andere deskriptive Analysen können Ergebnisse einfacher arithmetischer Operationen sein, wie z. B. Stimmanteil, durchschnittliche Antwortzeit, %-Index, durchschnittliche Anzahl der Antworten pro Beitrag etc. Dies alles findet heute in einem Großteil der Unternehmen statt und wird zumeist als Business Intelligence bezeichnet. Meistens wird unter dem Stichwort Advanced Analytics die Erweiterung dieses Reporting um einige Filter auf den Daten, bevor die deskriptiven Statistiken berechnet werden, verstanden. Wenn Sie beispielsweise für Social-Media-Analysen zuerst einen Geofilter anwenden, können Sie Kennzahlen wie den durchschnittlichen Post pro Woche aus Deutschland und den durchschnittlichen Post pro Woche aus den Niederlanden erhalten. Und Sie können diese Daten auf einer schicken Karte für alle Länder, in denen Sie aktiv sind, anzeigen. Dann erhalten Sie plötzlich eine erweiterte Analytik, benutzen aber eigentlich auch nur Mathematik für die Grundschule.

Für einen Wettbewerbsvorteil gegenüber den Wettbewerbern reicht diese rudimentäre Analytik allerdings nicht. Vor allem, wenn Sie plötzlich mit Digital Natives wie Google, Amazon oder Alibaba konkurrieren müssen. Im Zeitalter der Digitalisierung ist dies aber plötzlich die Herausforderung für viele Branchen. Amazon hat den Bücherhandel und danach den Handel an sich umgekrempelt. Google steigt plötzlich mit selbstfahrenden Autos in den Automobilmarkt ein, Uber degradiert Branchengrößen in der Automobilindustrie (Volvo und Toyota) zu reinen Zulieferern und Airbnb übernimmt ohne eigene Immobilien einen großen Marktanteil in der Hotelindustrie. So unterschiedlich diese Beispiele auch sind, so basieren diese nicht nur auf Software und Plattformen, sondern vor allem einer ausgereiften Analytik. Uber hat eine riesige Datenbank von Fahrern, sodass, sobald Sie ein Auto anfordern, Ubers Algorithmus sofort einsatzbereit ist – in 15 Sekunden oder weniger passt er Sie mit dem Fahrer in Ihrer Nähe zusammen. Im Hintergrund speichert Uber Daten über jede Fahrt – auch wenn der Fahrer keine Fahrgäste hat. Alle diese Daten werden gespeichert und genutzt, um Angebot und Nachfrage vorherzusagen und die Tarife festzulegen. Uber untersucht auch, wie der Transport zwischen den Städten abgewickelt wird, und versucht, sich auf Engpässe und andere häufige Probleme einzustellen.

Zum Ziel des Buchs

Das Ziel dieses Buches soll nicht die Ausbildung zum Data Scientist oder Data Analyst sein, auch wird sich niemand nach der Lektüre als Experte für Künstliche Intelligenz oder

Maschinelles Lernen bezeichnen können – auch wenn das der eine oder andere Unternehmensberater tun wird. Vielmehr führt das Buch in komprimierter Form in die essenziellen Aspekte der Business Analytics und den Einsatz von Methoden der Künstlichen Intelligenz dort ein. Zunächst einmal werden die grundlegenden Begriffe und Denkmuster der Analytics der deskriptiven, prädiktiven bis zur präskriptiven Analytik vorgestellt Abschn. 1.3. Es folgt mit dem Business Analytics Model for Artificial Intelligence (BAM. AI) ein Vorgehensmodell zur Umsetzung von Business-Analytics-Projekten Abschn. 1.5 und ein Technologieframework inklusive der Vorstellung der wichtigsten Frameworks, Programmiersprachen und Architekturen Kap. 3. Nach einer Einführung in die Künstliche Intelligenz Kap. 2 und vor allem den Teilbereich des Maschinellen Lernens werden die wichtigsten Problemkategorien beschrieben und die einsetzbaren Algorithmen grob, aber in verständlicher Form vorgestellt Abschn. 2.2. Danach folgt eine detaillierte Übersicht über die gängigen Cloud-Plattformen Abschn. 3.2, die eine schnelle Umsetzung eines BA-Projektes ermöglichen. Hier wird dem Leser ein Leitfaden an die Hand gegeben, der ihm eine Übersicht über das umfangreiche Angebot der großen Anbieter ermöglicht. Abschließend zeigen mehrere Anwendungsszenarien aus verschiedenen Perspektiven die Einsatzmöglichkeit von KI und BA in verschiedenen Branchen als Fallstudien auf Abschn. 3.3.

Da sich das Buch definitiv als Einstiegs- und Übersichtswerk für Entscheider und Umsetzer in der IT und den Einsatzdomänen sieht, wird an vielen Stellen auf tiefergehende Literatur verweisen.

Essen, Deutschland

Felix Weber

Inhaltsverzeichnis

1 Business Analytics und Analytics	1
1.1 Notwendigkeit einer zunehmenden analytischen Entscheidungsunterstützung	1
1.2 Abgrenzung zwischen Business Intelligence und Business Analytics	6
1.3 Kategorisierung von analytischen Methoden und Modellen	8
1.3.1 Deskriptive Analytik	8
1.3.2 Prädiktive Analytik	9
1.3.3 Präskriptive Analytik	11
1.4 Business Analytics Technologieframework (BA.TF)	12
1.4.1 Datenquellen	14
1.4.2 Data Preparation	15
1.4.3 Datenspeicherung	16
1.4.4 Analyse	18
1.4.5 Zugriff und Nutzung	19
1.4.6 (Big)-Data Management and Governance	20
1.5 Vorgehensmodell: Business Analytics Model for Artificial Intelligence (BAM.AI)	22
1.5.1 Development Cycle	24
1.5.1.1 Business Understanding	24
1.5.1.2 Data Discovery	26
1.5.1.3 Data Wrangling	26
1.5.1.4 Analyse	27
1.5.1.5 Validierung	28
1.5.1.6 New Data Aquisition	29
1.5.2 Deployment Cycle	29
1.5.2.1 Publish	30
1.5.2.2 Analytic Deployment	31
1.5.2.3 Application Integration	33

1.5.2.4	Test	34
1.5.2.5	Production/Operations	34
1.5.2.6	Continuous Improvement	35
	Literatur	35
2	Künstliche Intelligenz	37
2.1	Maschinelles Lernen	39
2.1.1	Überwachtes Lernen/Supervised Learning	40
2.1.2	Unüberwachtes Lernen/Unsupervised Learning	42
2.1.3	Bestärkendes Lernen/Reinforcement Learning	43
2.1.4	Übersicht über die Arten des Maschinellen Lernens	44
2.1.5	Neuronale Netze	44
2.2	Problemtypen der Künstlichen Intelligenz und deren Algorithmen	47
2.2.1	Klassifizierung	47
2.2.2	Abhängigkeiten und Assoziationen	50
2.2.3	Clustering	54
2.2.4	Regression, Prediction oder Vorhersage	56
2.2.5	Optimierung	59
2.2.6	Erkennung von Anomalien (Outliner)	61
2.2.7	Empfehlung oder Recommender-Systems	63
2.2.8	Wann welchen Algorithmus nutzen?	66
	Literatur	71
3	KI- und BA-Plattformen	73
3.1	Grundbegriffe und Softwareframeworks	73
3.1.1	Datenhaltung	73
3.1.1.1	Datawarehouse	73
3.1.1.2	Data Lake	74
3.1.1.3	Data Streaming und Message Queuing	76
3.1.1.4	Datenbankmanagementsystem	78
3.1.1.5	Apache Hadoop	79
3.1.2	Datenanalyse und Programmiersprachen	82
3.1.2.1	Python	82
3.1.2.2	R	83
3.1.2.3	SQL	84
3.1.2.4	Scala	85
3.1.2.5	Julia	85
3.1.3	KI-Frameworks	86
3.1.3.1	Tensorflow	87
3.1.3.2	Theano	88
3.1.3.3	Torch	88
3.1.3.4	Scikit-learn	89
3.1.3.5	Jupyter Notebook	90

3.2	Business Analytics und Machine Learning as a Service (Cloud-Plattformen)	91
3.2.1	Amazon AWS	92
3.2.1.1	Data-Services von Amazon AWS	93
3.2.1.2	ML-Services von Amazon AWS	98
3.2.2	Google Cloud Platform	103
3.2.2.1	Data-Services von Google	104
3.2.2.2	ML-Services von Google	107
3.2.3	IBM Watson	108
3.2.4	Microsoft Azure	109
3.2.4.1	Data-Services von Microsoft Azure	109
3.2.4.2	ML-Services von Microsoft Azure	111
3.2.4.3	Weitere Services von Microsoft Azure in der Übersicht	112
3.2.5	SAP Services und SAP HANA Cloud Plattform (SCP)	112
3.2.5.1	Data-Services von SAP	115
3.2.5.2	ML-Services von SAP	118
3.2.5.3	SAP-HANA-Datenbankplattform	122
3.3	Build or Buy?	123
	Literatur	125
4	Fallstudien zum Einsatz von KI-basierter Business Analytics	127
4.1	Fallstudie: Kundenstimmung in Echtzeit analysieren mit Streaming Analytics	127
4.1.1	Kundenzufriedenheit im Einzelhandel	127
4.1.2	Mit Technologieakzeptanz und Omnichannel zu mehr Daten	128
4.1.3	Customer Satisfaction Streaming Index (CSSI)	130
4.1.4	Implementierung in einer Retail-Architektur	131
4.1.5	Ergebnisse	133
4.2	Fallstudie: Marktsegmentierung und -automatisierung im Einzelhandel mit neuronalen Netzen	136
4.2.1	Die Standortentscheidung im stationären Handel	137
4.2.2	Marketing-Segmentierung und Einzugsgebiet	138
4.2.3	Klassische Clustering-Ansätze und Growing Neural Gas	139
4.2.4	Projektaufbau	142
4.2.5	Die Daten und Quellen	143
4.2.6	Implementierung	147
4.3	Ergebnisse	148
	Literatur	151
	Stichwortverzeichnis	155



1.1 Notwendigkeit einer zunehmenden analytischen Entscheidungsunterstützung

Die Globalisierung, eine möglicherweise entstehende Ressourcenknappheit, deutlich erhöhte Komplexität der Märkte und der Aufstieg der BRICS-Länder sind die größten Herausforderungen für die führenden Industrieländer der letzten Jahre. Für diese Nationen und die dort beheimateten Unternehmen besteht die Hauptaufgabe für die nächsten Dekaden darin, die vorhandenen Produktionskapazitäten wesentlich effizienter auszunutzen und ein Umfeld für hoch entwickelte Industrieprodukte zu gewährleisten. Um diesen Herausforderungen gerecht zu werden, konzentriert man sich vornehmlich auf Subventionspolitik und Forschungsaktivitäten zu komplexen Konzepten wie der „Digitalen Fabrik“ [1], „Industrie 4.0“ [2] oder generell „Intelligente Produktionssysteme“ [3]. Neben dieser großen Veränderung liegt ein weiterer Fokus auf der Einführung einer Vielzahl von Systemen zur Steuerung, Optimierung und Kontrolle der bestehenden Betriebsprozesse. Das Hauptziel dieser Maßnahmen ist dabei, die vollständige Digitalisierung¹ und Integration aller Prozesse des Produktlebenszyklus inklusive der Lieferketten anzustreben.

Die Analytik ist dabei zum Technologietreiber dieses Jahrzehnts geworden. Unternehmen wie IBM, Oracle, Microsoft und andere haben ganze Organisationseinheiten geschaffen, die sich nur auf Analytik konzentrieren und die zu beratenden Unternehmen dabei unterstützen, effektiver und effizienter zu arbeiten. Entscheidungsträger nutzen mehr computergestützte Tools, um ihre Arbeit zu unterstützen. Ganze Bereiche im operativen Management und der Verwaltung könnten durch automatisierte analytische Systeme ersetzt werden. Und selbst die Verbraucher nutzen Analysetools direkt oder indirekt, um Entscheidungen über Routineaktivitäten wie Einkaufsentscheidungen (Stichwort:

¹ Im Sinne einer digitalen Abbildung der Realität (Stichwort „Digital Twin“).

Preisvergleichstools), Gesundheitswesen (Stichwort Health Apps) und Unterhaltung (das angesprochene Beispiel von Netflix) zu treffen. Business Analytics entwickelt sich rasant weiter und konzentriert sich zunehmend auf innovative Anwendungen von Daten, die vor einiger Zeit noch nicht einmal erfasst, geschweige denn signifikant analysiert wurden.

Was Unternehmen in diesem Zusammenhang brauchen, ist eine Möglichkeit, dass die richtigen Personen die richtigen Daten und Informationen zur richtigen Zeit verfügbar haben und somit eine Grundlage für rationale Entscheidungsfindung gewinnen, die den strategischen und operativen Marktbedingungen entspricht. Und das ist genau der Teil der Anforderungen und Leitlinien an Analytik in diesem Buch:

- ▶ Anforderung: Die richtigen Daten für die richtigen Personen zur richtigen Zeit zur Entscheidungsunterstützung bereitstellen.

In dieser Anforderung wurde bewusst der Begriff Entscheidungsunterstützung gewählt, da dem Anwender Daten, Informationen oder Wissen zu einem bestimmten Zweck zur Verfügung gestellt werden: zur Erleichterung der Entscheidungen, die so oder so getroffen werden müssen. Hier wäre das markante Beispiel die Analyse der Kassensbons im Supermarkt und die Kenntnis darüber, welche Produkte im Verlauf eines Jahres wie oft verkauft werden. Die Kenntnis darüber ermöglicht es den Verantwortlichen, beispielsweise dem Filialleiter, seine Entscheidungen über Regalplatzierungen, Nachbestellungen oder Preisänderungen zu unterstützen. Diese Entscheidungen müssen so oder so getroffen werden, nur ermöglicht es Analytik, dass diese Entscheidungen nicht nur vom reinen „Bauchgefühl“ oder der jahrelangen Erfahrung abhängig sind.

Unabhängig davon, ob Prognosemodelle verwendet werden, enthalten die historischen Daten eines Unternehmens immerhin einen Hinweis darauf, warum das Unternehmen in der jetzigen Situation ist, da diese Daten die vergangenen Situationen und Entscheidungen abbilden. Aufgrund der technischen Nähe und Komplexität wird Analytik meistens als reine IT-Disziplin betrachtet, die in erster Linie durch das technische Umfeld des Unternehmens getrieben wird. Diese Verortung vernachlässigt allerdings immer das notwendige Wissen der Domäne. So wird eine Analytik ohne Berücksichtigung der Organisation (Mission, Vision, Strategie und Ziele) und der genauen Kenntnis der realen Geschäftsprozesse, die meistens nicht dokumentiert oder in IT-Systemen abgebildet sind, kaum die optimale Lösung darstellen. Betrachtet man nur das zuvor genannte Beispiel des Filialleiters, so lassen sich schnell eine Fülle von Einflussfaktoren ableiten, die ein reiner IT-Fokus – auch aufgrund nicht erfasster Daten – vernachlässigt hätte: Demografie, sozio-ökonomisches Umfeld (lokal und gesamtwirtschaftlich), Kundenstimmungen oder gar die Eigenheiten der Konsumenten vor Ort (bekannt dem Filialleiter durch seinen täglichen Umgang, aber kaum in einem IT-System abgebildet).

Ein gewisses Grundgerüst an Analytik war und ist eigentlich in jedem Unternehmen vorhanden und seien es nur die allgegenwärtigen Excel-Dateien. In den letzten Jahren haben die zugrunde liegenden IT-Systeme jedoch einige wichtige Entwicklungen erfahren.

Eine große Änderung wird als „Big Data“ bezeichnet. Dabei ist die Größe der Daten² die erste und manchmal die einzige Dimension, die bei der Erwähnung großer Daten hervortritt. An dieser Stelle wollen wir nicht zu tief in die Herkunft und Hintergründe des Begriffs abtauchen und dokumentieren nur das grundlegende Konzept im Zusammenhang mit großen Datenmengen. Bereits 2001 schlug Laney [4] vor, dass Volume, Variety und Velocity (oder die drei Vs) die drei Dimensionen der Herausforderungen im Datenmanagement sind. Die 3Vs haben sich seitdem zu einem gemeinsamen Rahmen für die Beschreibung von Big Data entwickelt. Auch definiert Gartner [5] den Begriff in ähnlicher Weise:

► „Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enables enhanced insight, decision making, and process automation“.

Das **Volumen** bezieht sich auf die Größe der Daten. Große Datenmengen werden in mehreren Terabyte und Petabyte angegeben. Ein Terabyte speichert so viele Daten wie auf 1500 CDs oder 220 DVDs passen würden, genug, um etwa 16 Millionen Facebook-Fotos zu speichern. Facebook verarbeitet bis zu einer Million Fotos pro Sekunde [6]. Ein Petabyte entspricht dabei 1024 Terabyte. Die Definitionen von Big Data hängen aber von der Branche und dem Typ der Daten ab und erlauben es nicht einfach, einen bestimmten Schwellenwert für Big Data zu definieren. So können zwei Datensätze gleicher Größe je nach Typ (Tabellen- vs. Videodaten) unterschiedliche Technologien zur Verarbeitung erfordern.

Die **Variety** bezieht sich auf die strukturelle Heterogenität in einem Datensatz. Moderne Technologien ermöglichen es dabei Unternehmen, verschiedene Arten von strukturierten, halbstrukturierten und unstrukturierten Daten zu verwenden. Strukturierte Daten beziehen sich auf die tabellarischen Daten in Tabellenkalkulationen oder relationalen Datenbanken. Text, Bilder, Audio und Video sind Beispiele für unstrukturierte Daten, denen die strukturelle Ordnung fehlt, welche aber von Programmen zur Analyse benötigt wird. Über ein Kontinuum zwischen vollständig strukturierten und unstrukturierten Daten hinweg entspricht das Format der teilstrukturierten Daten nicht den strengen Standards auf der einen noch der anderen Seite. Extensible Markup Language (XML) ist eine textuelle Sprache für den Datenaustausch im Web und ist ein typisches Beispiel für semi-strukturierte Daten. XML-Dokumente enthalten benutzerdefinierte Daten-Tags, die sie maschinenlesbar machen.

Die **Velocity** bezieht sich auf die Geschwindigkeit, mit der Daten erzeugt werden, und die Geschwindigkeit, mit der sie analysiert und bearbeitet werden sollen. Die Verbreitung digitaler Endgeräte wie Smartphones und Sensoren hat zu einer beispiellosen Datenerstellungsrate geführt und führt weiterhin zu einem immer stärker anwachsenden Bedarf an Echtzeitanalysen. Auch konventionelle Einzelhändler generieren hochfrequente Daten, Wal-Mart zum Beispiel verarbeitet mehr als eine Million Transaktionen pro Stunde [7].

²Eigentlich gemeint ist die Masse oder Menge.

Auch neue Technologien, wie In-Memory-Datenbanken (hier befinden sich die Daten dauerhaft im physischen Hauptspeicher (RAM) des Computers ermöglichen es, nicht nur größere Datenmengen zu verarbeiten, sondern sogar in kürzerer Zeit. In einem konventionellen Datenbanksystem sind die Daten plattenresident und die benötigten Daten können für den Zugriff im Hauptspeicher zwischengespeichert und dort bearbeitet werden, wohingegen bei einer In-Memory-Datenbank die Daten speicherresident als Sicherungskopie auf der Festplatte abgelegt werden, ansonsten aber im Hauptspeicher verweilen. In beiden Fällen kann ein bestimmtes Objekt Kopien sowohl im Speicher als auch auf der Festplatte haben. Der Hauptunterschied besteht aber darin, dass bei der In-Memory-Datenbank die Primärkopie dauerhaft im Hauptspeicher verbleibt. Da in den letzten Jahren ein Trend zu beobachten ist, dass der Hauptspeicher immer billiger wird, ist es jetzt schon möglich, immer größere Datenbanken in den Hauptspeicher zu verlagern. Da auf Daten direkt im Speicher zugegriffen werden kann, können wesentlich bessere Antwortzeiten und Transaktionsdurchsätze ermöglicht werden. Dies ist besonders wichtig für Echtzeitanwendungen, bei denen Transaktionen zu den angegebenen Zeitfristen abgeschlossen werden müssen.

Auch ändert sich momentan das vorherrschende Paradigma des Softwarebezugs. Der zunehmende Einsatz von **Cloud-Lösungen** (bei denen Software und Daten nicht am Anwenderstandort gehostet werden), ermöglicht tendenziell eine kürzere Time-to-Market und die Möglichkeit, mit neuen Technologien früher erste Tests und Prototypen zu erstellen.

Die genannten Änderungen in der Verfügbarkeit der Daten, das größere Datenvolumen und die Verfügbarkeit von neuer Software und Softwarebezugsmodellen zur Speicherung und Verarbeitung dienen einem weiteren Trend als Grundlage: dem verstärkten Einsatz von Analysemodellen zur automatisierten Steuerung von ganzen Betriebsprozessen. So ist der entscheidende Schritt, der uns dazu veranlasst, von Business Analytics statt von Business Intelligence zu sprechen (siehe die Ausführung in Abschn. 1.2) der der Überführung der Entscheidungen von Menschen auf die Systeme. Hier sind einige Beispiele:

- Bei **rein digitalen Prozessen** wie dem Omnichannel-Marketing sind heute schon die Entscheidungen auf das System übertragen worden. Die Kundenkommunikation wird direkt vom System an die Kunden versendet, basierend auf der systemischen Einschätzung des individuellen Kunden. Beispielhaft sind hier die Werbe-E-Mails von Amazon oder die Empfehlungen von Netflix. Basierend auf der Datenhistorie des Kunden optimieren Recommender-Systeme die Kommunikation mit dem Kunden. Aber auch der Handel mit Aktien und Währungen ist inzwischen fast vollkommen automatisiert und die Algorithmen der unterschiedlichen Händler arbeiten gegeneinander. Natürlich ist hier der erfolgreichste Investor derjenige, der den besten Algorithmus verwendet.
- **Semiphysisch digitalisierte Prozesse** sind Prozesse, in denen Analytik eingesetzt wird, um beispielsweise die zukünftige Nachfrage vorherzusagen und automatisch die Ware nachzubestellen. Auch in diesem Fall wird der Gewinner im Markt das Unternehmen sein, das die Prozesse mit Hilfe der am besten optimierten Algorithmen ausführt.

Das Internet der Dinge³ ist ein weiterer neuer Begriff, der die Abbildbarkeit von bisher rein physischen Prozessen durch Sensoren und Sensordaten in allen Arten von Alltagsgegenständen beschreibt. So gibt es Milchbauern, die ihre Kühe fast ausschließlich automatisiert von Robotern melken lassen. Menschen werden nur bei Bedarf herangezogen, wie bei festgestellten Krankheiten der Kühe und der notwendigen Behandlung, was nicht durch Maschinen erledigt werden kann. Dazu werden verschiedenste Sensoren aus dem Stall, der Umgebung und an den einzelnen Tieren selber genutzt und ausgewertet.

- **Vollständig digital gesteuerte physikalische Prozesse**, wie der Einsatz von Robotern in der automatisierten Fertigung von Waren oder Autos. Diese Roboter reagieren dabei auf externen physikalischen Input und Algorithmen entscheiden über die notwendige Reaktion. Sie müssen in der Lage sein, auf der Grundlage von Algorithmen autonom zu entscheiden und nutzen dazu Sprach- und Videoerkennung, um die physiologische Umgebung zu verstehen, in der sie agieren.

In den letzten Jahren wurde eine große Vielzahl von Prozessen digital abgebildet, digitalisiert oder komplett automatisiert, und die damit verbundenen manuellen Entscheidungen sind verschwunden. In vielerlei Hinsicht sehen wir heute, was die Menschen während der Dotcom-Ära erwartet haben, in der es um die Möglichkeiten neuer automatisierter und digitalisierter Geschäftsprozesse ging, die es Unternehmen ermöglichten, auf der Grundlage extrem skalierbarer Geschäftsmodelle global zu konkurrieren. Schon damals haben neue Marktteilnehmer wie [Amazon.com](https://www.amazon.com) den Verkauf von Büchern neu definiert, indem man einen rein physikalischen Prozess (der Buchladen um die Ecke) in einen physisch digitalisierten Prozess (Online-Kauf von physischer Ware) umgewandelt hat. Später begannen Apple und Amazon, physische Geräte zu produzieren, um die Möglichkeiten zum Konsum von Inhalten (Büchern, Musik und Filmen) über das Internet und damit den Grad der digitalen Wertschöpfung weiter zu erhöhen. Weniger beachtet von Öffentlichkeit ist die Entwicklung bei den physischen Produktionsprozessen zunehmend digitalisiert worden. In den letzten zehn Jahren wurden immer mehr Geschäftsprozesse so weit digitalisiert, dass der nächste Wettbewerber nur eine App entfernt ist (Apple iBooks vs. Google Play Books). Die marktführende App ist oft diejenige, die auf der führenden Plattform integriert ist, die beste Benutzererfahrung bietet und auf das Individuum optimierte Empfehlungen, basierend auf kundenbezogenen Datenanalysen, enthält.

Da Analytik zunehmend in digitalen Prozessen eingesetzt und diese Prozesse durch die Analytik auch automatisiert werden können, ist Business Analytics heute auch viel mehr als die Entscheidungshilfe für den Menschen innerhalb eines Unternehmens. Es geht auch um die Bereitstellung von Daten, aber vor allem darum, digitalisierte Prozesse intelligent und automatisiert zu steuern. Die Unterstützung für den Menschen rückt dabei deutlich in den Hintergrund. So kommen wir auch zu unserer Definition für Business Analytics:

³Auch Internet of Things (IoT).

Business Analytics bezeichnet (1) Technologien, Systeme, Methoden und Anwendungen, die (2) Daten erfassen, verarbeiten und analysieren, um damit (3) eine Entscheidungsunterstützung oder Automatisierung mit einem Mehrwert für das Unternehmen zu ermöglichen.

Wobei im weiteren Verlauf diese Definition mit Leben gefüllt wird:

1. Grundlage für alle Datenerfassung, -verarbeitung, -analyse und die darauf basierten Entscheidungen sind immer IT-Systeme: von den IoT-Sensoren, die ihre Daten an die zentralen oder auch verteilten Systeme zu Weiterverarbeitung geben, bis hin zum zentralistischen ERP-System, welches die Entscheidungen umsetzt. Die Methoden werden in Abschn. 2.2, die Technologien, Systeme und Anwendungen in Kap. 3 beschrieben.
2. Die Erfassung von Daten, deren Verarbeitung und das Zusammenspiel wird in Abschn. 1.4 in einem umfassenden Technologie-Framework beschrieben.
3. Beispiele für die Entscheidungsunterstützung und -automatisierung finden Sie dabei im ganzen Buch und besonders in Kap. 3 und vor allem in Kap. 4.

1.2 Abgrenzung zwischen Business Intelligence und Business Analytics

Die eingennommene Perspektive – Gegenwart bis Zukunft – Ein Teil der vorhergehenden Ausführungen treffen auch auf den bisher bekannteren Begriff der Business Intelligence (BI) zu. Im Folgenden wollen wir einmal ausführen, worin sich die Begrifflichkeiten der BI und der BA unterscheiden und wo es eine Schnittmenge gibt. So gibt es in den einnehmbaren Perspektiven im Rahmen der Analytik einen gleichartigen Rahmen (siehe Tab. 1.1).

Die meisten Autoren ziehen die Linie zwischen Business Intelligence (BI) und Business Analytics in Bezug auf die zeitliche Zielstellung der Anwendung. So wird BI generell eine reine Ex-post- und BA eine Ex-ante-Perspektive unterstellt. Diese Abgrenzung ist aus einer technischen Sicht, wenn man beispielsweise rein die eingesetzten Algorithmen betrachtet, sicherlich richtig, auch wenn man die Betrachtung auf die grundsätzlich gestellten Fragen, die den Gegenstand und die Rechtfertigung für den Einsatz von BI bilden, ansieht. Diese stützen sich tatsächlich im Hauptaugenmerk auf Kennzahlensysteme (KPIs)

Tab. 1.1 Perspektiven der Analytik

	Vergangenheit	Gegenwart	Zukunft
Informationen	Was ist passiert? (Reporting)	Was passiert gerade? (Alerts)	Was wird passieren? (Hochrechnung)
Erkenntnisse	Wie und warum ist etwas passiert? (Modellierung)	Was ist der beste nächste Schritt (Empfehlungen)	Was wird im besten und schlechtesten Fall passieren? (Vorhersage, Optimierung)

Im Rahmen der Analytik können verschiedene Perspektiven eingenommen werden. Die führende Struktur bildet dabei der zeitliche Rahmen zwischen der Vergangenheit und der Zukunft.

aus der Aggregation von historischen Daten. Erweitern wir die Betrachtung nun aber auf eine Gesamtperspektive und betrachten BI aus einer gesamtunternehmerischen Sicht, losgelöst von der technischen und operativen Ebene, so gliedert sich BI doch sicherlich in einen größeren Kontext der betrieblichen Entscheidungsunterstützung ein. BI ist niemals ein Selbstzweck, sondern ein Unterstützungsinstrument für Entscheidungen, durchgeführt von Menschen. Jede ermittelte KPI dient im Grunde ja nur dazu, dass, basierend auf dieser, eine Beurteilung und Entscheidung für zukünftige Veränderungen (oder eben auch nicht) getroffen wird. Die KPI zum Umsatz nach geografischen Absatzregionen und die dazugehörige KPI der Veränderung derselben kann mehreren Zwecken dienen. Abb. 1.1 verdeutlicht die zeitlogische Perspektive anhand einer Zeitleiste mit Daten, die zum Erstellen von Vorhersagemodellen oder Business-Intelligence-Berichten verwendet werden. Die vertikale Linie in der Mitte stellt die Zeit dar, in der das Modell gebaut wird (heute/jetzt). Die Daten, die für den Aufbau der Modelle verwendet werden, befinden sich auf der linken Seite, da diese immer historische Daten darstellen – es können logisch keine Daten aus der Zukunft existieren. Wenn prädiktive Modelle, die die Grundlage von Business Analytics bilden, erstellt werden, um ein „zukünftiges“ Ereignis vorherzusagen, werden die für den Aufbau der prädiktiven Modelle ausgewählten Daten auf eine Zeit vor dem Datum, an dem das zukünftige Ereignis bekannt ist, zurückgreifen. Wenn man beispielsweise ein Modell erstellen will, um vorherzusagen, ob ein Kunde auf eine E-Mail-Kampagne antwortet, so beginnt man mit dem Datum, an dem die Kampagne erstellt wurde (wenn alle Antworten eingegangen sind), um alle Teilnehmer zu identifizieren. Dies ist das Datum für die Bezeichnung „Zielvariable, die aus diesem Datum berechnet wurde“ in der Abbildung Abb. 1.1. Die als Eingaben verwendeten Attribute müssen vor dem Datum des Mailings selbst bekannt sein, sodass diese Werte links vom Datum der Sammlung der Zielvariablen gesammelt werden. Mit anderen Worten, die Daten werden mit allen Modellierungsdaten in der Vergangenheit erstellt, aber die Zielvariable liegt noch in der Zukunft bis zum Datum, an dem die Attribute in der Zeitachse der für die Modellierung verwendeten Daten gesammelt werden. Es ist jedoch wichtig klarzustellen, dass sowohl die Analysen der Business Intelligence als auch der Business Analytics auf den gleichen Daten basieren und

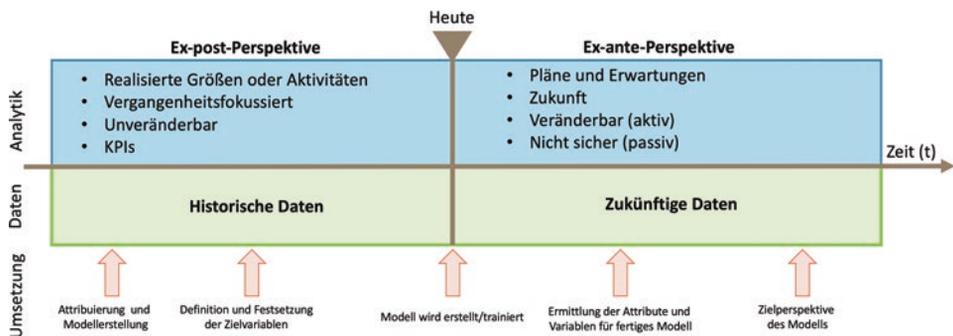


Abb. 1.1 Die Analytik-, Daten- und Umsetzungsperspektiven der Business Analytics (eigene Darstellung)