

Engels, Barbara; Goecke, Henry

**Research Report**

## Big Data in Wirtschaft und Wissenschaft: Eine Bestandsaufnahme

IW-Analysen, No. 130

**Provided in Cooperation with:**

German Economic Institute (IW), Cologne

*Suggested Citation:* Engels, Barbara; Goecke, Henry (2019) : Big Data in Wirtschaft und Wissenschaft: Eine Bestandsaufnahme, IW-Analysen, No. 130, ISBN 978-3-602-45624-6, Institut der deutschen Wirtschaft (IW), Köln

This Version is available at:

<https://hdl.handle.net/10419/201760>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



# **IW-Analysen 130**

## **Big Data in Wirtschaft und Wissenschaft**

**Eine Bestandsaufnahme**

Barbara Engels / Henry Goecke



Forschungsberichte aus dem  
Institut der deutschen Wirtschaft

# IW-Analysen 130

## Big Data in Wirtschaft und Wissenschaft

Eine Bestandsaufnahme

Barbara Engels / Henry Goecke

**Bibliografische Information der Deutschen Nationalbibliothek.**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie. Detaillierte bibliografische Daten sind im Internet über <http://www.dnb.de> abrufbar.

ISBN 978-3-602-15008-3 (Druckausgabe)

ISBN 978-3-602-45624-6 (E-Book|PDF)

Herausgegeben vom Institut der deutschen Wirtschaft Köln e.V.

Grafik: Dorothe Harren

© 2019 Institut der deutschen Wirtschaft Köln Medien GmbH

Postfach 10 18 63, 50458 Köln

Konrad-Adenauer-Ufer 21, 50668 Köln

Telefon: 0221 4981-452

Fax: 0221 4981-445

[iwmedien@iwkoeln.de](mailto:iwmedien@iwkoeln.de)

[www.iwmedien.de](http://www.iwmedien.de)

Druck: Elanders GmbH, Waiblingen



# Inhalt

<b>Kurzdarstellung</b>	4
<b>1 Daten im digitalen Zeitalter</b>	5
1.1 Definition von Big Data	6
1.2 Abgrenzung von Big Data und künstlicher Intelligenz	8
<b>2 Big Data in der Wissenschaft</b>	10
<b>3 Big Data in der Wirtschaft</b>	16
3.1 Veränderte Produktionsbedingungen und neue Geschäftsmodelle	24
3.2 Empirische Validierung der Chancen und Herausforderungen von Big-Data-Analysen	32
<b>4 Fazit</b>	50
<b>Literatur</b>	54
<b>Autoren</b>	58
<b>Abstract</b>	59

## Kurzdarstellung

Die immer größere Anzahl und Vielfalt der im Zuge der fortschreitenden Digitalisierung entstehenden Daten ist eng mit dem Begriff „Big Data“ verbunden. Damit nimmt aus ökonomischer Perspektive neben der Bedeutung von Daten auch die Relevanz von Big Data zu. Die vorliegende IW-Analyse erfasst die Effekte von Big Data über die zwei Wirkkanäle Wirtschaft und Wissenschaft. In der Wissenschaft steigt die Bedeutung von Big Data und verwandten Themenfeldern stetig: Seit 2015 hat sich dazu der Anteil an allen Publikationen nahezu verdoppelt. Wissenschaftliche Erkenntnisse münden häufig in Innovationen und haben so einen indirekten ökonomischen Effekt. Der direkte ökonomische Effekt, der sich unter anderem darin äußert, inwiefern Unternehmen Big-Data-Technologien einsetzen, nimmt ebenfalls zu, ist aber insgesamt noch gering. Lediglich gut 8 Prozent der 1.235 deutschen Unternehmen aus Branchen der Industrie und der industrienahen Dienstleistungen, die exklusiv für diese Untersuchung befragt wurden, führen Big-Data-Analysen durch. Die Umfrage ergab auch, dass sich viele Unternehmen derzeit noch zahlreichen Herausforderungen gegenübersehen, die die Nutzung von Big Data behindern. Zu diesen gehören vor allem das fehlende Know-how im Unternehmen, die mangelnde Qualität und Quantität der Daten sowie zu geringe Budgets. Beim fehlenden Know-how ist eine Kooperation mit Start-ups zu empfehlen. Die Bedeutung der Datenbewirtschaftung sollte in Unternehmen generell bewusster sein, um der mangelnden Qualität und Quantität von Daten in Unternehmen entgegenzuwirken. Hierdurch ist auch zu erwarten, dass das Bewusstsein für das tatsächliche Kosten-Nutzen-Verhältnis des Einsatzes von Big-Data-Technologien geschärft wird und folgerichtig mehr in Big Data investiert wird.

# 1 Daten im digitalen Zeitalter

Die wichtigste Ressource der Digitalisierung sind Daten. Die zahlreichen neuen Technologien und Anwendungen, die sich in Wirtschaft, Wissenschaft und Gesellschaft ausbreiten, haben eins gemeinsam: Sie basieren auf Daten und produzieren selbst große Datenmengen. Die Anwendungen der vernetzten Fabrik (Industrie 4.0) oder von Smart Mobility, Smart Home und E-Commerce, um nur wenige Beispiele zu nennen, sind nicht ohne das Aufzeichnen und Auswerten von Daten möglich. Der Begriff „Smart“, der im Zusammenhang mit der Digitalisierung immer wieder genutzt wird, bezieht sich auch darauf, dass Technologien auf Basis von Daten ökonomisch sinnvolle und nutzerorientierte Entscheidungen treffen. Daten werden auch zum Kern von immer mehr Geschäftsmodellen. Im Gartner Hype Cycle der Emerging Technologies basieren fast alle der genannten Technologien auf Daten (Pannetta, 2018). Auch in Unternehmen, die ihr Geschäftsmodell nicht originär auf Daten gründen, spielen diese als Produktionsfaktor eine immer größere Rolle.

Im Jahr 2018 wurden im Zuge der fortschreitenden Digitalisierung weltweit 33 Zettabyte (Trillionen Gigabyte) an vielfältigen Daten produziert, das waren fünfmal mehr als fünf Jahre zuvor. Für die kommenden fünf Jahre wird erwartet, dass dieser Wert sich auf etwa 100 Zettabyte verdreifacht – dabei ist kein Ende dieser Dynamik in Sicht (Reinsel et al., 2018). Auch die Anzahl an Datenlieferanten und die Vielfalt an Datenquellen nehmen zu: Immer mehr Sensoren an immer mehr vernetzten Geräten tragen zu der wachsenden Datenmenge bei. Die immense Anzahl neuer Datensätze bedeutet jedoch nicht, dass all diese Daten auch sinnvoll genutzt werden. Vom aufgezeichneten Datum bis zur tatsächlich verwertbaren und nutzbringenden Information ist es oft ein weiter Weg. Die Datenmengen müssen sinnvoll gereinigt, aufbereitet und kombiniert werden, um tatsächlich Informationen aus ihnen gewinnen zu können.

Datenmenge und Datenvielfalt steigen im Zuge der wirtschaftlichen Produktion, aber auch durch die gesellschaftliche Nutzung des Internets und all seiner Angebote. Dadurch gewinnen die Aufbereitung, die Auswertung und die Übersetzung der Daten in Wertschöpfung durch Unternehmen immer mehr an Bedeutung. All diese Schritte sind eng verknüpft mit dem Begriff von Big Data.

Die vorliegende IW-Analyse bietet einen Überblick über die vielfachen Auswirkungen von Big Data auf Wirtschaft und Wissenschaft und liefert gleichermaßen eine Bestandsaufnahme über die bisherige Entwicklung und Verbreitung von Big Data. Anders als Studien, die die volkswirtschaftliche Bedeutung von Big Data durch eine prozentuale Änderung im Bruttoinlandsprodukt (BIP) angeben (vgl. beispielsweise Centre for Economics and Business Research, 2016; Europäisches Parlament, 2016), erfasst diese IW-Analyse die Bedeutung durch verschiedene Kennzahlen im Bereich Wirtschaft und Wissenschaft. Dies geschieht vor dem Hintergrund, dass die Auswirkungen von Big Data mannigfaltig und teilweise kaum quantifizierbar sind. Das Abstellen auf eine einzige Schätzung oder eine einzige Kennzahl würde dem nicht Rechnung tragen und würde vermutlich zu Fehleinschätzungen führen. Diese Studie bezieht stattdessen Statistiken aus unterschiedlichen Quellen ein und liefert eigene empirische Beiträge. Mit den Bereichen Wirtschaft und Wissenschaft werden zudem die direkten und indirekten zentralen ökonomischen Dimensionen einer Volkswirtschaft abgedeckt.

Um zu klären, was hier unter Big Data verstanden wird und was nicht, erfolgt in einem ersten Schritt eine Definition von Big Data und eine Abgrenzung zum Thema der künstlichen Intelligenz. Anschließend wird der Status quo von Big Data in der Wissenschaft aufgezeigt und mit einer eigenen Analyse erweitert (Kapitel 2). Die wirtschaftliche Bedeutung von Big Data ist Gegenstand von Kapitel 3. Nach einer Betrachtung auf europäischer Ebene liegt der Fokus auf der Bedeutung von Big Data in den deutschen Unternehmen und auf einer theoretischen Einordnung der Auswirkungen von Big Data auf die Produktionsbedingungen der Unternehmen (Kapitel 3.1). Erweitert werden diese Aspekte mithilfe einer empirischen Analyse, die auf einer eigens für diese Studie angefertigten Unternehmensbefragung basiert (Kapitel 3.2). Abschließend werden aus den theoretischen Einordnungen und den empirischen Befunden wirtschaftspolitische Schlüsse gezogen (Kapitel 4).

## **1.1 Definition von Big Data**

Das Kernelement der Datenökonomie, die Daten selbst, werden in der wissenschaftlichen Literatur und der politischen Diskussion oft unter dem Begriff „Big Data“ zusammengefasst. Hierbei erfolgt in der Literatur die Abgrenzung zwi-

schen Big Data und kleineren Datenmengen anhand spezieller Eigenschaften. Diese speziellen Eigenschaften von Big Data werden häufig als die V-Eigenschaften (V's) bezeichnet. Eine eindeutige und verständliche Definition, die in ähnlicher Form bei vielen weiteren Autoren zu finden ist, liefert die TechAmerica Foundation (2012, 10): „Big Data is a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.“

Je nach Quelle und Autor schwankt die Anzahl der zur Definition von Big Data genutzten Eigenschaften zwischen drei und zehn. Es sei allerdings darauf hingewiesen, dass diese V-Eigenschaften der umfassenden Mehrdimensionalität von Big Data nicht vollständig Rechnung tragen können: Die vielfältigen Dimensionen von Big Data reichen von der ökonomischen und technischen über die gesellschaftliche und rechtliche bis hin zur politischen Perspektive, die alle zu anderen Definitionen, Implikationen und Handlungsempfehlungen führen können (Dorschel, 2015).

Die erste Beschreibung nach den V-Eigenschaften findet sich bei Laney (2001), der auf Volume, Variety und Velocity von Daten abstellt. Diese drei Eigenschaften sind auch Bestandteil der Abgrenzungen von Big Data, die auf mehr Eigenschaften beruhen, werden aber durch weitere ergänzt (vgl. hierzu unter anderem Baron, 2013; Dorschel, 2015; King, 2014; vbw, 2016). In der vorliegenden IW-Analyse wird Big Data anhand dieser drei zentralen V-Eigenschaften charakterisiert: Data Volume, Data Variety und Data Velocity.

- Der Begriff **Data Volume** stellt auf die Quantität, die bloße Menge an Daten, ab. Wie groß diese Datenmenge sein muss, damit ihre Verwendung einer Big-Data-Anwendung zuzurechnen ist, lässt sich nicht eindeutig festlegen. Der Grundgedanke ist, dass die Datenmenge so groß ist, dass sie nicht mehr mit den herkömmlichen statistischen Methoden und technologischen Standards verarbeitet werden kann. Beispielsweise werden bei Big Data verteilte IT-Systeme notwendig. Die Datenmenge ist zu groß, um auf einer einzelnen Maschine verarbeitet zu werden. Da sich die Methoden und Technologien stetig weiterentwickeln, kann sich die Zuordnung einer Anwendung zu Big Data unter dem Aspekt des Data Volume im Zeitverlauf ändern.

- **Data Variety** beschreibt die Heterogenität von Daten. Daten liegen in unterschiedlichen Formaten wie beispielsweise Texten, Audio- oder Videodateien vor. Viele dieser Daten sind zudem unstrukturiert und passen nicht in ein vordefiniertes Datenmodell. Die Aufgabe von Big-Data-Anwendungen ist es, aus dieser Vielfalt von unterschiedlichen Daten und ihrer Unstrukturiertheit Informationen und Zusammenhänge zu generieren.
- Bei der **Data Velocity** geht es um die Geschwindigkeit der Daten. Differenziert wird hierbei nach der Geschwindigkeit, mit der Daten entstehen, nach der Geschwindigkeit, mit der Daten verändert werden müssen, oder nach der Verarbeitungsgeschwindigkeit von Big-Data-Systemen (Dorschel, 2015). Bei einer Big-Data-Anwendung werden beispielsweise die Daten mit einer so hohen Geschwindigkeit erzeugt, dass diese dadurch unter Umständen nicht komplett gespeichert werden können.

Neben Volume, Variety und Velocity von Daten werden in der wissenschaftlichen Literatur noch folgende V-Eigenschaften von Big Data definiert: Data Veracity/Validity (Richtigkeit der Daten), Data Visualization (Darstellung der Daten), Data Variability (Veränderbarkeit der Daten), Data Vulnerability (Anfälligkeit im Sinne von Sicherheitsaspekten), Data Volatility (Unbeständigkeit) und Data Value (Wertigkeit der Daten); für eine genauere Erläuterung dieser bis zu zehn Eigenschaften vgl. Firican (2017). In der vorliegenden IW-Analyse werden unter Big Data sowohl Datensätze mit den oben beschriebenen Eigenschaften verstanden als auch deren Verarbeitung im Sinne von Speichern, Bereinigen, Aufbereiten etc.

## 1.2 Abgrenzung von Big Data und künstlicher Intelligenz

Die Begriffe Big Data und künstliche Intelligenz (KI) werden im politischen und gesellschaftlichen Diskurs häufig in einem Atemzug verwendet. Um nicht den Anschein zu erwecken, dass diese beiden Begriffe Synonyme sind, geht dieses Kapitel auf den Begriff der KI ein, um die Unterscheidung zu dem Begriff „Big Data“ zu erleichtern.

Der Begriff von Big Data bezieht sich zunächst auf eine Beschreibung von Daten, basierend auf verschiedenen Dateneigenschaften. Er wird aber oft sy-

onym für deren Verarbeitung, Anwendung und Analyse verwendet. Die Begrifflichkeit der künstlichen Intelligenz stellt hingegen nicht ab auf Daten oder eine Datenmenge, sondern auf Algorithmen, die diese Daten als Inputfaktoren nutzen. Kurz formuliert: Big Data ist eine Voraussetzung für künstliche Intelligenz; aber künstliche Intelligenz ist keine Voraussetzung für Big Data. Big Data kann demnach ohne KI existieren. Für gute Ergebnisse im Sinne ausreichender Datenmengen zum Lernen kommt KI aber nicht ohne Big Data aus.

Es gibt keine genaue oder einheitliche Definition von KI. Zwar haben immer mehr Staaten eine KI-Strategie herausgegeben, allerdings ohne eine einheitliche KI-Definition zu nutzen (Dutton, 2018; Konrad-Adenauer-Stiftung, 2018). Bei einer Meta-Analyse dieser und weiterer Dokumente zu nationalen und supranationalen KI-Strategien, die im Zuge dieser IW-Analyse durchgeführt wurde, zeigt sich, dass nur in einigen dieser Dokumente überhaupt eine Definition von KI vorgenommen wird. Viele beinhalten nur eine eher schwammige Beschreibung. Einige unternehmen gar keinen Versuch, künstliche Intelligenz zu beschreiben oder abzugrenzen – und das in Dokumenten, die nationale KI-Strategien formulieren.

Ein Beispiel einer Definition von KI liefert die Datenethikkommission der Bundesregierung: „Wir verstehen ‚Künstliche Intelligenz‘ [...] als Sammelbegriff für diejenigen Technologien und ihre Anwendungen, die durch digitale Methoden auf der Grundlage potenziell sehr großer und heterogener Datensätze in einem komplexen und die menschliche Intelligenz gleichsam nachahmenden maschinellen Verarbeitungsprozess ein Ergebnis ermitteln, das ggf. automatisiert zur Anwendung gebracht wird. Die wichtigsten Grundlagen für KI als Teilgebiet der Informatik sind die subsymbolische Mustererkennung, das maschinelle Lernen, die computergerechte Wissensrepräsentation und die Wissensverarbeitung, welche Methoden der heuristischen Suche, der Inferenz und der Handlungsplanung umfasst“ (Datenethikkommission, 2018, 1).

Hierbei zeigt sich zum einen die Eigenschaft, dass KI eine Technologie beziehungsweise Anwendung ist. Zum anderen verwendet diese Technologie in ihrem Verarbeitungsprozess Datensätze. Dies ist der wesentliche Unterschied zwischen den beiden Begriffen künstliche Intelligenz und Big Data, der bereits kurz angerissen wurde: In dieser IW-Analyse besteht der Unterschied zwischen

den beiden Begriffen KI und Big Data vor allem darin, dass KI Anwendung beziehungsweise Algorithmus ist, während Big Data Daten und ihre Aufbereitung beschreibt. KI wird als regel- oder datenbasierte Anwendung verstanden, die Entscheidungen trifft, während Big Data in erster Linie die Informationsgenerierung umfasst. In der Definition der Datenethikkommission ist mit dem Verweis auf „sehr große und heterogene Datensätze“ des Weiteren zu erkennen, dass Big Data eine bedeutende Rolle als einem nötigen Input für die KI eingeräumt wird.

Künstliche Intelligenz ist keine neue Erscheinung, sondern wurde schon vor Jahrzehnten eingesetzt. In den Anfängen gab es viele Anwendungen von KI für menschliche Spiele wie beispielsweise Schach (Samuel, 1959). Dieser Anwendungsbereich eignet sich aufgrund des einfachen Regelsystems und der damit klar beschreibbaren Handlungsoptionen. Diese werden von dem Algorithmus so lange in ihren Kombinationen durchsucht, bis ein gewünschtes Ergebnis erreicht wird. Darauf folgten zeitlich die KI-Anwendungen des maschinellen Lernens. Hierbei lernt der Algorithmus selbstständig aus den Ergebnissen, die er erzeugt hat. Der Algorithmus erfährt ein Feedback, mit dessen Information er Optimierungen vornimmt. Die neueste Entwicklung ist Deep Learning, welches mithilfe von neuronalen Netzen funktioniert. Der strukturelle Aufbau der neuronalen Netze orientiert sich an den Nervenzellenverbindungen des menschlichen Gehirns (vbw, 2016). Der Lernprozess des Algorithmus verwendet mehrere Ebenen, die miteinander verbunden sind. Aus den Daten wird gelernt und es werden Ergebnisse berechnet (für weitere Erläuterungen vgl. IIT, 2018, 29 ff.).

## 2 Big Data in der Wissenschaft

Über den positiven Effekt von wissenschaftlicher Forschung auf eine zusätzliche Wertschöpfung in Unternehmen und im Aggregat auf eine gesamte Volkswirtschaft existiert eine breite Literatur. Diese liefert sowohl empirische Evidenz als auch eine theoretische Fundierung (für empirische Befunde vgl. Jaffe, 1989; Mansfield, 1998; Becheikh et al., 2006; für die Theorie vgl. Romer, 1986; 1990; Acs et al., 1994). Eines der Hauptargumente der Wirkungskanäle von Forschung

auf Wertschöpfung sind sogenannte Spill-over-Effekte. Diese beschreiben den Wissenstransfer von Forschung in die Unternehmen und manifestieren sich unter anderem in Patentanmeldungen und Innovationen (Acs et al., 1994; Audretsch/Feldman, 1996).

Vor diesem Hintergrund sind die Aktivitäten in der wissenschaftlichen Forschung im Bereich von Big Data besonders interessant (für einen Exkurs der Bedeutung von Big-Data-Analysen in der Volkswirtschaftslehre vgl. zum Beispiel Engels, 2016; Einav/Levin, 2014). Zur Messung der Forschungsaktivität in einem Bereich ist die Anzahl an wissenschaftlichen Publikationen ein relevanter Indikator.

Die Zeitschrift „The Economist“ nutzte diesen Indikator, um die Bedeutung von Big Data in der wissenschaftlichen Forschung zu messen. Mithilfe von Bibliometrie wurde analysiert, in wie vielen wissenschaftlichen Papieren die Methoden von Big Data und maschinellem Lernen verwendet wurden (The Economist, 2016). Hierzu wurden alle Working Paper des National Bureau of Economic Research (NBER) dahingehend untersucht, ob in den jeweiligen Abstracts eines der Schlagwörter „Big Data“ oder „Machine Learning“ vorkam. Dieses Vorgehen lässt sich damit begründen, dass die Anzahl an Veröffentlichungen zu einem bestimmten Suchbegriff innerhalb eines Jahres einen Anhaltspunkt zu der Wichtigkeit dieses Begriffs im wissenschaftlichen Diskurs darstellt. Die Anzahl an Publikationen, bei denen das der Fall war, wurde in Relation zu der jährlichen Gesamtzahl an Publikationen gesetzt und mit klassischen Methoden der empirischen Wirtschaftswissenschaft wie der Difference-in-Differences-Methode verglichen.

Im Ergebnis zeigt sich, dass in der NBER-Publikationsreihe erste Papiere zu dem Thema im Jahr 2013 zu finden waren. Danach nahm die Anzahl der Papiere, bei denen Big Data oder maschinelles Lernen eine Rolle spielten, bis zum Ende des Beobachtungszeitraums im November 2016 stetig zu. Im Vergleich zu den klassischen Methoden lagen diese Häufigkeiten jedoch auf einem relativ niedrigen Niveau – im Jahr 2016 wurden viermal so viele Difference-in-Differences-Papiere publiziert. Zu bedenken ist allerdings, dass die Untersuchung von „The Economist“ nur die NBER-Working-Paper herangezogen hat. Die Gesamtzahl an veröffentlichten Artikeln ist damit für die wirtschaftswissenschaft-

liche Community zwar relevant, jedoch recht überschaubar. Beispielsweise wurden im Jahr 2018 dort gut 1.200 Artikel veröffentlicht – sehr wenige im Vergleich zu den knapp 3,1 Millionen Publikationen, die im Jahr 2018 in der Literaturdatenbank Scopus eingestellt wurden.

Eine Untersuchung mit einem sehr ähnlichen Ansatz ist bei Huber et al. (2018) zu finden. Die Autoren analysieren – ebenfalls bis 2016 – die in wirtschaftswissenschaftlichen Publikationen verwendeten Methoden mit dem Fokus, wie relevant dabei die Methode des maschinellen Lernens ist. Nach den entsprechenden Begriffen suchen sie nicht nur im Abstract, sondern auch in den jeweiligen Titeln und zugeordneten Schlüsselwörtern. Ein weiterer Unterschied ist die herangezogene Datenquelle. Huber et al. (2018) verwenden die Literaturdatenbank EconLit. Dadurch erhöht sich die Datenbasis auf über eine Million Veröffentlichungen, wovon knapp 800.000 Publikationen in die finale Auswertung über einen Zeitraum von 25 Jahren eingehen. Die Ergebnisse decken sich qualitativ mit denen von „The Economist“: Bis zum Anfang der 2000er Jahre kommt die Methode des maschinellen Lernens gar nicht vor. Etwa seit 2013 nimmt die Bedeutung stark zu, liegt aber bis zum Jahr 2016 noch auf einem relativ geringen Niveau von 0,3 Prozent.

Um ein genaueres und aktuelleres Bild der Bedeutung von Big Data in der wissenschaftlichen Forschung zu erhalten, wurde im Zuge der vorliegenden Analyse eine eigene Untersuchung durchgeführt. Hierzu wurde die Datenbank Scopus verwendet. Scopus ist eine der weltgrößten Literaturdatenbanken für alle wissenschaftlichen Richtungen mit insgesamt über 43,5 Millionen Publikationen im aktuellen Jahrtausend. Diese sind die Grundlage dieser Analyse. Die Literatur wurde nach verschiedenen Suchbegriffen im zeitlichen Verlauf durchsucht.

Die Analyse schließt alle auf Scopus veröffentlichten Publikationen ein (Fachartikel, Artikel aus Fachzeitschriften, Sammelbände etc.). Um ein möglichst breites Bild von der Anwendung von Big Data in der Wissenschaft zu erhalten, war die Suche nicht auf spezielle Fachrichtungen beschränkt. Für die Suche wurden Abstract, Titel sowie Schlüsselwörter einer jeden Publikation auf das Vorhandensein des entsprechenden Begriffs hin untersucht. Kommen die Begriffe in Abstract, Titel oder Schlüsselwörtern vor, wurde der Artikel dem je-

weiligen Thema zugeordnet. Bei der Suche wurden Suchabfragen benutzt, die verschiedene Suchbegriffe durch logische Operatoren wie UND und ODER verknüpfen.

Um eine Vergleichbarkeit mit den beiden oben genannten Studien zu ermöglichen, lauteten die Suchbegriffe, bei denen nicht zwischen Groß- und Kleinschreibung differenziert wurde:

- „Difference-in-Differences“ OR „Difference in Differences“ OR „DID“
- „Randomised Controlled Trial“ OR „RCT“
- „Instrumental Variable“
- „Regression Discontinuity Design“ OR „RDD“
- „Dynamic Stochastic General Equilibrium“ OR „DSGE“
- „Big Data“
- „Machine Learning“

Diese Suchbegriffe mussten mindestens einmal in einem der drei erwähnten Bereiche (Abstract, Titel, Schlüsselwörter) auftauchen, um so für das jeweilige Jahr gezählt zu werden. Im Wissen, dass Big Data und die Methode des maschinellen Lernens vor dem Jahr 2000 in der wirtschaftswissenschaftlichen Literatur keine Bedeutung hatten, wurde der Suchzeitraum auf die Jahre 2000 bis 2018 beschränkt.

In der Annahme, dass jeder Artikel zumindest einen Punkt als Satzzeichen im Abstract enthält, wurde der Punkt als Suchkriterium benutzt, um alle veröffentlichten Artikel in den untersuchten Jahren zu identifizieren. Dieser Schritt ist nötig, um im Folgenden die Anzahl der Artikel zu den jeweiligen Themen mit der Grundgesamtheit in Relation setzen zu können. Durch den so berechneten Anteil wird ausgeschlossen, dass ein Thema als eines mit steigender Relevanz identifiziert wird, nur weil es über die Zeit einen absoluten Anstieg in den themenspezifischen Publikationen gab. Wenn zeitgleich die Gesamtzahl an Publikationen überproportional anstieg, hätte sich vielmehr die Bedeutung des Themas verringert.

Die Suche auf Scopus identifizierte insgesamt gut 43,5 Millionen Veröffentlichungen in den 19 untersuchten Jahren. Die jährliche Gesamtzahl an Veröffent-

## Verwendung von Forschungsmethoden

Tabelle 1

Begriffe in Abstracts, Titeln oder Schlüsselwörtern der in der Scopus-Datenbank verfügbaren wissenschaftlichen Literatur (2000–2018)

Suchbegriff	Anzahl <sup>1)</sup>	In Prozent
Difference-in-Differences	3.268.000	7,5
Randomised Controlled Trial	625.000	1,4
Instrumental Variable	382.000	0,9
Regression Discontinuity Design	1.000	< 0,01
Dynamic Stochastic General Equilibrium	1.000	< 0,01
Big Data	58.000	0,1
Machine Learning	275.000	0,6

1) Auf 1.000 gerundet.

Quellen: Scopus, 2019; eigene Berechnungen

lichungen auf Scopus hat fast jedes Jahr zugenommen. Die Suche nach den oben genannten Methoden brachte insgesamt gut 4,6 Millionen Artikel hervor. Das entspricht 10,6 Prozent aller Artikel. Dabei wurde nicht differenziert, in welcher der drei Kategorien (Abstract, Titel, Schlüsselwörter) der Suchbegriff gefunden wurde. Es kann auch nicht ausgeschlossen werden, dass die Suchergebnisse sich überschneiden, dass sich also ein Artikel mit mehr als einer Methode befasst hat und deshalb mehrfach in die Zählung einging. Die Anzahl der Treffer der verschiedenen Methoden in den Jahren 2000 bis 2018 ist in Tabelle 1 dargestellt.

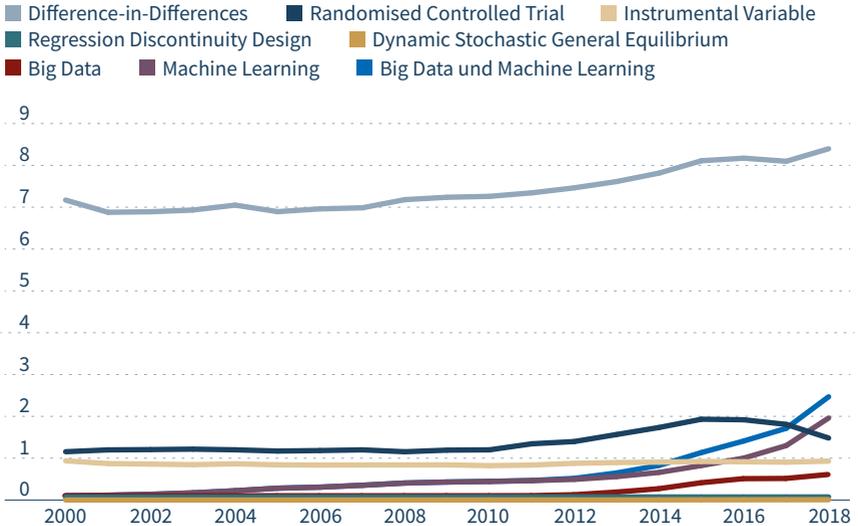
Es zeigt sich, dass in der wissenschaftlichen Literatur die Difference-in-Differences-Methode über die Jahre hinweg sehr stark angewendet wurde. Aktuell wird sie in einem recht konstanten Anteil an Publikationen von knapp 8,5 Prozent verwendet (Abbildung 1). In Bezug auf die Nutzung von Big Data ist zu erkennen, dass diese Methode vor dem Jahr 2012 noch sehr wenig zum Einsatz kam, seitdem jedoch stark an Bedeutung gewonnen hat. Im Jahr 2018 lag der Anteil an Publikationen, die Big Data nutzten, bei 0,5 Prozent der insgesamt knapp 3,1 Millionen Publikationen aus diesem Jahr. Big Data rangiert damit zwar noch weit entfernt von den am häufigsten verwendeten Methoden, ist jedoch vor dem Hintergrund der Gesamtzahl an Publikationen schon recht bedeutend.

Wenn zusätzlich zu den Studien, die sich mit Big Data befassen, noch das Anwendungsgebiet des Machine Learning berücksichtigt wird, zeigt sich eine

## Entwicklung der Verwendung von Forschungsmethoden

Abbildung 1

Begriffe in Abstracts, Titeln oder Schlüsselwörtern der in der Scopus-Datenbank verfügbaren wissenschaftlichen Literatur, in Prozent aller Publikationen



Daten: <http://dl.iwkoeln.de/index.php/s/rs36oFtiCTwoEsr>

Quellen: Scopus, 2019; eigene Berechnungen

noch stärkere Dynamik. Zum Machine Learning gab es zwar bereits Anfang dieses Jahrtausends Publikationen, jedoch in relativ überschaubarer Menge. Von 2000 bis 2012 stieg die Publikationszahl nahezu linear. Simultan mit der aufkommenden Relevanz von Big Data in wissenschaftlichen Publikationen nach 2012 drehte sich der lineare Trend bei Machine-Learning-Veröffentlichungen hin zu einer exponentiellen Entwicklung. Im Jahr 2018 hatten bereits knapp 2 Prozent aller Publikationen Machine Learning als Bestandteil. Das Aggregat von Machine Learning und Big Data steht damit aktuell für 2,5 Prozent aller wissenschaftlichen Publikationen, wobei beide Teilbereiche eine starke Dynamik aufweisen – in den letzten drei Jahren hat sich der Anteil dieser Publikationen nahezu verdoppelt. Damit werden in den wissenschaftlichen Papieren Machine Learning und Big Data zusammen aktuell häufiger verwendet als beispielsweise Randomised-Controlled-Trial-Ansätze. Ein Ende dieser Entwicklung ist, basierend auf der aktuellen Bedeutung von Big Data und maschinellem Lernen sowie den zahlreich ausgeschriebenen Forschungsarbeiten in diesen Bereichen, nicht in Sicht.

# 3 Big Data in der Wirtschaft

Für Unternehmen bietet die Anwendung von Big Data umfassende Potenziale: Damit lassen sich Prozesse verbessern, Produkte und Dienstleistungen anpassen und entwickeln sowie ganze Geschäftsmodelle aufbauen. Gleichzeitig stellen diese Anwendungen für die Wirtschaft aber hinsichtlich der benötigten Investitionen und Kompetenzen eine große Herausforderung dar. Generell lässt sich zeigen, dass Firmen, die datengetriebene Entscheidungen treffen, höhere Profite und höheren Output erzielen und produktiver sind (Brynjolfsson et al., 2011).

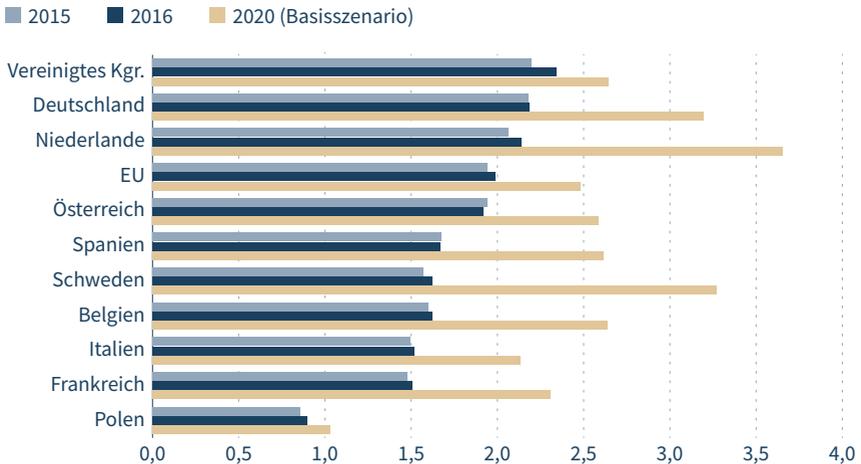
Vor diesem Hintergrund bietet dieses Kapitel eine Übersicht über die Verbreitung von Big Data und entsprechenden Anwendungen in der Wirtschaft. Zunächst werden Unternehmen im europäischen Vergleich betrachtet. Es wird jeweils die Empirie der zehn Volkswirtschaften mit dem größten Bruttoinlandsprodukt (Belgien, Deutschland, Frankreich, Italien, Niederlande, Österreich, Polen, Schweden, Spanien, Vereinigtes Königreich) und der Durchschnitt der EU-28 dargestellt. Anschließend beschränkt sich die Darstellung auf Big-Data-Anwendungen in Deutschland. Nach einer Branchenanalyse folgt eine genauere Betrachtung der Geschäftsfelder, in denen in Deutschland Big Data verwendet wird. Zur Vervollständigung des Bildes enthält die Analyse eine ausführliche Auswertung einer Umfrage, die unter deutschen Industrieunternehmen durchgeführt wurde (Kapitel 3.2).

Die Bedeutung von Big Data bettet sich in die übergeordnete Relevanz von Daten für Volkswirtschaften ein: Daten werden nicht nur als Inputfaktor für die Produktion von Waren und Dienstleistungen immer wichtiger, sondern stellen auch immer häufiger ein eigenständiges Produkt dar (DEMAND, 2019). Die aktuelle und zukünftige gesamtwirtschaftliche Bedeutung von Daten zeigen Statistiken, die für die Europäische Kommission erstellt wurden (Abbildung 2). Berechnet wurde hierbei die volkswirtschaftliche Bedeutung der Datenökonomie. Diese ist hier wie folgt definiert (für weitere Definitionen vgl. Möller/Karstedt-Meierrieks, 2017; Bundesverband Digitale Wirtschaft, 2018): „Die Datenökonomie misst die Gesamtauswirkungen des Datenmarktes auf die Gesamtwirtschaft. Sie umfasst die Erzeugung, Sammlung, Speicherung, Ver-

## Bedeutung der Datenökonomie<sup>1)</sup>

Abbildung 2

Anteil am jeweiligen Bruttoinlandsprodukt, in Prozent



1) Die Datenökonomie umfasst die Erzeugung, Sammlung, Speicherung, Verarbeitung, Verteilung, Analyse, Aufbereitung, Bereitstellung und Nutzung von Daten.

Daten: <http://dl.iwkoeln.de/index.php/s/myytGmSWzi9cQkb>

Quelle: IDC, 2017

arbeitung, Verteilung, Analyse, Aufbereitung, Bereitstellung und Nutzung von Daten, die durch digitale Technologien ermöglicht werden. Die Datenwirtschaft umfasst auch die direkten, indirekten und induzierten Auswirkungen des Datenmarktes auf die Wirtschaft“ (IDC, 2017, 123; eigene Übersetzung).

Es zeigt sich, dass die Datenökonomie im europäischen Durchschnitt bereits im Jahr 2016 einen Anteil am Bruttoinlandsprodukt von 2 Prozent hatte, in Deutschland lag der Wert sogar etwas höher. Es wird erwartet, dass unter der Annahme eines Basisszenarios (in dem unter anderem von einer Fortsetzung des moderaten ökonomischen Wachstums aus dem Jahr 2016 und einer stärkeren Relevanz des Internets der Dinge ausgegangen wird) sich die Bedeutung bis 2020 in allen europäischen Ländern erhöhen wird (IDC, 2017). In Deutschland beispielsweise soll die Datenökonomie im Jahr 2020 bereits für über 3 Prozent des BIP verantwortlich sein. In einem optimistischen Szenario soll der Wert gar auf 4,6 Prozent steigen (IDC, 2017). Damit hätte die Datenökonomie in Deutschland im Jahr 2020 ein ähnliches Gewicht erreicht wie die Baubranche im Jahr 2017 oder 2018 (Statistisches Bundesamt, 2018; 2019).

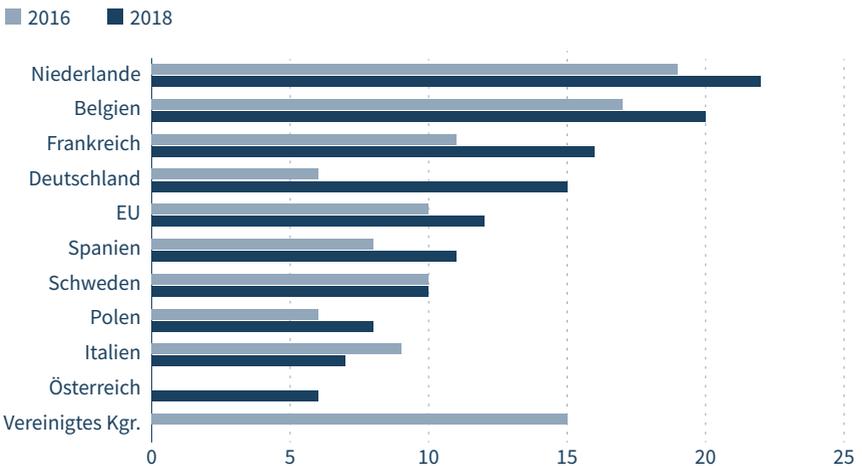
Eine erste hilfreiche Quelle zur Einordnung der Bedeutung von Big Data für die Unternehmen liefert Eurostat, das Statistikamt der Europäischen Union. Hier sind für die Jahre 2016 und 2018 Zahlen darüber zu finden, wie viele Unternehmen in den jeweiligen europäischen Ländern bereits Big-Data-Analysen durchführten, welche Datenart sie verwendeten und ob die Analysen von internen oder externen Mitarbeitern durchgeführt wurden (Eurostat, 2019). Ausgewiesen werden all diese Zahlen bis herunter auf die Ebene der Unternehmenssektoren.

Im europäischen Durchschnitt führte demnach im Jahr 2016 bereits jedes zehnte Unternehmen Big-Data-Analysen durch (Abbildung 3). Dieser Wert erhöhte sich zum Jahr 2018 auf 12 Prozent. Mit einem Wert von 22 Prozent nahmen die Niederlande den Spitzenplatz ein (hinter Malta mit einem Wert von 24 Prozent, der in Abbildung 3 nicht dargestellt ist). Deutschland hat in den Jahren von 2016 bis 2018 den größten Sprung gemacht: Analysierten im Jahr 2016 nur 6 Prozent der deutschen Unternehmen Big Data, waren es im Jahr 2018 mit 15 Prozent nach Prozentpunkten mehr als doppelt so viele Unter-

### Big-Data-Analysen in der Wirtschaft

Abbildung 3

Anteil der Unternehmen, die Big-Data-Analysen durchführen, an allen Unternehmen im jeweiligen Land<sup>1)</sup>, in Prozent



1) Ohne Bankensektor; Betriebe mit mindestens zehn Beschäftigten.

Daten: <http://dl.iwkoeln.de/index.php/s/2bqNyL8oMaJrcdF>

Quelle: Eurostat, 2019

nehmen. Im europäischen Durchschnitt betrug das Wachstum 2 Prozentpunkte. Damit lag Deutschland 2018 deutlich über dem europäischen Durchschnitt, ausgehend von einem unterdurchschnittlichen Niveau.

In Europa steigt der Anteil der Unternehmen, die Big-Data-Analysen durchführen, mit der Unternehmensgröße (Eurostat, 2019): Spitzenreiter bei den Unternehmen mit mehr als 250 Mitarbeitern ist Belgien mit einem Anteil von 55 Prozent, dicht gefolgt von den Niederlanden mit 53 Prozent. Deutschland liegt hier bei einem Wert von 34 Prozent. Von 2016 bis 2018 hat sich der Anteil der Unternehmen, die Big-Data-Analysen durchführen, in fast allen europäischen Ländern und über alle Unternehmensgrößenklassen erhöht.

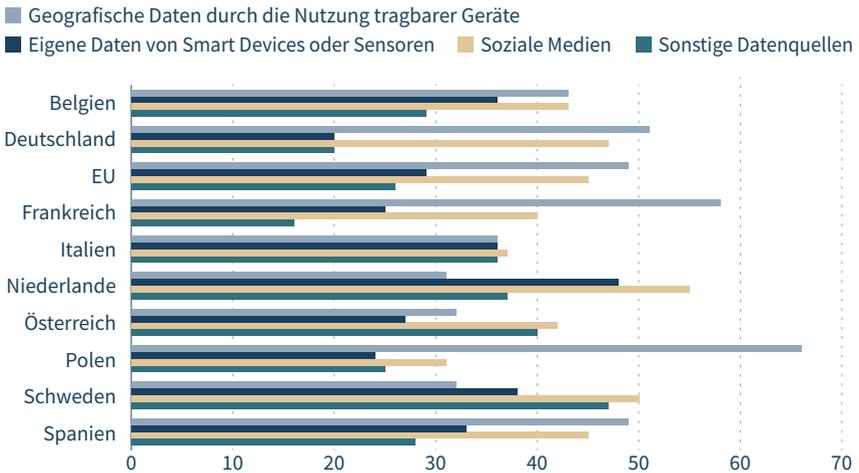
Wie bei der Definition von Big Data in Kapitel 1.1 erläutert, zeichnet sich Big Data unter anderem durch die Heterogenität der Datenarten aus (Data Variety). Dass diese breite Datenvielfalt auch in der Praxis relevant ist, zeigt Abbildung 4. Im europäischen Durchschnitt nutzen fast die Hälfte der Unternehmen, die Big-Data-Analysen durchführen, hierfür geografische Daten aus tragbaren Geräten. Etwas weniger Unternehmen (45 Prozent) nutzen Daten aus sozialen Medien. Smart Devices und Sensoren sowie sonstige Datenquellen folgen mit 29 beziehungsweise 26 Prozent. In Deutschland sind die Werte für die Nutzung der einzelnen Datenquellen ähnlich zum europäischen Durchschnitt. Spitzenreiter der Nutzung von geografischen Daten ist Polen (66 Prozent). Bei Daten aus den sozialen Medien führen die Niederlande (hinter Irland mit einem Wert von 66 Prozent, in Abbildung 4 nicht dargestellt). Bei den Sensordaten liegen ebenfalls die Niederlande vorn (48 Prozent) und bei sonstigen Datenquellen Schweden (47 Prozent). Insgesamt ist bei den europäischen Unternehmen, die Big-Data-Analysen durchführen, jedoch die Verwendung von mehreren Datenquellen weit verbreitet und eher die Regel als die Ausnahme.

Die meisten Unternehmen führen die Big-Data-Analysen selbst durch – im europäischen Durchschnitt 76 Prozent aller Unternehmen, die im Jahr 2018 Big-Data-Analysen anwendeten (Eurostat, 2019). Häufig kommen auch externe Dienstleister für Big-Data-Analysen zum Einsatz, im europäischen Durchschnitt in 42 Prozent der relevanten Unternehmen. Die Nutzung sowohl interner als auch externer Mitarbeiter gilt für fast alle europäischen Länder, für die Daten vorliegen (keine Daten für Deutschland). In den Niederlanden werden

## Datenquellen von Big-Data-Analysen

Abbildung 4

Anteil der Unternehmen, die für ihre Big-Data-Analysen im Jahr 2018 die jeweiligen Datenquellen verwendeten, an allen Big Data analysierenden Unternehmen im jeweiligen Land<sup>1)</sup>, in Prozent



1) Ohne Bankensektor; Betriebe mit mindestens zehn Beschäftigten; keine Angaben für das Vereinigte Königreich.

Daten: <http://dl.iwkoeln.de/index.php/s/74GLWKp5NdWeSyq>

Quelle: Eurostat, 2019

in 85 Prozent der Unternehmen die Big-Data-Analysen von eigenen Mitarbeitern erstellt. In Österreich sind diese Werte fast pari – hier lassen 60 Prozent der Unternehmen die Big-Data-Analysen von externen Dienstleistern durchführen und 63 Prozent von internen Mitarbeitern.

Für eine Differenzierung nach Branchen und Bereichen, in denen deutsche Unternehmen Big-Data-Analysen verwenden, wird auf die fachspezifische Plattform Crunchbase zurückgegriffen. Auf dieser Plattform liegen Daten über die Nutzung von Big Data in Unternehmen vor. Sie fokussiert sich nach eigenen Aussagen auf innovative Unternehmen und damit auf Unternehmen, die neue und innovative Technologien wie Big Data verwenden (Crunchbase, 2019). Die Plattform bekommt ihre Daten nach eigener Beschreibung unter anderem von Investitionsfirmen. Mehr als 4.000 international aktive Investitionsfirmen senden die Veränderung ihrer Portfolios monatlich an Crunchbase. Des Weiteren werden die Daten von hauseigenen Experten validiert.

Crunchbase hat die Unternehmen in seiner Datenbank nach knapp 50 Gruppenkategorien strukturiert, denen jeweils eine Vielzahl von Einzelkategorien zugeordnet ist. Eine dieser Einzelkategorien ist „Big Data“, die in der Gruppenkategorie „Data and Analytics“ zu finden ist (Crunchbase, 2019). Für jedes gelistete Unternehmen lassen sich zahlreiche Informationen abrufen, etwa eine Beschreibung der Unternehmensaktivität, der Standort oder auch die unternehmenseigene URL. In Deutschland gibt es 168 Unternehmen in der Einzelkategorie Big Data (Stand: 18. Februar 2019). Deren geografische Verteilung zeigt Abbildung 5. Es fällt auf, dass die Region Berlin der größte deutsche Standort von Big-Data-Unternehmen ist. Insgesamt gibt es, ebenso wie auch bei Unternehmen, die sich mit KI befassen (Goecke/Thiele, 2018), eine geografische Streuung mit einer Konzentration auf die großen Städte in Deutschland.

Durch die Einbeziehung des Kontexts, in dem die Big-Data-Unternehmen tätig sind, wird ein tieferer Einblick bezüglich der Anwendungsfelder in den Unternehmen möglich. Hierzu werden alle weiteren Einzelkategorien verwendet, die den Unternehmen bei Crunchbase zugewiesen sind. Abbildung 6 visualisiert die Ergebnisse: Jede Verbindung zwischen zwei Punkten stellt ein simultanes Vorkommen von zwei Einzelkategorien dar. Die Darstellung berücksichtigt, wie häufig welche Einzelkategorie gleichzeitig mit der Einzelkategorie „Big Data“ vorkommt. Somit repräsentieren die Punkte in Abbildung 6 die Einzelkategorien, die auf die Aktivitäten der Big-Data-Unternehmen in Deutschland zutreffen.

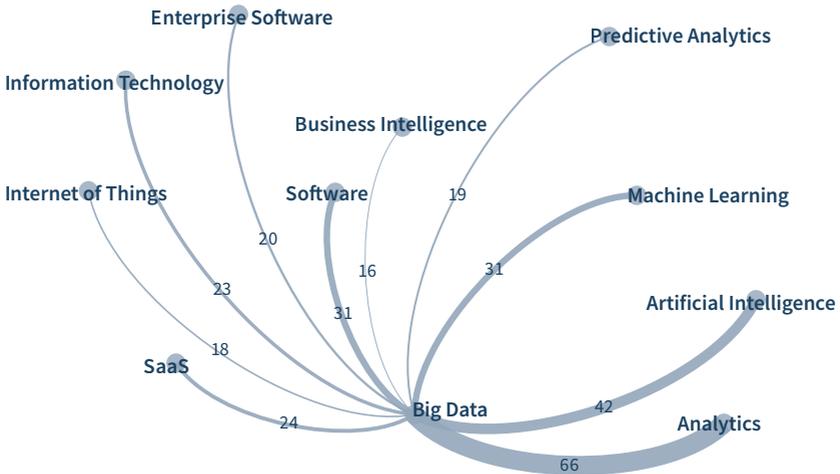
Die zehn Einzelkategorien, die am häufigsten bei den Big-Data-Unternehmen ebenfalls vorkommen, sind dargestellt. Es ist zu erkennen, dass die deutschen Big-Data-Unternehmen häufig in weiteren technischen Anwendungen wie der künstlichen Intelligenz (Artificial Intelligence: 42 Unternehmen) und dem maschinellen Lernen (Machine Learning: 31 Unternehmen) aktiv sind. Erfolgt die Betrachtung nicht von der technischen Seite, sondern aus dem Blickwinkel des Geschäftsmodells, zeigt sich, dass Big-Data-Anwendungen häufig bei der Analyse (Analytics: 66 Unternehmen) beziehungsweise Prognose (Predictive Analytics: 19 Unternehmen) zum Einsatz kommen, ebenso bei Unternehmen, die im Softwarebereich tätig sind (Software: 31 Unternehmen; Enterprise Software: 20 Unternehmen).



Stand: Februar 2019.  
Quellen: Crunchbase, 2019; eigene Berechnungen

## Weitere Anwendungsgebiete deutscher Big-Data-Unternehmen Abbildung 6

Von 168 Unternehmen, die sich mit Big Data befassen, nutzen so viele auch diese digitalen Technologien



Die Abbildung zeigt die zehn Einzelkategorien in der Crunchbase-Datenbank, die am häufigsten gemeinsam mit der Einzelkategorie „Big Data“ vorkommen. An den Verbindungslinien steht die jeweilige Anzahl der Unternehmen.

SaaS: Software as a Service.

Quellen: Crunchbase, 2019; eigene Berechnungen; Grafik erstellt mit Gephi

Die bisherigen Ausführungen sind rein deskriptiv und ermöglichen daher noch keine weiterführenden Ableitungen. Für fundierte politische Implikationen ist zunächst eine theoretische Überlegung zu den Auswirkungen von Big Data im Produktionsprozess der Unternehmen anzustellen (Kapitel 3.1). Erst im Anschluss lässt sich die Relevanz von Big Data für die Unternehmen und damit in der Breite auch für die gesamte Volkswirtschaft im Sinne von zusätzlicher Wertschöpfung verstehen. Vor dem Hintergrund dieses Wissens ist für eine positive Gestaltung der ökonomischen Zukunft ein genauerer Blick in die Unternehmen notwendig. Diesen ermöglicht die für diese IW-Analyse erstellte empirische Untersuchung auf Basis einer umfassenden Unternehmensbefragung (Kapitel 3.2).

### 3.1 Veränderte Produktionsbedingungen und neue Geschäftsmodelle

Big Data eröffnet eine Vielfalt von Optionen für Unternehmen, die an unterschiedlichen Punkten der Wertschöpfungskette ansetzen können. Big Data durchdringt zahlreiche Technologiebereiche, darunter

- industrielle Produktion,
- Energie,
- Biologie,
- Ernährungs- und Lebensmittel,
- Gesundheit und Medizin,
- Mobilität,
- Luft- und Raumfahrt,
- neue Werkstoffe und Materialien sowie
- Nanotechnologie (vbw, 2016, 14).

Die Branchen in Deutschland profitieren unterschiedlich stark von den durch Big Data entstehenden Möglichkeiten der Wertsteigerung und Wertschöpfungs-optimierung. Markl et al. (2013, 12) identifizieren folgende Bereiche als herausragend in Bezug auf ihr Innovationspotenzial durch Big Data:

- Industrie 4.0,
- Life Sciences,
- Marktforschung und Medien,
- Mobilitätsdienstleistungen,
- Energiewirtschaft,
- Risikomanagement und Versicherungswesen,
- öffentlicher Sektor.

Insbesondere im Bereich Industrie 4.0 gibt es viele Ansatzpunkte für Big-Data-Anwendungen. Jede vierte Maschine in deutschen Fabriken ist smart und arbeitet vernetzt. Das ergab eine repräsentative Befragung von 554 Industrieunternehmen ab 100 Mitarbeitern im Auftrag des Digitalverbands Bitkom (2018). Laut der Studie nutzt jedes zweite Unternehmen aus dem Produzierenden Gewerbe (49 Prozent) Industrie-4.0-Anwendungen, weitere 22 Prozent haben konkrete Pläne für den Einsatz. Das heißt: Mehr als sieben von zehn

deutschen Industrieunternehmen sind bereits im Bereich Industrie 4.0 aktiv und produzieren im Rahmen ihrer Wertschöpfung Big Data. Eng mit Industrie 4.0 verknüpft ist das Internet der Dinge. Durch das Internet der Dinge entstehen Daten in völlig neuem Umfang und in neuen Formen. Mit der Ausbreitung des Internets der Dinge nehmen sowohl die Notwendigkeit der Anwendung von Big-Data-Technologien als auch die ökonomischen Potenziale zu, die exponentiell wachsenden Datenströme zu verarbeiten und mit weiteren Daten zu kombinieren. Dementsprechend ist eine stärkere Verbreitung des Internets der Dinge unmittelbar mit einem Bedeutungsgewinn von Big-Data-Technologien verbunden (vbw, 2016, 77).

Aufgrund der sehr dynamischen und umfassenden Entwicklung von Big Data ist es schier unmöglich, ein vollständiges oder nahezu vollständiges Bild der Auswirkungen von Big Data auf die Wirtschaft im Allgemeinen und auf die Unternehmen und ihre Arbeits- und Produktionsweisen im Besonderen zu zeichnen. Vor allem lassen sich potenzielle zukünftige Anwendungsfelder und Wirkungsweisen kaum abschätzen.

In der Literatur gibt es unterschiedliche Ansätze, die Effekte von Big Data auf Unternehmen zu kategorisieren. Die OECD (2013) sieht für datenintensive Sektoren durch Big Data die Vorteile datengetriebener Forschung und Prozesse, von datengetriebenem Marketing und datengetriebener Unternehmensorganisation. Buchholtz et al. (2014) und Krotova et al. (2019) unterscheiden dagegen drei Arten, auf die Big Data für die Unternehmen einen Mehrwert schafft:

- **Data to Information** bedeutet, dass Daten genutzt werden, um bestimmte Informationen aus ihnen zu ziehen; dazu zählt unter anderem die Analyse des Kundenverhaltens, welche zu effizienteren Werbemaßnahmen beitragen kann.
- **Data to Product/Service** heißt, dass Erkenntnisse aus der Datenanalyse in der Praxis in einem Produkt oder einer Dienstleistung implementiert werden; dazu zählen die zahlreichen datengetriebenen Geschäftsmodelle.
- **Data to Management** nutzt die aus den Daten gewonnenen Informationen bei der Entscheidungsfindung und hilft etwa dabei, Verluste durch den Abbau von Entscheidungsunsicherheiten zu vermeiden.

Darüber hinaus ist es denkbar, dass Unternehmen Einnahmen aus Daten erzielen können, wenn sie diese an andere Marktteilnehmer vertreiben. Dafür ist die Bewertung der Daten Voraussetzung. Die Methoden für eine solche Einschätzung des Preises von Daten sind vielfältig (für einen Überblick vgl. Krotova et al., 2019).

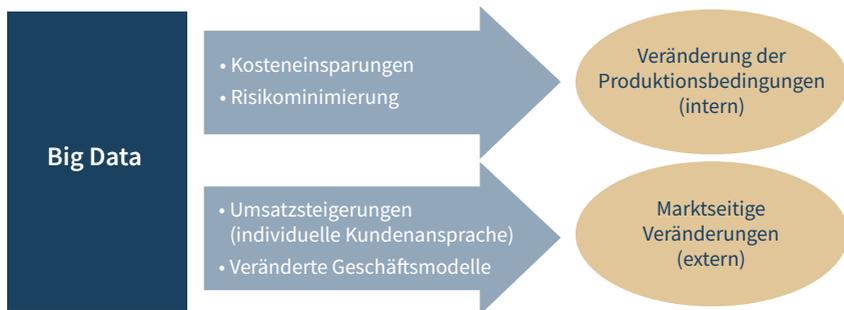
Um die Effekte von Big Data auf Unternehmen und damit die möglichen Zwecke von Big-Data-Anwendungen zu präzisieren, wird auf vier zentrale Wirkungsweisen abgestellt, die sich laut Literatur als am umfassendsten und dennoch präzisesten herausgestellt haben (vbw, 2016). Sie lehnen sich an die Kategorisierungen der OECD (2013) und von Buchholtz et al. (2014) an. Eine scharfe Trennung der einzelnen Dimensionen ist nicht möglich, da an vielen Stellen Überschneidungen und Korrelationen bestehen (Abbildung 7).

Der Einsatz von Big-Data-Anwendungen beeinflusst vor allem die Produktionsbedingungen in Unternehmen und wirkt dort über Kosteneinsparungen (erste Dimension) und die Risikominimierung (zweite Dimension). Marktseitig ist vor allem die Umsatzsteigerung durch Individualisierung der Kundenansprache (dritte Dimension) relevant. Als vierte Wirkungsweise von Big Data wird die der veränderten Geschäftsmodelle (vierte Dimension) diskutiert. Zu den Wirkungsweisen werden jeweils Beispiele genannt, die jedoch nicht auch nur annähernd ein vollständiges Bild der möglichen Breite der Big-Data-Anwendungen geben können.

### Wirkungsweisen von Big Data

Abbildung 7

Big-Data-Anwendungen in Unternehmen



Eigene Darstellung

## **Kosteneinsparungen**

Big-Data-Anwendungen werden von vielen Unternehmen genutzt, um Kosten einzusparen und so langfristig den Gewinn zu steigern. Dies geschieht, indem durch die Erhebung und Auswertung von Daten Prozesse und Verfahren sicherer, stabiler, robuster, fehlertoleranter und damit effizienter gestaltet werden. Dies ist auch ein Ziel der vernetzten Fabrik mit ihrer datengetriebenen Prozesssteuerung, ihren autonom handelnden Maschinen, cyberphysischen Systemen und selbstlernenden Produktionsprozessen (vbw, 2016, 14).

Beispielsweise ermöglichen Predictive Analytics (die Datenauswertung zu Vorhersagezwecken) und Predictive Maintenance (die Datenauswertung zur vorhersehenden Wartung von Maschinen) die Überwachung und Wartung der Produktionsanlagen. Auf der Basis von Sensordaten können Rückschlüsse auf kritische Zustände wie die Überhitzung einer Produktionsanlage gezogen und es kann proaktiv auf mögliche Ausfälle reagiert werden (IIT, 2018, 14). Durch die frühzeitige Erkennung von Maschinendefekten kann die Lebensdauer von Maschinenteilen verlängert werden. Ausfallzeiten in der Produktion können eingeschränkt oder sogar ganz vermieden werden. Laut einer Studie von McKinsey (2017, 25) ist damit zu rechnen, dass die jährlichen Instandhaltungskosten dadurch um bis zu 10 Prozent fallen. Inspektionskosten können um bis zu 25 Prozent sinken, Ausfallzeiten um 20 Prozent.

Kosten können auch eingespart werden, indem Ressourcen besser gemanagt werden. Big-Data-Analysen ermöglichen die Optimierung von Produktions- und Fertigungsplänen und eine passgenaue Personalplanung. Auslastungen der einzelnen Produktions- und Prozessschritte können besser analysiert und Überlastungen verhindert werden. Auch der Energieverbrauch kann mit der Hilfe von Big Data intelligenter und damit kostensparend gesteuert werden. Anhand von Video-, Bild- und Sensordaten lässt sich auch die Korrektheit der Prozesse in Form einer datenbasierten Qualitätskontrolle überprüfen. Die Eingangs- und Ausgangslogistik lässt sich auf Basis von Big Data ebenfalls optimieren.

Des Weiteren hat Big Data in der Logistik vor allem bei der Routen- und Lade-raumoptimierung eine hohe Relevanz, denn durch kürzere oder schnellere Strecken und höhere Auslastungen sinken die Transportkosten. Mittels Ver-

kehr-, Wetter- und Telematikdaten können Vorhersagen über voraussichtliche Ankunftszeiten eines Lkws getroffen werden, was auch eine effiziente Entladeplanung ermöglicht. Lange Wartezeiten der Fahrer vor Ort werden auf diese Weise vermieden und somit Personalkosten gesenkt. Eine sichere Just-in-time-Belieferung wird ermöglicht (Uckelmann et al., 2018).

Die durch Datenanalysen steigende Transparenz in der Lieferkette ermöglicht eine Senkung der Sicherheits- und Lagerbestände bei Lieferant und Produzent. Durch Verknüpfung von RFID- und Positionsdaten können überlastete und nicht ausgelastete Lagerbereiche identifiziert und mithilfe von Simulationen effizienter gestaltet werden (ebenda).

Bei Versicherungen können Predictive Analytics dafür sorgen, mögliche Betrugsfälle zu identifizieren, bevor Ansprüche an die Versicherten gezahlt werden. Auch dies wirkt kostensparend (Europäisches Parlament, 2016, 3). Die Minderung von Betrugsrisiken kommt auch in anderen Teilen der Finanzbranche zum Tragen.

Es ist letztlich schwierig, die Kosten- und Gewinneffekte von Big Data zu quantifizieren. Eine Umfrage des Business Application Research Center (2015) unter 32 internationalen Unternehmen hat ergeben, dass die Unternehmen im Durchschnitt mit Big Data 16 Prozent Kosteneinsparung und 13 Prozent Umsatzsteigerung erzielen. Auch aufgrund der niedrigen Fallzahlen sind diese Ergebnisse jedoch allenfalls ein Anhaltspunkt für die tatsächlichen Effekte. Es ist zu erwarten, dass diese von Unternehmen zu Unternehmen stark variieren. Konsens ist allerdings, dass Big-Data-Technologien erheblich zur Reduktion von Kosten und damit letztlich zur Steigerung des Gewinns beitragen (vbw, 2016; Forbes Insight/Teradata, 2015).

### **Risikominimierung**

Verbesserte Steuerungsmöglichkeiten von Prozessen und Verfahren, darunter auch die automatische Anpassung von Materialflüssen in Route und Menge durch Echtzeitdaten, senken nicht nur Kosten, sondern reduzieren auch Risiken. Big Data kann dadurch zur Risikominimierung beitragen, dass es mithilfe von Datenanalysen zu einer höheren Objektivität bei kritischen Entscheidungen kommt. Je mehr Daten zu einem Sachverhalt gesammelt werden, desto rea-

listischer und präziser kann ein Problem erfasst und eine mögliche Lösung gefunden werden. Eine datengestützte Entscheidung tritt an die Stelle von Entscheidungen, die auf subjektiven Empfindungen beruhen und oft verzerrt sind. Auch Compliance-Fälle wie die Einschätzung der Kreditwürdigkeit sind auf einer breiteren Datenbasis präziser durchführbar.

Big-Data-Auswertungen im Rahmen von Predictive Analytics ermöglichen eine genauere Beschaffungsplanung unter der Berücksichtigung von Absatzschwankungen oder unter einer verbesserten Absatzvorhersage. Dadurch wird das Risiko einer Überproduktion reduziert. Auf der anderen Seite kann auch eine Out-of-Stock-Situation („Regallücke“) eher vermieden werden. Die Unternehmensberatung McKinsey (2017, 33) geht davon aus, dass durch den Einsatz von Big Data und speziell künstlicher Intelligenz im Supply-Chain-Management Vorhersagefehler um 20 bis 50 Prozent reduziert werden können. Durch diese Risikominimierung können Umsatzverluste durch die fehlende Verfügbarkeit von Produkten um bis zu 65 Prozent sinken. Insgesamt ließe sich der Warenbestand um 20 bis 50 Prozent verringern.

In der risikobehafteten Landwirtschaft mit ihrer ausgeprägten Saisonalität und Abhängigkeit von Wetter-, Boden-, Wasser- und Nährstoffdynamik ermöglicht das Big-Data-basierte Precision Farming eine bessere Planung und Risikoreduktion. Precision Farming umfasst unter anderem variable Ausbringungstechniken und Fernerkundung. Dadurch kann eine teilflächenspezifische Verteilung von Betriebsmitteln wie Saatgut und Dünger erfolgen, was Erträge steigert und das Risiko von Ernteausfällen verringert.

Auch in der Logistik können natürliche Risiken wie Unwetter mittels Big Data besser in der Planung berücksichtigt werden. Politische und gesellschaftliche Veränderungen können durch die Erstellung von Szenarien ebenfalls erfasst werden, sodass sich frühzeitig strategische und operative Maßnahmen einleiten lassen. Beispielsweise können Prognosen zu politischen Lagen erstellt werden, damit Lieferausfälle und -engpässe verhindert werden können (Uckelmann et al., 2018). Ein Hedgefonds in den USA etwa kalkuliert die potenziellen Folgen von 500 „Doomsday Szenarios“ („Weltuntergangsszenarien“) pro Tag, um das potenzielle Risiko geopolitischer und sonstiger Konflikte einschätzen zu können (Desjardins, 2017).

Insgesamt dürfte die Risikoreduktion durch Big Data noch schwieriger zu quantifizieren sein als die Effekte der Kostenreduktionen. Die Praxisbeispiele zeigen, dass vielfältige Effekte realisierbar sind.

### **Umsatzsteigerungen**

Big Data ermöglicht Unternehmen eine weitgehend individuelle Kundenansprache (Microtargeting). Das Kundenverhalten kann mittels Datenanalysen granular ausgewertet werden, sodass Angebote im sogenannten Behavioral Advertising genau auf den Kunden zugeschnitten werden können (Acquisti/ Varian, 2005). Außerdem können Kundenwünsche datenbasiert besser antizipiert werden. Die Reaktionszeiten auf sich verändernde Marktsignale werden so verkürzt, die Marketingkosten können sinken. Daneben kann der Umsatz durch die Ansprache neuer Kundengruppen und die effizientere Adressierung bestehender Kundengruppen gesteigert werden. Datenschutzbestimmungen wie die Europäische Datenschutz-Grundverordnung (DSGVO) bestimmen die Grenzen dieser Auswertungen. Auch für diese Wirkungsweise spielen Predictive Analytics eine große Rolle. Beispielsweise schlagen Amazon und Netflix dem Nutzer aufgrund seiner Klick- und Konsumhistorie Produkte vor. Experten schätzen, dass 35 Prozent des Umsatzes von Amazon dadurch generiert werden, dass Amazon die Kundenwünsche präzise antizipieren kann (Desjardins, 2017).

Im Einzelhandel sorgen Kundentreueprogramme und entsprechende Bonuskarten für zahlreiche Informationen über das Kaufverhalten der Kunden und ermöglichen eine individuelle Anpassung des Angebots. Auch individualisierte und dynamische Pricing-Strategien werden auf Basis der Kundendatenauswertung erstellt. Hierdurch können Unternehmen nicht nur die verkauften Mengen erhöhen, sondern auch die individuellen maximalen Zahlungsbereitschaften abschöpfen und so den Umsatz steigern. Allerdings ist individualisiertes Pricing aus Verbraucherschutzrechtlicher Sicht nicht unkritisch. Diese Wirkungsweise von Big Data ähnelt sich über Branchen und Unternehmen hinweg und ist in ihrer Ausprägung nicht so vielfältig wie die der Risikominimierung und der Kosteneinsparungen. Dennoch ist sie für Unternehmen ein wichtiger Hebel zur Ertragssteigerung. Bei vielen ist diese kundenseitige Wirkungsweise auch einer der bekanntesten Big-Data-Effekte (zum Beispiel Heidkamp, 2016). Dies liegt mit daran, dass die Verarbeitung personenbezogener Daten in diesem Zusammenhang in der Öffentlichkeit breit diskutiert wird.

## **Veränderte Geschäftsmodelle**

Schließlich ermöglichen Big-Data-Anwendungen neue Geschäftsmodelle und Produkte, die entweder Daten als Produkt oder als Hauptbestandteil eines Produkts nutzen. Unternehmen können beispielsweise ihr Produktportfolio um datenzentrierte Dienstleistungen (Smart Services) erweitern, die an ihre bestehenden Produkte gekoppelt sind. Es sind aber auch Geschäftsmodelle denkbar, die keine Verbindungen zu dem bestehenden Geschäftsmodell haben und daher für das jeweilige Unternehmen neue datengetriebene Geschäftsmodelle darstellen.

Es gibt immer mehr Unternehmen, deren Geschäftsmodell sich vollständig auf Big Data stützt. So ist das Silicon-Valley-Unternehmen Palantir damit erfolgreich geworden, Algorithmen zu entwickeln, die terroristische Gefahren auf Basis von Kommunikationsdaten zwischen Individuen und anderen Daten vorhersagen (Einav/Levin, 2014). In der Autoversicherung sind Telematiktarife möglich, die den Beitragssatz an das Fahrverhalten des Versicherungsnehmers anpassen.

Generell sind auf Big Data fußende Geschäftsmodelle leicht skalierbar. Die Vervielfältigungskosten sind gering und die fortlaufende Weiterentwicklung ist möglich. Diese Wirkungsweise ist deshalb auch am wenigsten statistisch einzuschätzen. Big Data kann in kurzer Zeit Geschäftsmodelle zutage bringen, die aus heutiger Sicht undenkbar sind.

## **Zusammenfassung**

Insgesamt wird Big-Data-Anwendungen durch ihre Auswirkungen auf Unternehmen ein sehr großes Potenzial zugeschrieben, die Bruttowertschöpfung zu steigern. Allerdings lassen sich die tatsächlichen Effekte von Big Data auf die Bruttowertschöpfung kaum präzise quantifizieren, was unter anderem aus dem Querschnittscharakter von Big Data resultiert (vbw, 2016). Die von zahlreichen Studien errechneten Effekte bieten allenfalls Anhaltspunkte.

Dass Big Data tatsächlich auf die beschriebenen Weisen auf die Unternehmen wirkt, zeigt unter anderem eine Studie von KPMG/Bitkom (2017): Sechs von zehn Unternehmen, die Big-Data-Lösungen einsetzen, erreichen mit diesen mindestens eins ihrer Geschäftsziele. 41 Prozent gaben an, sie hätten Risiken

minimiert, 27 Prozent haben den Umsatz erhöht und 19 Prozent Kosten reduziert. Die Unternehmen nutzen die Datenanalysen vor allem, um ihre Geschäftsentwicklung zu überwachen (40 Prozent), ihre Kunden besser kennenzulernen (37 Prozent) und um das eigene Personal effizienter einsetzen zu können (33 Prozent). 21 Prozent nutzen Analysen, um Geschäftsrisiken zu analysieren und zu managen.

### 3.2 Empirische Validierung der Chancen und Herausforderungen von Big-Data-Analysen

Um die im vorherigen Kapitel skizzierten möglichen Wirkungsweisen von Big Data auf Unternehmen über veränderte Produktionsbedingungen und marktseitige Veränderungen empirisch zu validieren, wurden im Rahmen des IW-Zukunftspanels 1.235 Unternehmen aus Branchen der Industrie und der industrienahen Dienstleistungen zu Big-Data-Analysen in ihrem Unternehmen befragt.

#### Datengrundlage

Die verwendete Stichprobe basiert auf dem IW-Zukunftspanel 2018. Das Zukunftspanel ist eine regelmäßige repräsentative Befragung der Geschäftsführer von Industrie- und Dienstleistungsunternehmen in Deutschland. Die Umfrage erfasst allgemeine Daten (zum Beispiel Umsatz, Branche, Alter des Unternehmens und Führungsstruktur) und Antworten auf Fragen zur Internationalisierung, zu Forschung und Entwicklung, Innovationen, zum Marktumfeld, zu Lage- und Zukunftseinschätzungen sowie zu Erfolg und Erfolgsfaktoren.

Die Stichprobe als solche ist nicht repräsentativ. Große Unternehmen sind relativ häufig vertreten, denn um auch über dieses Subsample Aussagen treffen zu können, ist eine hinreichend große Fallzahl erforderlich. Diese Verzerrung fällt bei den verwendeten Regressionsanalysen allerdings nicht ins Gewicht, weil bei diesen um Branche und Größe kontrolliert wird.

Das Branchenspektrum der befragten Unternehmen ist breit. Ein Viertel der 1.235 befragten Unternehmen ist der Metall- und Elektroindustrie zuzuordnen, knapp ein Fünftel der Branche der unternehmensnahen Dienst-

leistungen wie Informationsdienstleistungen, Unternehmensberatung, Wirtschaftsprüfung, Forschung und Entwicklung und Marketing. Als industriennahe Branchen sind neben den unternehmensnahen Dienstleistungen auch Logistik und Großhandel (14 Prozent) sowie Medien und Informations- und Kommunikationstechnologie (6 Prozent) vertreten. Außerdem stammen die Unternehmen aus den Branchen Sonstige Industrie (ohne Bergbau), Bau, Chemie/Pharma/Kunststoff und Maschinenbau. Die meisten der Unternehmen haben weniger als 50 Mitarbeiter (68 Prozent), 32 Prozent sogar weniger als zehn, sind also eher als kleine Unternehmen zu betrachten. Ein Zehntel der Unternehmen hat mindestens 250 Mitarbeiter. Das Gros der Unternehmen, die sich zu ihren Umsätzen äußerten, erwirtschaftet einen Jahresumsatz von weniger als 50 Millionen Euro (88 Prozent), 27 Prozent einen Jahresumsatz von bis zu 1 Million Euro. Zwischen Mitarbeiterzahl und Umsatz besteht mit einem Korrelationskoeffizienten von 0,7 ein deutlicher Zusammenhang.

Für die Analyse der Rolle von Big Data in diesen Unternehmen ist außerdem relevant, dass die Unternehmen hinsichtlich des Grads ihrer Digitalisierung eingeordnet werden. Die digitale Reife der Unternehmen wurde anhand der Stärke der Digitalisierung der Unternehmensabläufe und -produkte abgefragt. 1.177 der befragten Unternehmen machten Angaben dazu, ob sie ihre Produkte, Prozesse oder Werkzeuge ganz, teilweise oder gar nicht virtualisiert haben und mit digitalen Modellen arbeiten oder ob ihre Geschäftsmodelle auf Datenmodellen, Datenanalysen oder Algorithmen basieren. 984 der Unternehmen (84 Prozent von 1.177) sind demnach nicht als digitalisiert einzustufen, sondern lediglich als computerisiert. Sie nutzen die Digitalisierung also eher nur stellenweise und unterstützend. Eine deutliche Minderheit von 16 Prozent (193 Unternehmen) ist digitalisiert.

Um die Digitalisierung des Unternehmens in Bezug auf die angebotenen Produkte zu bestimmen, wurde die Verteilung des Gesamtumsatzes auf rein digitale Waren und Dienstleistungen (zum Beispiel Software/-updates, Datenmodelle/Datenanalyse, Webdesign), auf Produkte mit digitalen Komponenten sowie auf nicht digitale Produkte abgefragt. Rein nicht digitale Unternehmen sind demnach solche, deren Umsatz zu 100 Prozent auf nicht digitale Produk-

te entfällt. Dazu zählt ein gutes Drittel der Unternehmen. Nur 4 Prozent der befragten Unternehmen erzielen ihren Umsatz fast vollständig mit digitalen Produkten. Bei gut einem Zehntel der Unternehmen liegt der Umsatzanteil digitaler Produkte bei mehr als 50 Prozent. Insgesamt bieten die befragten Unternehmen nur in geringem bis mittlerem Maße digitale Produkte und Dienstleistungen an.

### **Verbreitung von Big-Data-Analysen in deutschen Unternehmen**

Im Rahmen der Befragung wurde Big Data wie folgt definiert: „Big Data sind Daten, die aufgrund ihrer Größe, Komplexität, (Un-)Strukturiertheit oder Schnelllebigkeit nicht mit herkömmlichen Methoden und Technologien der Datenanalyse bearbeitet werden können.“ Diese breite Definition trägt dem Umstand Rechnung, dass Big-Data-Technologien einen starken Querschnittscharakter haben und teils sehr unterschiedliche Verfahren und Methoden beinhalten (vbw, 2016, 13). Dabei geht es im Rahmen dieser Analyse ausschließlich um nicht personenbezogene Daten, also zum Beispiel maschinengenerierte Daten.

Die Antworten auf die allgemein gehaltene Frage „Hat sich Ihr Unternehmen schon mit der Thematik Big Data befasst?“ deuten darauf hin, dass viele Unternehmen in den Industriebranchen bislang wenig mit der Analyse großer Datensätze, etwa in Echtzeit, zu tun hatten. Lediglich etwa 8 Prozent der befragten Unternehmen geben an, bereits Big-Data-Analysen durchzuführen (Abbildung 8). Das sind 7 Prozentpunkte weniger als im deutschen Durchschnitt, der von Eurostat (2019) ermittelt wurde. Beide Werte sind gering, wenn man die Potenziale von Big Data und den Fortschritt der Digitalisierung betrachtet. Immerhin gut 11 Prozent planen derartige Analysen. Die restlichen rund 81 Prozent der befragten Unternehmen führen aus verschiedenen Gründen keine Analysen durch. Die Ergebnisse decken sich mit denen einer Studie der Commerzbank (2018): Demnach analysieren 8 Prozent der befragten Unternehmen ihre Daten systematisch. Laut einer älteren repräsentativen Umfrage des Zentrums für Europäische Wirtschaftsforschung (ZEW, 2015, 4) unter 4.500 Unternehmen in Deutschland wenden 18 Prozent Big-Data-Analysen an.

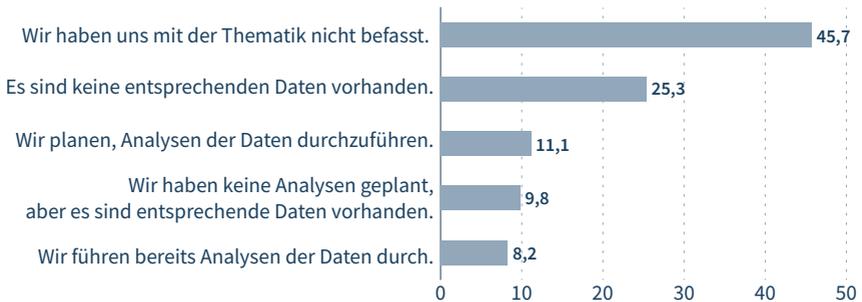
Unterteilt man die Unternehmen in solche, die Big-Data-Analysen durchführen oder dies zumindest planen, und solche, die dies nicht tun, ergibt sich bezüg-

## Big-Data-Analysen in deutschen Unternehmen

Abbildung 8

Antworten von 1.235 im Jahr 2018 befragten Unternehmen auf die Frage:

„Hat sich Ihr Unternehmen schon mit der Thematik Big Data befasst?“, in Prozent



Rundungsdifferenzen.

Daten: <http://dl.iwkoeln.de/index.php/s/9c83ETq2bMzENgz>

Quellen: IW-Zukunftspanel 2018, 1. Befragungswelle; eigene Berechnungen

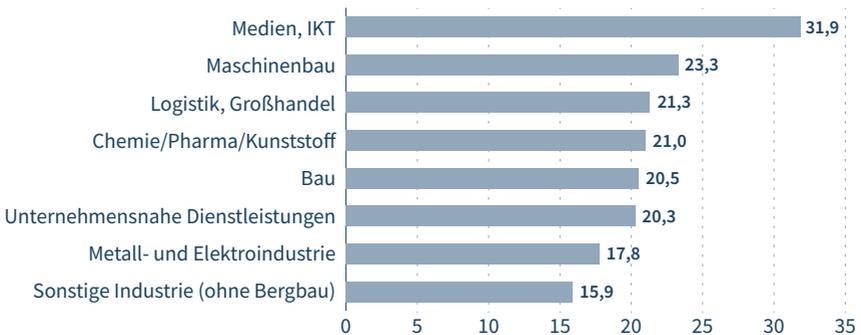
lich der Branchen kein eindeutiges Bild (Abbildung 9). Spitzenreiter bei der Big-Data-Analyse ist die Branche der Medien und Informations- und Kommunikationstechnologie (IKT) – knapp 32 Prozent der Unternehmen arbeiten bereits mit Big Data oder planen dies. Die restlichen Branchen liegen mit rund 20 Prozent Anwendern relativ gleichauf, das Schlusslicht bildet die sonstige Industrie mit fast 16 Prozent. Auffällig ist, dass in keiner der Branchen mehrheitlich Big-Data-Analysen durchgeführt werden oder in Planung sind. Dies deutet auf ein großes Entwicklungspotenzial hin. Laut Expertengesprächen (vbw, 2016, 63) werden Big-Data-Technologien vor allem in den traditionell sehr datenintensiven Branchen wie im Finanzwesen, in der Automobilindustrie, der Logistik oder im Online-Einzelhandel genutzt. Im Medienbereich und im Mittelstand und hierbei besonders im Produzierenden Gewerbe werden Potenziale bisher nur unzureichend erschlossen.

Eine Betrachtung der deskriptiven Statistiken nach Unternehmenseigenschaften (Abbildung 10) zeigt, dass vor allem Unternehmen mit vielen Mitarbeitern und hohem Umsatz bereits im Bereich Big Data aktiv sind oder dies zumindest planen. Der Anteil der Anwender steigt mit beiden Kennzahlen. Auch laut einer repräsentativen Umfrage des ZEW (2015, 4) ist Big Data überwiegend eine Domäne von Großunternehmen.

## Big-Data-Analysen nach Branche

Abbildung 9

Anteil der Unternehmen der jeweiligen Branche im Jahr 2018, die Big-Data-Analysen durchführen oder dies planen, in Prozent



IKT: Informations- und Kommunikationstechnologie; N = 1.235 Unternehmen.

Daten: <http://dl.iwkoeln.de/index.php/s/FcpgQm4FfyQp5So>

Quellen: IW-Zukunftspanel 2018, 1. Befragungswelle; eigene Berechnungen

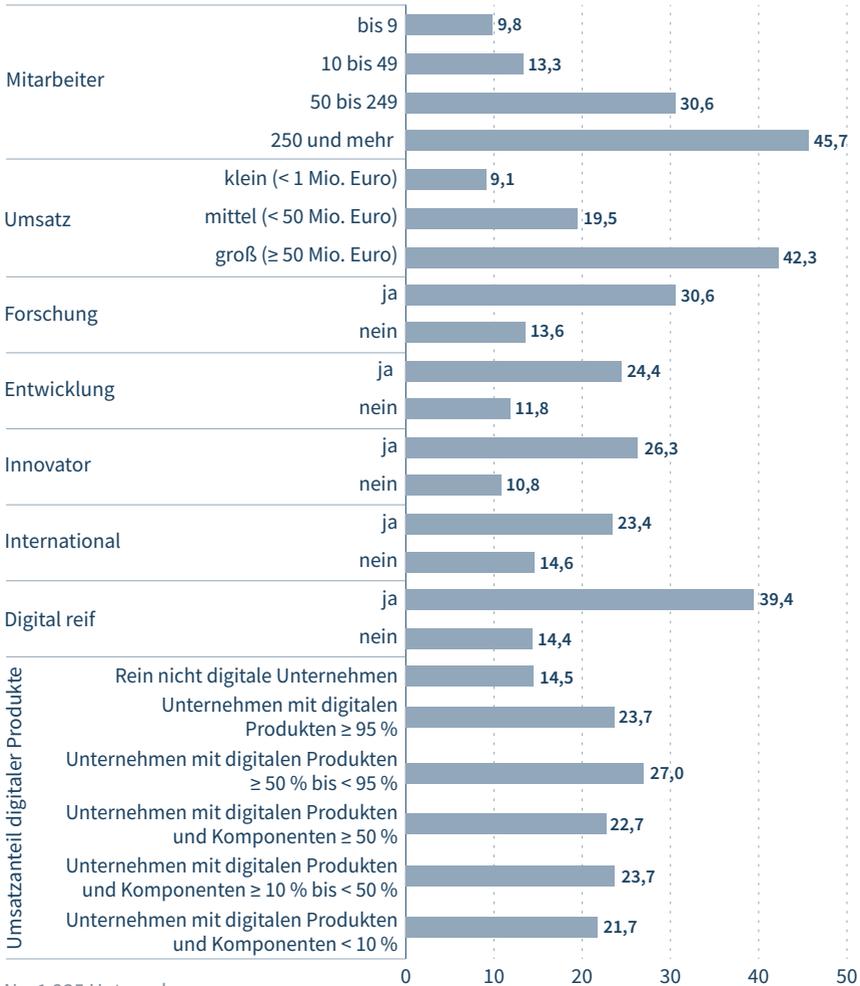
Außerdem ist der Anteil der Big-Data-Anwender unter Unternehmen, die in der Forschung und Entwicklung tätig sind, höher als unter Unternehmen, die sich nicht in diesen Bereichen engagieren. Auch international wirtschaftende Unternehmen sind eher Anwender als nur national tätige Firmen. Insgesamt beschäftigen sich als fortschrittlich geltende Unternehmen eher mit Big Data: Unter den Innovatoren gibt es mehr Anwender als unter den Nichtinnovatoren. Von den digitalisierten beziehungsweise digital reifen Unternehmen sind es sogar knapp 40 Prozent der Unternehmen, die sich mit Big Data befassen – im Vergleich zu nur 14 Prozent unter denjenigen Unternehmen, die sich selbst nicht als digitalisiert, sondern als computerisiert einordnen. Digitalisierung scheint demnach mit Big-Data-Analysen einherzugehen.

Dies gilt indes nicht für digitale Produkte: Unternehmen, die einen höheren Anteil ihres Umsatzes mit digitalen oder teilweise digitalen Produkten machen, beschäftigen sich nicht häufiger mit Big Data als solche mit einem niedrigeren Umsatzanteil. Die Anteile der Anwender bewegen sich zwischen 21 und 24 Prozent. Ausreißer gibt es nach unten durch die Unternehmen, die rein nicht digitale Produkte anbieten – von ihnen befassten sich nur knapp 15 Prozent mit Big Data. Ein kleiner Ausreißer nach oben ist den Unternehmen zuzurechnen, die zwischen 50 und 95 Prozent ihres Umsatzes mit digitalen Produkten ma-

## Big-Data-Analysen nach Unternehmenseigenschaft

Abbildung 10

Anteil der Unternehmen im Jahr 2018 mit der jeweiligen Eigenschaft, die Big-Data-Analysen durchführen oder dies planen, in Prozent



N = 1.235 Unternehmen.

Rein nicht digitale Unternehmen: Unternehmen mit rein nicht digitalen Produkten;  
 Unternehmen mit digitalen Produkten: Unternehmen, bei denen digitale Produkte für über 95 Prozent oder zwischen 50 und 95 Prozent des Umsatzes sorgen; Unternehmen mit digitalen Produkten und Komponenten: Unternehmen, bei denen digitale Produkte und Produkte mit digitalen Komponenten für über 50 Prozent, zwischen 10 und 50 Prozent oder unter 10 Prozent des Umsatzes sorgen; bei Überschneidungen dominieren rein digitale Produkte den Mix aus rein digitalen und teilweise digitalen Produkten (digitale Komponenten).

Daten: <http://dl.iwkoeln.de/index.php/s/9MB3GbXX2T5xfow>

Quellen: IW-Zukunftspanel 2018, 1. Befragungswelle; eigene Berechnungen

chen. Sie beschäftigen sich zu 27 Prozent mit Big Data. Digitale Produkte erfordern zumindest derzeit und auf die befragten Unternehmen bezogen keine Big-Data-Analysen.

Insgesamt ist Big Data für die befragten Unternehmen noch ein relativ neues Phänomen. Mehr als die Hälfte der 101 Unternehmen, die angeben, bereits Big-Data-Analysen durchzuführen, hat damit ab 2015 begonnen. Ein Fünftel hat jedoch bereits vor 2011 damit angefangen. An dieser Stelle kommen unter Umständen auch unterschiedliche Definitionen von Big Data und Big-Data-Analysen zum Tragen. Die Auswertung großer Datensätze ist auch mit älteren Technologien möglich; neuartige Big-Data-Anwendungen erfordern jedoch Technologien, die erst in den vergangenen Jahren auf den Markt gekommen sind. Es ist dennoch davon auszugehen, dass die Unternehmen, die sehr früh mit der Analyse großer Datensätze begonnen haben, sich entsprechend weiterentwickelt haben und auch neuartige Big-Data-Anwendungen nutzen.

Im Folgenden werden die deskriptiven Statistiken durch eine Regressionsanalyse ergänzt. Daraus ergibt sich ein umfassenderes Bild darüber, welche Faktoren und Unternehmenseigenschaften es begünstigen, dass Unternehmen in Deutschland sich mit Big Data beschäftigen. Aufgrund der geringen Zahl an Unternehmen, die bereits Big-Data-Analysen durchführen (N = 101), ist diese Regressionsanalyse mit starken Einschränkungen behaftet. Um dies abzumildern und die Fallzahl zu erhöhen, wurde im Folgenden zwischen Unternehmen, die Big-Data-Analysen durchführen oder dies planen (N = 238), sowie Unternehmen, die diese weder durchführen noch planen (N = 997), unterschieden.

Tabelle 2 zeigt die Ergebnisse einer logistischen Regressionsanalyse. Abhängige Variable ist die binäre Variable, dass Big-Data-Analysen bereits durchgeführt werden oder zumindest in Planung sind (1, wenn dies der Fall ist; 0, wenn dies nicht der Fall ist). Die Kontrollvariablen, allesamt Dummies, die den Wert 1 annehmen, falls das Unternehmen in die jeweilige Kategorie fällt, sind die Mitarbeiterzahl (verschiedene Größenkategorien), der Umsatz (verschiedene Größenkategorien), die Branchenzugehörigkeit, ob Unternehmen in der Forschung und Entwicklung tätig sind, ob sie als Innovatoren gelten, international tätig sind, digital reif sind und welcher Anteil ihres Umsatzes auf digitale Produkte entfällt (verschiedene Größenkategorien). Die jeweils nicht aufgeführte Kategorie ist die

## Durchführung von Big-Data-Analysen

Tabelle 2

Ergebnisse einer logistischen Regression inklusive Odds-Ratio<sup>1)</sup>

Unabhängige Variable	Odds-Ratio	Standardfehler
<b>Mitarbeiter</b> (Referenz: bis 9 Mitarbeiter)		
10 bis 49	1,407	0,469
50 bis 249	2,537**	0,894
250 und mehr	3,742**	1,640
<b>Umsatz</b> (Referenz: bis 1 Million Euro; klein)		
mittel (1 bis unter 50 Millionen Euro)	1,411	0,492
groß (mindestens 50 Millionen Euro)	1,699	0,773
<b>Branche</b> (Referenz: Chemie/Pharma/Kunststoff)		
Metall- und Elektroindustrie	0,842	0,294
Sonstige Industrie (ohne Bergbau)	1,190	0,485
Bau	0,457	0,220
Logistik, Großhandel	1,538	0,598
Medien, IKT	1,203	0,614
Unternehmensnahe Dienstleistungen	0,993	0,393
Maschinenbau	0,955	0,399
<b>Forschung</b>	1,625*	0,358
<b>Entwicklung</b>	1,202	0,299
<b>Innovator</b>	1,679**	0,354
<b>International</b>	0,817	0,173
<b>Digital reif</b>	1,596***	0,172
<b>Umsatzanteil digitaler Produkte und Komponenten</b> (Referenz: Unternehmen mit rein nicht digitalen Produkten)		
Unternehmen mit digitalen Produkten $\geq 95$ %	1,157	0,667
Unternehmen mit digitalen Produkten $\geq 50$ % bis $< 95$ %	1,944*	0,760
Unternehmen mit digitalen Produkten und Komponenten $\geq 50$ %	1,222	0,326
Unternehmen mit digitalen Produkten und Komponenten $\geq 10$ % bis $< 50$ %	1,498	0,372
Unternehmen mit digitalen Produkten und Komponenten $< 10$ %	1,333	0,408

\*\*\*/\*\*/\*: signifikant auf dem 1-/5-/10-Prozent-Niveau. Abhängige Variable: Big-Data-Analysen werden bereits durchgeführt oder das Unternehmen plant, dies zu tun; N = 971.

1) Die Odds-Ratio ist ein Assoziationsmaß, bei dem zwei Odds miteinander verglichen werden. Mathematisch berechnen sich Odds als Quotienten aus der Wahrscheinlichkeit, dass ein Ereignis eintritt, und der Wahrscheinlichkeit, dass es nicht eintritt. Die Odds-Ratio für Big-Data-Analysen ist ein Maß dafür, um wie viel größer die Chance ist, dass ein Unternehmen Big-Data-Analysen durchführt oder diese plant (im Sinne einer Quote), wenn dieses Unternehmen eine bestimmte binäre Kontrollvariable erfüllt (Wert 1; zum Beispiel Unternehmen, die digital reif sind), verglichen mit der Gruppe, für welche die Kontrollvariable den Wert 0 annimmt (Unternehmen, die nicht digital reif sind).

Quellen: IW-Zukunftspanel 2018, 1. Befragungswelle; eigene Berechnungen

Basiskategorie, auf die sich die Schätzergebnisse beziehen. Das bedeutet, dass für den Dummy „Mitarbeiterzahl 10 bis 49“ der Schätzwert verglichen mit der Mitarbeiterzahl der nicht aufgeführten Basiskategorie „bis 9 Mitarbeiter“ dargestellt ist. Um die Ergebnisse nicht nur qualitativ, sondern auch quantitativ zu interpretieren, wurden Odds-Ratios, also Quotenverhältnisse, gebildet. Folgende Faktoren steigern nach den Regressionsergebnissen statistisch signifikant die Wahrscheinlichkeit, dass ein Unternehmen Big-Data-Analysen durchführt oder dies plant (zukünftig durchführt):

- Die Wahrscheinlichkeit erhöht sich mit der Anzahl an Mitarbeitern. Sie steigt im Vergleich zu Unternehmen mit weniger als zehn Mitarbeitern um das 2,5-Fache (ist also 2,5-mal so hoch), wenn ein Unternehmen zwischen 50 und 249 Mitarbeitern hat, und sogar um das 3,7-Fache, wenn ein Unternehmen mehr Mitarbeiter hat. Der Umsatz wirkt sich nicht signifikant auf die Wahrscheinlichkeit aus, dass ein Unternehmen Big Data (zukünftig) nutzt. An dieser Stelle ist davon auszugehen, dass zwischen Umsatz und Mitarbeiterzahl Kollinearitäten vorliegen, die zu diesem Ergebnis führen. Eine Testanalyse, in der nicht für die Mitarbeiterzahl kontrolliert wird, bestätigt den Zusammenhang – dann werden für Umsätze signifikante Werte erzielt: Ein mittlerer Umsatz erhöht die Wahrscheinlichkeit der (zukünftigen) Big-Data-Anwendung um das 2,3-Fache, ein großer Umsatz sogar um das 4,8-Fache. Bezieht man nur die Mitarbeiterzahl in das Schätzmodell ein, ergeben sich signifikante positive Werte. Für Unternehmen mit 10 bis 49 Mitarbeitern steigt die Wahrscheinlichkeit um das 1,7-Fache, für solche mit 50 bis 249 Mitarbeitern um das 3,5-Fache und für solche mit 250 und mehr Mitarbeitern sogar um das 5,8-Fache.
- Wenn ein Unternehmen in der Forschung aktiv ist oder als Innovator gilt, wirkt sich das positiv auf die Wahrscheinlichkeit aus, dass dieses Unternehmen Big Data (zukünftig) nutzt. Die Wahrscheinlichkeit steigt im Vergleich zu Unternehmen, die nicht forschen oder innovativ sind, um das 1,6- beziehungsweise 1,7-Fache.
- Auch die Digitalisierung wirkt sich positiv aus. Digitale Unternehmen haben eine 1,6-fach erhöhte Wahrscheinlichkeit im Vergleich zu nicht digitalen Unternehmen, sich mit Big Data (zukünftig) zu beschäftigen.

- Erwirtschaftet ein Unternehmen mindestens 50 Prozent, aber weniger als 95 Prozent seines Umsatzes mit digitalen Produkten, ist es im Vergleich zu einem Unternehmen mit rein nicht digitalen Produkten fast doppelt so wahrscheinlich, dass dieses Unternehmen sich mit Big Data beschäftigt oder dies plant.

Insgesamt haben also die Mitarbeiterzahl beziehungsweise die Umsatzhöhe und die Digitalisierung eines Unternehmens den größten Einfluss darauf, ob das Unternehmen im Bereich Big Data aktiv ist oder dies plant. Diese Ergebnisse werden auch erzielt, wenn man das Modell ohne die Kontrollvariablen der Branchen schätzt.

Dass digitale Unternehmen sich eher mit Big Data beschäftigen (oder dies planen) als nicht digitale, liegt zumindest teilweise daran, dass ein Faktor der Klassifizierung eines Unternehmens als digital in dieser Analyse die Durchführung von Datenanalysen selbst ist. Dementsprechend erwartbar ist das Ergebnis.

Big Data zeigt sich insgesamt als Thema, das derzeit vor allem von großen und umsatzstarken Unternehmen besetzt wird. Dies mag zum einen daran liegen, dass oft in größeren Unternehmen auch mehr Daten anfallen, etwa weil es mehr Punkte entlang der unternehmensinternen Wertschöpfungskette gibt, an denen Daten aufgezeichnet werden. Zum anderen sind es unter Umständen vor allem die umsatzstarken Unternehmen, die sich Big-Data-Analysen überhaupt leisten können, weil sie die finanziellen und personellen Ressourcen dazu haben (mehr zu den Hemmnissen im Bereich Big Data weiter unten in diesem Kapitel).

### **Zwecke von Big-Data-Analysen in deutschen Unternehmen**

Die 101 Unternehmen aus der Panelumfrage, die bereits Big-Data-Analysen durchführen, wurden nach der Zielsetzung beziehungsweise dem Zweck dieses Befassens mit Big Data gefragt. In Kapitel 3.1 wurde zwischen den Wirkungsweisen von Big Data unterschieden, vor allem zwischen den Kosteneinsparungen und Umsatzsteigerungen, der Risikominimierung sowie den veränderten Geschäftsmodellen. Die in der Panelbefragung abgefragten Zwecke knüpfen an diese Kategorien an, lassen sich aber oftmals nicht eindeutig und trenn-

scharf zuteilen. So sind die abgefragten Zwecke unterschiedlichen Phasen des Produktionsprozesses und auch eines möglichen Veränderungsprozesses zuzuordnen. Eine Kostensenkung etwa kann das Resultat einer Prozessoptimierung sein; daraus kann wiederum eine Umsatzsteigerung resultieren. Insgesamt verfolgen die Unternehmen mit ihren Big-Data-Analysen deutlich mehr als einen Zweck. Im Schnitt nannten die Unternehmen vier verschiedene Zwecke. Lediglich drei Unternehmen gaben nur einen Zweck an, 14 Unternehmen gaben sieben verschiedene Zwecke an.

Der am häufigsten genannte Zweck von Big-Data-Analysen ist die Prozessoptimierung. Über 83 Prozent der befragten Unternehmen, die Big-Data-Analysen durchführen, geben diese Zielsetzung an (Abbildung 11). Wachstums- oder Umsatzsteigerung sind für knapp 72 Prozent der Unternehmen ein Zweck der Big-Data-Analysen. Die Kostensenkung ist mit fast 62 Prozent schon deutlich weniger häufig genannt als die Prozessoptimierung, allerdings ist davon auszugehen, dass beide korreliert sind.

Herkömmliche Datenanalysezwecke wie Wettbewerbs- oder Trendanalyse und Marketing nennen rund 56 beziehungsweise 52 Prozent der Unternehmen. Die Entwicklung neuer Geschäftsmodelle streben immerhin rund 56 Prozent der

### Zwecke von Big-Data-Analysen

Abbildung 11

Anteil der Unternehmen im Jahr 2018, die für ihre Big-Data-Analysen den jeweiligen Zweck angeben, in Prozent



N = 101; Mehrfachantworten möglich.

Daten: <http://dl.iwkoeln.de/index.php/s/GdimSwEfFYDSCRT>

Quellen: IW-Zukunftspanel 2018, 1. Befragungswelle; eigene Berechnungen

Unternehmen an, das heißt, Big Data wird durchaus auch als Möglichkeit gesehen, sich neue Geschäftsfelder zu erschließen und sich als Unternehmen weiterzuentwickeln. Individualisierte Produkte, ein zentraler Nutzen der Industrie 4.0, wird nur von gut 47 Prozent der Unternehmen genannt und ist somit der am wenigsten verfolgte konkrete Zweck (neben den zusammengefassten „sonstigen Zwecken“). Dies ist insofern sachlogisch, als individualisierte Produkte vor allem auf die Präferenzen des einzelnen Endverbrauchers abstellen, wofür ein Datensatz benötigt wird statt viele. Insgesamt nutzen die befragten Unternehmen Big-Data-Analysen vor allem, um interne Verbesserungen zu erreichen.

Die Ergebnisse sind mit denen von Capgemini (2018) vergleichbar. Laut der Studie unter 121 Unternehmen aus Deutschland, Österreich und der Schweiz wird Big Data hauptsächlich genutzt, um die Effizienz zu erhöhen (Capgemini, 2018, 25). Dabei stehen kundenzentrierte Prozesse im Vordergrund. Interne Prozesse im Personal- und Finanzbereich oder rund um Produktion und Lieferkette werden von etwas weniger Unternehmen mithilfe von Big Data optimiert. Das Aufspüren und Umsetzen neuer Geschäftsmodelle wird von gut 58 Prozent der Unternehmen verfolgt, erreicht also ähnliche Werte wie unter den befragten Unternehmen der vorliegenden Analyse.

Tabelle 3 enthält die Ergebnisse von Chi-Quadrat-Tests über die Gleichheit der Mittelwerte der Antwortwerte unterschiedlicher Subsamples der befragten Unternehmen. Diese Tests können Unterschiede zwischen Unternehmen je nach Unternehmenseigenschaft aufdecken. Damit soll geklärt werden, ob bestimmte Zwecke für Unternehmen mit bestimmten Eigenschaften eine höhere oder niedrigere Relevanz haben als für andere. Dafür wurden die abgefragten Unternehmenscharakteristika (außer der Branche) so definiert, dass es jeweils zwei Kategorien pro Variable gibt. Eine Aufteilung der Variablen in mehr als zwei Kategorien hätte im Rahmen der Mittelwerttests zu sehr kleinen Subsamples geführt. Statt zwischen vier verschiedenen Mitarbeitergrößen zu unterscheiden, wurden die Unternehmen in solche mit unter 49 und solche mit mindestens 50 Mitarbeitern (viele Mitarbeiter) unterteilt. Diese Grenze wurde auch gewählt, um in den Größenklassen eine hinreichend hohe Unternehmenszahl zu erhalten. Außerdem wurde zwischen Unternehmen mit kleinem und solchen mit mittlerem oder großem Umsatz (hoher Umsatz) unter-

## Zwecke von Big-Data-Analysen nach Unternehmenseigenschaft

Tabelle 3

Unternehmens-eigenschaft	Kosten-senkung	Wachstums-oder Umsatz-steigerung	Prozess-optimie-rung	Individua-lisierte Produkte	Entwick-lung neuer Geschäfts-modelle	Wettbe-werbs-oder Trend-analyse	Marke-ting
Forschung	0,224**	0,057	<b>0,048</b>	0,085	0,112	0,113	0,016
Entwicklung	0,273***	0,076	<b>0,181**</b>	0,127	0,218**	0,146	-0,020
Innovator	0,076	0,203**	<b>-0,033</b>	0,183*	0,317***	0,192*	0,315***
International	0,208**	0,146	<b>-0,011</b>	0,064	0,136	0,072	0,203**
Digital reif	0,023	0,082	<b>0,057</b>	0,088	0,163	0,042	0,057
Viele Mitarbeiter	0,291***	0,208**	<b>0,084</b>	-0,004	0,084	0,081	0,142
Hoher Umsatz	0,396***	0,150	<b>-0,015</b>	-0,164	-0,047	0,295**	-0,067
Unterneh-men mit digitalen Produkten	0,025	0,053	<b>0,096</b>	0,212*	0,113	-0,195	0,038

\*\*\*/\*\*/\*: signifikant auf dem 1-/5-/10-Prozent-Niveau; Chi-Quadrat-Tests; Differenz der Mittelwerte verschiedener Subsamples, je nachdem, ob die Zeilenvariablenausprägung 1 oder 0 ist (das Unternehmen in die jeweilige Kategorie fällt oder nicht); Faltungen zeigen den jeweils meistgenannten Zweck je nach Unternehmenseigenschaft (ersichtlich aus hier nicht aufgeführten deskriptiven Statistiken).

Quellen: IW-Zukunftspanel 2018, 1. Befragungswelle; eigene Berechnungen

schieden. Als Unternehmen mit digitalen Produkten werden solche betrachtet, die mindestens 10 Prozent ihres Umsatzes mit digitalen Produkten erzielen. Aus der deskriptiven Statistik ergibt sich, dass für alle positiven Unternehmensausprägungen (das Unternehmen forscht und entwickelt, ist Innovator, agiert international etc.) die Prozessoptimierung der meistgenannte Zweck ist.

Die ökonomische Analyse zeigt darüber hinaus folgende statistisch signifikanten Unterschiede (Tabelle 3):

- Forschende Unternehmen bezwecken mit Big-Data-Analysen eher als nicht forschende Unternehmen, ihre Kosten zu senken. Dies gilt auch für Firmen, die in der Entwicklung tätig sind. Diese nennen außerdem die Zwecke Prozessoptimierung und Entwicklung neuer Geschäftsmodelle häufiger als nicht entwickelnde Unternehmen.

- Als innovativ geltende Unternehmen verfolgen einige der abgefragten Zwecke deutlich häufiger als nicht innovative Unternehmen: Wachstums- oder Umsatzsteigerung, individualisierte Produkte, Entwicklung neuer Geschäftsmodelle, Wettbewerbs- oder Trendanalyse und Marketing. Innovative Unternehmen führen ihre Big-Data-Analysen mit vielfältigen Zielsetzungen durch.
- International tätige Unternehmen beschäftigen sich häufiger als nicht internationale Unternehmen mit Big Data, um ihre Kosten zu senken und um ihr Marketing zu verbessern. Unter Umständen hilft gerade ihnen eine Vielzahl von Daten, ihr Marketing auf verschiedene Länder mit unterschiedlichen Kundenpräferenzen auszurollen.
- Für Unternehmen mit vielen Mitarbeitern sind die Kostensenkung sowie Wachstums- oder Umsatzsteigerung deutlich wichtiger bei der Big-Data-Analyse als für Unternehmen mit wenig Mitarbeitern.
- Unternehmen mit hohem Umsatz nennen eher die Kostensenkung als Zweck als Unternehmen mit niedrigem Umsatz. Außerdem sehen sie besonders die Wettbewerbs- oder Trendanalyse eher als Zweck von Big-Data-Analysen als ihre Pendant.
- Unternehmen mit digitalen Produkten nennen individualisierte Produkte eher als Zweck als Unternehmen ohne diese. Dies ist sachlogisch, da viele digitale Produkte personalisiert sind. Digital reife Unternehmen hingegen zeigen keine statistisch signifikanten Unterschiede in der Zwecksetzung im Vergleich zu nicht digitalisierten Unternehmen.

Aus der Teilanalyse kann man den Schluss ziehen, dass gerade die Kostensenkung für umsatzstarke, große, forschende, international tätige Unternehmen ein erklärter Zweck der Big-Data-Analysen ist. Volkswirtschaftlich wäre es wünschenswert, wenn die Unternehmen den Zweck noch mehr in der Entwicklung neuer Geschäftsmodelle sähen. Zahlreiche Studien und Trendanalysen sehen dieses Potenzial von Big Data, es scheint sich aber noch nicht in der Realität manifestiert zu haben. Positiv ist anzumerken, dass Big Data mit vielen Zielsetzungen angewandt wird; die Vielfältigkeit des Einsatzes von Big-Data-Analysen wird also von den Unternehmen wahrgenommen.

## **Hemmnisse für die Durchführung von Big-Data-Analysen**

Trotz der vielfältigen Ziele, die Unternehmen mithilfe von Big Data verfolgen können, führen rund 81 Prozent der befragten Unternehmen derzeit keine Big-Data-Analysen durch und planen dies auch nicht. Um möglichen Hindernissen dafür auf den Grund zu gehen, wurden Unternehmen, die momentan noch keine Big-Data-Analysen anstellen, obwohl entsprechende Daten vorhanden sind, gefragt, welche Faktoren sie davon abhalten.

Fast 62 Prozent der Unternehmen fehlt das nötige Know-how für Big-Data-Analysen (Abbildung 12). Diese Unternehmen haben also keine geeigneten personellen Ressourcen. Es gibt zahlreiche Berichte darüber, dass es im Bereich der Datenanalyse zu wenige Fachkräfte gibt (zum Beispiel Capgemini, 2018; Commerzbank, 2018; KPMG/Bitkom, 2017). Außerdem schließen sich viele Datenanalysten eher originären Technologie-beziehungsweise Digitalunternehmen an und finden sich seltener in der Industrie, geschweige denn im ländlichen Mittelstand.

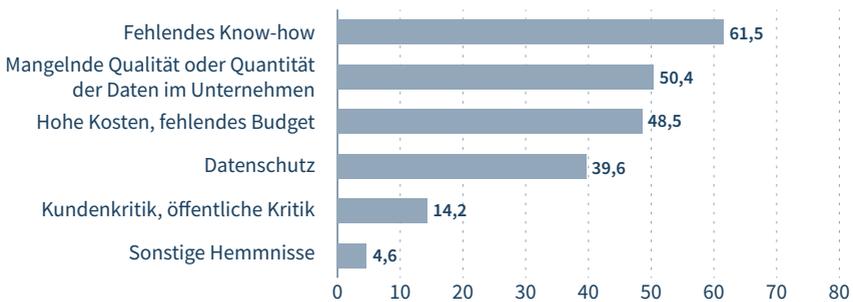
Gut die Hälfte der befragten Unternehmen gibt an, es mangle an der Qualität oder Quantität der Daten. Um Big-Data-Analysen durchführen zu können, werden Datensätze benötigt, die idealerweise an vielen Stellen des Wertschöpfungsprozesses sehr zahlreich auftreten. Daten müssen kontinuierlich und gleichförmig aufgezeichnet werden, um entsprechend auswertbar und belastbar zu sein. In der Regel müssen die Rohdaten zunächst aufwendig aufgearbeitet werden, um in einer Qualität vorzuliegen, die dann auch eine Auswertung ermöglicht. Daran mangelt es vielen deutschen Unternehmen noch (Experian, 2016).

Die Auswertung von Big Data ist oft mit hohen Kosten verbunden (Markl et al., 2013) – ein Hemmnis, das von knapp 49 Prozent der Unternehmen genannt wird. Zum Teil schrecken die hohen Investitionskosten die Unternehmen ab, obwohl sie sich diese leisten könnten. Andere können die hohen Kosten gar nicht stemmen, da ihnen das Budget fehlt, um Big-Data-Analysen durchzuführen. Das kann zum einen die Fixkosten betreffen, etwa die Installation von datenaufzeichnenden Sensoren oder die Investition in Rechnerleistung und Cloudspeicherplatz. Zum anderen können hohe variable Kosten wie Personalkosten für die Datenanalyse anfallen. Zum Vergleich: In der Studie von KPMG/

## Hemmnisse für Big-Data-Analysen

Abbildung 12

Anteil der Unternehmen im Jahr 2018, die den jeweiligen Faktor als Hindernis für Big-Data-Analysen nennen, obwohl entsprechende Daten vorhanden sind, in Prozent



N = 260; Mehrfachantworten möglich.

Daten: <http://dl.iwkoeln.de/index.php/s/QzZqSoH2eYRS67z>

Quellen: IW-Zukunftspanel 2018, 1. Befragungswelle; eigene Berechnungen

Bitkom (2017) sahen sich 53 Prozent der befragten Unternehmen bei der Nutzung von Datenanalysen mit unzureichenden Finanzmitteln und fehlenden personellen Ressourcen konfrontiert.

Datenschutzrechtliche Hemmnisse führen knapp 40 Prozent der Unternehmen an. Diese Zahl ist insbesondere vor dem Hintergrund als sehr hoch einzustufen, dass bei Big Data in dieser Befragung vor allem auf nicht personenbezogene Daten abgestellt wurde. Die Europäische Datenschutz-Grundverordnung (DSGVO) hingegen bezieht sich nur auf personenbezogene Daten. Allerdings herrscht einerseits unter vielen Unternehmen in Deutschland Unsicherheit hinsichtlich des Datenschutzes (Bitkom, 2017; Steier/Duisberg, 2018), weshalb zu vermuten ist, dass sie zum Teil mit überhöhter Vorsicht agieren. Andererseits sind auch nicht personenbezogene Daten oft unternehmensintern mit sehr hohen Datenschutzbestimmungen versehen, etwa um einen Abfluss von erfolgskritischen Daten und Know-how an Wettbewerber und Partner zu verhindern (Lichtblau et al., 2017; Jöns, 2016). Für Unternehmen ist es sehr schwierig einzuschätzen, welche Daten in ihrem Unternehmen besonders schützenswert sind (Filkins, 2017). Es ist anzunehmen, dass Verbesserungen im Bereich der Data Governance, besonders eine Prüfung des notwendigen Datenschutzniveaus für alle Datensätze, das Hemmnis der datenschutzrechtlichen Bestimmungen für Big-Data-Analysen lindern könnte. Bei der Studie von KPMG/

**Hemmnisse für Big-Data-Analysen nach Unternehmenseigenschaft** Tabelle 4

Unternehmens- eigenschaft	Hohe Kosten, fehlendes Budget	Fehlendes Know-how	Mangelnde Qualität oder Quantität der Daten	Kundenkritik, öffentliche Kritik	Daten- schutz
Forschung	0,052	<b>0,043</b>	-0,015	-0,059	-0,102
Entwicklung	0,010	<b>-0,056</b>	0,001	-0,139**	-0,175**
Innovator	0,073	<b>-0,084</b>	0,069	-0,004	-0,100
International	-0,051	<b>-0,018</b>	-0,011	-0,136***	-0,085
Digital reif	-0,085	<b>-0,254***</b>	-0,060	0,016	<b>0,143*</b>
Viele Mitarbeiter	-0,090	<b>0,031</b>	-0,036	-0,057	-0,105*
Hoher Umsatz	-0,007	<b>0,030</b>	0,013	-0,195***	-0,178**
Unternehmen mit digitalen Produkten	-0,011	<b>-0,047</b>	0,011	0,093*	0,109

\*\*\*/\*\*/\*: signifikant auf dem 1-/5-/10-Prozent-Niveau; Chi-Quadrat-Tests; Differenz der Mittelwerte verschiedener Subsamples, je nachdem, ob die Zeilenvariablenausprägung 1 oder 0 ist (das Unternehmen in die jeweilige Kategorie fällt oder nicht); Fettungen zeigen das jeweils meistgenannte Hemmnis je nach Unternehmenseigenschaft (ersichtlich aus hier nicht aufgeführten deskriptiven Statistiken).

Quellen: IW-Zukunftspanel 2018, 1. Befragungswelle; eigene Berechnungen

Bitkom (2017) gaben 61 Prozent der Unternehmen datenschutzrechtliche Bedenken als Hemmnis an.

Kundenkritik oder öffentliche Kritik spielen für fast 86 Prozent der befragten Unternehmen keine Rolle bei Big-Data-Analysen. Das kann damit zusammenhängen, dass eventuelle Kritik wahrscheinlich gerade auf mangelnden Datenschutz abzielen würde – dieser aber von Unternehmen gewährleistet wird.

Auch für die Hemmnisse wird mithilfe von Chi-Quadrat-Tests geprüft, ob sie für Unternehmen mit bestimmten Eigenschaften eine höhere oder niedrigere Relevanz haben als für andere (Tabelle 4). Die deskriptive Analyse ergibt, dass für alle positiven Unternehmensausprägungen außer der Digitalisierung (das Unternehmen forscht und entwickelt, ist Innovator, agiert international etc.) das fehlende Know-how das meistgenannte Hemmnis ist. Das zeigt, dass auch die umsatzstarken, innovativen Unternehmen nicht genug personelle Ressourcen für Big-Data-Analysen haben. Die ökonomische Analyse legt darüber hinaus folgende statistisch signifikanten Unterschiede offen (Tabelle 4):

- Für Unternehmen, die in der Entwicklung tätig sind, sind öffentliche und Kundenkritik und Datenschutzbestimmungen weniger ein Hemmnis als für nicht entwickelnde Unternehmen. Das könnte darauf zurückzuführen sein, dass in der Entwicklung mehr Möglichkeiten gegeben sind, gerade auch mit Big Data zu experimentieren, ohne gleich vor der Öffentlichkeit Zeugnis ablegen zu müssen. Warum sich dies nicht auch auf forschende Unternehmen auswirkt, kann an dieser Stelle nicht geklärt werden.
- Für Unternehmen mit hohem Umsatz sind öffentliche und Kundenkritik und Datenschutzbestimmungen eher weniger ein Problem als für Unternehmen mit geringerem Umsatz. Der Grund dafür könnte sein, dass sie mehrere Umsatzquellen und eine breitere Kundschaft haben. Die Abhängigkeit von einzelnen Kundengruppen ist somit geringer.
- Für internationale Unternehmen sind Kundenkritik und öffentliche Kritik weniger ein Hemmnis als für national agierende Unternehmen. Das kann auch daran liegen, dass in anderen Ländern unterschiedliche Datenschutzpräferenzen herrschen. Kundenkritik und Datenschutzpräferenzen sind an dieser Stelle nicht trennscharf, denn ein Teil der Kundenkritik könnten Befürchtungen sein, der Datenschutz würde nicht gewahrt. In den USA beispielsweise sind die Bestimmungen und auch das öffentliche Bewusstsein für Datenschutz generell niedriger als in Europa (zum Beispiel Brogger, 2018).
- Für digitalisierte Unternehmen ist fehlendes Know-how weniger ein Hemmnis als für nicht digitalisierte. Eine Begründung dafür lautet, dass digitale Unternehmen tendenziell eher über personelle Ressourcen für Datenanalyse und Programmierung verfügen. Dafür sehen digitale Unternehmen eher Datenschutz als Hemmnis als nicht digitale Unternehmen. Insgesamt ist Datenschutz das Haupthemmnis für digitale Unternehmen. Das könnte darauf hindeuten, dass Datenschutz gerade in der Digitalisierung bestimmte Geschäftsmodelle und Prozessoptimierungen behindert. Es könnte auch ein Hinweis darauf sein, dass sich digitale Unternehmen eher über eventuelle Datenschutzprobleme im Klaren sind als nicht digitale.
- Für Unternehmen mit einer ausgeprägten Produktdigitalisierung ist öffentliche und Kundenkritik eher ein Hemmnis als für Unternehmen, die nicht

so viele digitale Produkte verkaufen. Das kann damit zusammenhängen, dass digitale Produkte oft datenintensiver sind und die Kunden daher bei diesen mehr Wert auf Datenschutz legen. Geht es um öffentliche Kritik, ist diese bei digitalen Produkten häufig intensiver. Denn die damit verbundenen neuen Geschäftsmodelle müssen sich oft erst noch etablieren.

Insgesamt fällt auf, dass fehlendes Know-how das zentrale Hemmnis ist – über alle Unternehmen hinweg. Lediglich die digitalen Unternehmen haben damit weniger zu kämpfen. Das fehlende Know-how könnte auch das zweithäufigste Hemmnis der fehlenden Qualität oder Quantität von Daten begünstigen. Oft liegen Datenschätze vor, werden aber nicht von den Unternehmen als solche erkannt oder können nicht genutzt werden, weil sie nicht hinreichend aufgearbeitet werden können. Sie liegen dann zwar als Daten vor, bieten aber noch keine Informationen, die zweckdienlich sind und beispielsweise zur Kostensenkung beitragen.

Gelänge es, die Datensouveränität der Unternehmen durch entsprechende personelle Ressourcen zu stärken, würden sich mehr Unternehmen mit Big Data beschäftigen, als dies derzeit der Fall ist. Bei der Verbreitung von Big Data gibt es noch viel Ausbaupotenzial. Außerdem ist zu erwarten, dass die Anwendungen und Zielsetzungen von Big Data dann vielfältiger und innovativer werden, also auch öfter als derzeit dazu genutzt werden, um neue Geschäftsbereiche und Geschäftsmodelle zu entdecken und entwickeln.

## 4 Fazit

Big Data und darauf basierende Anwendungen und Technologien bieten große Chancen für die Unternehmen in Deutschland. Dass diese Potenziale in der deutschen Wirtschaft bislang nur partiell genutzt werden, zeigt die vorliegende Bestandsaufnahme des Status quo von Big Data in Wirtschaft und Wissenschaft. Die empirische Untersuchung von 1.235 deutschen Unternehmen hat ergeben, dass viele Firmen noch keine Big-Data-Analysen durchführen. Oft ist es das fehlende Know-how, das sie daran hindert, punktuell oder sogar in der Breite auf entsprechende Lösungen zu setzen.

Die stark zunehmende Anzahl an Publikationen mit Bezug auf Big Data allein in der wissenschaftlichen Forschung deutet zudem auf einen entsprechenden Forschungsfokus wie auch Forschungsbedarf hin. Diese Entwicklung dürfte sowohl für die wissenschaftliche als auch für die anwendungsorientierte Forschung gelten.

Aus den von den Unternehmen angeführten Hemmnissen ergeben sich folgende Handlungsempfehlungen:

- Um die Unternehmen mit den erforderlichen Kenntnissen für Big-Data-Anwendungen auszustatten, sollten Kooperationen mit Start-ups durch entsprechende politische Maßnahmen gefördert werden. Oft sind Start-ups gerade im digitalen Bereich mit einem Know-how und einer Risikokultur ausgestattet, welche sich komplementär zu den Kenntnissen und der Unternehmenskultur der etablierten Unternehmen verhalten. Von der Politik initiierte und von Wissenschaft und Wirtschaft umgesetzte Plattformen zur Kooperationsanbahnung sowie steuerliche Förderungen für Mittelstand-Start-up-Kooperationen wären Möglichkeiten, das Know-how für Big-Data-Anwendungen in die Breite zu tragen und insgesamt zu erhöhen.
- Um der mangelnden Qualität und Quantität von Daten in Unternehmen entgegenzuwirken, sollte das Bewusstsein für die Bedeutung der Datenbewirtschaftung in Unternehmen generell gestärkt werden. Denkbar wären entsprechende Informationsveranstaltungen und Weiterbildungsmaßnahmen, die von der Politik initiiert und durch wirtschaftsnahe Institutionen wie Industrie- und Handelskammern und Verbänden sowie Forschungseinrichtungen umgesetzt werden. Auch Förderungen und steuerliche Vorzüge für Unternehmen, die ein umfassendes Datenmanagement und eine entsprechende Data Governance darlegen können, sind empfehlenswerte Optionen. Je mehr Aufmerksamkeit die Unternehmen auf Daten als strategischen Vermögenswert legen und je eher sie eine Strategie für die Datenbewirtschaftung entwickeln, desto eher sind auch Verbesserungen in der Qualität und der Quantität der Daten zu erwarten. Im Rahmen von Projekten wie dem International Data Space (IDS) oder dem durch das Bundesministerium für Wirtschaft und Energie geförderte Projekt „Data Economics and Management of Data-Driven Business (DEMAND)“ gibt es für Unterneh-

men bereits viele Hilfestellungen für den Datenaustausch untereinander und für die Datenbewirtschaftung. Hier gilt es, eine noch größere Zahl von Unternehmen – gerade auch von kleinen und mittleren – einzubinden.

- Mit einer stärkeren Beschäftigung der Firmen mit Daten ist auch zu erwarten, dass deren Bewusstsein für das tatsächliche Kosten-Nutzen-Verhältnis für den Einsatz von Big-Data-Technologien geschärft wird. Unternehmen wägen vor allem die kurzfristigen Kosten gegen den mittelfristig kaum einschätzbaren Nutzen von Big Data ab und gelangen deshalb zu dem Ergebnis, die Kosten für den Big-Data-Einsatz seien zu hoch. An dieser Stelle muss, angeleitet durch die Führungskräfte, bei allen Mitarbeitenden eines Unternehmens ein Bewusstsein über die Vorteile und Herausforderungen von Big Data geschaffen werden.
- Um der Unklarheit beziehungsweise der Komplexität der datenschutzrechtlichen Bestimmungen entgegenzuwirken, sollte die Politik die Unternehmen darüber informieren, wie sich der Datenschutz auf Big-Data-Anwendungen auswirken könnte. Derzeit ist davon auszugehen, dass Unternehmen bestimmte Datenschutzvorgaben fürchten, selbst wenn diese in der Praxis bei einem Einsatz von Big-Data-Anwendungen im Unternehmen gar keine Anwendung finden würden, weil dort viele Daten keinen Personenbezug gemäß der Europäischen Datenschutz-Grundverordnung haben. Während eine konsequente Umsetzung dieser Verordnung alternativlos ist, muss auch klar sein, in welchen Fällen diese irrelevant ist. Informationskampagnen und -plattformen für die Unternehmen sind eine mögliche Lösung für diese Herausforderung.
- Eine bessere Vernetzung der in Bezug auf Big Data bereits sehr aktiven Wissenschaftsgemeinschaft mit der noch in geringem Umfang aktiven Wirtschaft könnte dazu beitragen, die vorhandenen Potenziale in der Anwendung und Weiterentwicklung von Big Data in Deutschland besser auszuschöpfen. Die vielfältigen Chancen, die Big Data für Unternehmen, Branchen, die Wirtschaft insgesamt und damit auch den Standort Deutschland bietet, können nur dann genutzt werden, wenn Wirtschaft und Wissenschaft Hand in Hand arbeiten. Ergebnisse aus der Forschung im Bereich Big Data müssen auf der Unternehmensebene umgesetzt werden können. Inputs aus den Unter-

nehmen wiederum sollten den Anstoß für weitere Forschung geben. Obwohl diese Vernetzung vielerorts bereits stattfindet, sind ein weiterer Ausbau und eine verstärkte Sichtbarmachung der Aktivitäten wichtig, um die Potenziale von Big Data in Deutschland heben zu können. Digitale Plattformen könnten die dafür geeigneten Informations- und Kommunikationsorte bilden.

# Literatur

- Acquisti, Alessandro / Varian, Hal, 2005, Conditioning Prices on Purchase History, in: Marketing Science, 24. Jg., Nr. 3, S. 367–381
- Acs, Zoltan / Audretsch, David / Feldman, Maryann, 1994, R&D Spillovers and Recipient Firm Size, in: Review of Economics and Statistics, 76. Jg., Nr. 2, S. 336–340
- Audretsch, David / Feldman, Maryann, 1996, R&D Spillover and the Geography of Innovation and Production, in: American Economic Review, 86. Jg., Nr. 3, S. 630–640
- Baron, Pavlo, 2013, Big Data für IT-Entscheider. Riesige Datenmengen und moderne Technologien gewinnbringend nutzen, München
- Becheikh, Nizar / Landry, Réjean / Amara, Nabil, 2006, Lessons from innovation empirical studies in the manufacturing sector. A systematic review of the literature from 1993–2003, in: Technovation, 26. Jg., Nr. 5/6, S. 644–664
- Bitkom – Bundesverband Informationswirtschaft, Telekommunikation und neue Medien, 2017, DS-GVO: Jedes dritte Unternehmen hat sich noch nicht mit der Datenschutzgrundverordnung beschäftigt, <https://www.bitkom.org/Presse/Presseinformation/Jedes-dritte-Unternehmen-hat-sich-noch-nicht-mit-der-Datenschutzgrundverordnung-beschaefigt.html> [7.3.2019]
- Bitkom, 2018, Industrie 4.0 – Wo steht Deutschland?, Hannover
- Brogger, Morten, 2018, Jetzt müssen auch die USA beim Datenschutz nachziehen, <https://www.welt.de/wirtschaft/bilanz/article181020862/US-DSGVO-Jetzt-muessen-auch-die-USA-beim-Datenschutz-nachziehen.html> [24.1.2019]
- Brynjolfsson, Erik / Hitt, Lorin / Kim, Heekyung, 2011, Strength in Numbers. How does data-driven decision-making affect firm performance?, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1819486](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486) [4.6.2019]
- Buchholtz, Sonia / Bukowski, Maciej / Śniegocki, Aleksander, 2014, Big and open data in Europe. A growth engine or a missed opportunity?, Warschau
- Bundesverband Digitale Wirtschaft, 2018, Data Economy. Datenwertschöpfung und Qualität von Daten, Berlin
- Business Application Research Center, 2015, Big Data Use Cases. Getting real on data monetization, Würzburg
- Capgemini, 2018, Studie IT Trends 2018. Digitalisierung: Aus Ideen werden Ergebnisse, <https://www.capgemini.com/de-de/wp-content/uploads/sites/5/2018/02/it-trends-studie-2018.pdf> [24.5.2019]
- Centre for Economics and Business Research, 2016, The Value of Big Data and the Internet of Things to the UK Economy, London
- Commerzbank, 2018, Der Rohstoff des 21. Jahrhunderts. Big Data, Smart Data – Lost Data?, [https://www.unternehmerperspektiven.de/portal/media/unternehmerperspektiven/up-studien/up-studien-einzelseiten/up-pdf/2018\\_04\\_17\\_FL\\_UP\\_Studie\\_online\\_2018\\_V2.pdf](https://www.unternehmerperspektiven.de/portal/media/unternehmerperspektiven/up-studien/up-studien-einzelseiten/up-pdf/2018_04_17_FL_UP_Studie_online_2018_V2.pdf) [7.12.2018]

- Crunchbase, 2019, Datenbank, <https://www.crunchbase.com/> [15.3.2019]
- Datenethikkommission, 2018, Empfehlungen der Datenethikkommission für die Strategie Künstliche Intelligenz der Bundesregierung, <https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/themen/it-digitalpolitik/datenethikkommission/empfehlungen-datenethikkommission.html> [11.4.2019]
- DEMAND – Data Economics and Management für Data-driven Business, 2019, Data Economy. Status quo der deutschen Wirtschaft & Handlungsfelder in der Data Economy. White Paper, [https://www.demand-projekt.de/paper/DEMAND-DataEconomicsAndManagementOfDataDrivenBusiness\(WhitePaper\).pdf](https://www.demand-projekt.de/paper/DEMAND-DataEconomicsAndManagementOfDataDrivenBusiness(WhitePaper).pdf) [4.6.2019]
- Desjardins, Jeff, 2017, Can Predictive Data Revolutionize Capital Markets?, <https://www.visualcapitalist.com/predictive-data-analytics-capital-markets/> [7.3.2019]
- Dorschel, Joachim (Hrsg.), 2015, Praxishandbuch Big Data. Wirtschaft – Recht – Technik, Wiesbaden
- Dutton, Tim, 2018, An Overview of National AI Strategies, <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd> [16.4.2019]
- Einav, Liran / Levin, Jonathan, 2014, The Data Revolution and Economic Analysis, in: Innovation Policy and the Economy, 14. Jg., Nr. 1, S. 1–24
- Engels, Barbara, 2016, Big-Data-Analyse. Ein Einstieg für Ökonomen, IW-Kurzbericht, Nr. 78, Köln
- Europäisches Parlament, 2016, Big data and data analytics. The potential for innovation and growth, Brüssel
- Eurostat, 2019, Digitale Wirtschaft und Gesellschaft, Datenbank, <https://ec.europa.eu/eurostat/de/web/digital-economy-and-society/data/database> [6.3.2019]
- Experian, 2016, Datenqualität und -management, <http://www.experian.de/assets/marketing-services/white-papers/data-management-trend-report-2016-DE.pdf?ref=emailmarketingtipps.de> [7.3.2019]
- Filkins, Barbara, 2017, Sensitive Data at Risk. The SANS 2017 Data Protection Survey, Swansea
- Firican, George, 2017, The 10 Vs of Big Data, <https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx> [11.3.2019]
- Forbes Insights / Teradata, 2015, Betting on Big Data. How the right culture, strategy and investments can help you leapfrog the competition, New York
- Goecke, Henry / Thiele, Christopher, 2018, KI-Forschung und Start-ups in Deutschland. Zahlreiche Zentren, IW-Kurzbericht, Nr. 75, Köln
- Heidkamp, Peter, 2016, Big Data 2016: So weit sind Unternehmen in Deutschland, <https://www.bitkom-akademie.de/news/big-data-2016-so-weit-sind-unternehmen-deutschland> [7.3.2019]
- Huber, Matthias / Schüller, Simone / Stöckli, Marc / Wohlrabe, Klaus, 2018, Maschinelles Lernen in der ökonomischen Forschung, in: ifo Schnelldienst, 71. Jg., Nr. 7, S. 74–76

- IDC – International Data Corporation, 2017, European Data Market, SMART 2013/0063, Framingham
- IIT – Institut für Innovation und Technik, 2018, Potenziale der Künstlichen Intelligenz im produzierenden Gewerbe in Deutschland, Berlin
- Jaffe, Adam, 1989, Real Effects of Academic Research, in: ABI/INFORM Global, 79. Jg., Nr. 5, S. 957–970
- Jöns, Johanna, 2016, Daten als Handelsware, Hamburg
- King, Stefanie, 2014, Big Data. Potential und Barrieren der Nutzung im Unternehmenskontext, Wiesbaden
- Konrad-Adenauer-Stiftung, 2018, Vergleich nationaler Strategien zur Förderung von künstlicher Intelligenz, Sankt Augustin
- KPMG / Bitkom, 2017, Mit Daten Werte schaffen, <https://hub.kpmg.de/mit-daten-werte-schaffen-17?hsCtaTracking=35bd247a-db5a-422c-a350-9a2fac0910f2%7C4190b7e8-a27b-43cf-b8bc-adfc718c49e3> [24.5.2019]
- Krotova, Alevtina / Rusche, Christian / Spiekermann, Markus, 2019, Die ökonomische Bewertung von Daten. Verfahren, Beispiele und Anwendungen, IW-Analysen, Nr. 129, Köln
- Laney, Doug, 2001, 3-D Data Management. Controlling Data Volume, Velocity, and Variety, Stanford
- Lichtblau, Karl et al., 2017, Ökonomische Aspekte der Digitalisierung, in: vbw – Vereinigung der Bayerischen Wirtschaft (Hrsg.), Neue Wertschöpfung durch Digitalisierung, München, S. 45–117
- Mansfield, Edwin, 1998, Academic research and industrial innovation. An update of empirical findings, in: Research Policy, 26. Jg., Nr. 7/8, S. 773–776
- Markl, Volker et al., 2013, Innovationspotenzialanalyse für die neuen Technologien für das Verwalten und Analysieren von großen Datenmengen (Big Data Management), Berlin
- McKinsey, 2017, Smartening up with Artificial Intelligence (AI). What's in it for Germany and its Industrial Sector?, Frankfurt am Main
- Möller, Doris / Karstedt-Meierrieks, Annette, 2017, Rechtliche Aspekte der Datenökonomie, Thesenpapier, Berlin
- OECD – Organisation for Economic Co-operation and Development, 2013, Exploring Data-Driven Innovation as a New Source of Growth. Mapping the policy issues raised by “big data”, OECD Digital Economy Paper, Nr. 222, Paris
- Pannetta, Kasey, 2018, 5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, <https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/> [11.3.2019]
- Reinsel, David / Gantz, John / Rydning, John, 2018, The Digitization of the World. From Edge to Core, IDC White Paper, Framingham

- Romer, Paul M., 1986, Increasing returns and long-run growth, in: Journal of Political Economy, 95. Jg., Nr. 5, S. 1002–1037
- Romer, Paul M., 1990, Endogenous Technological Change, in: Journal of Political Economy, 98. Jg., Nr. 5, S. 71–102
- Samuel, Arthur, 1959, Some Studies in Machine Learning Using the Game of Checkers, in: IBM Journal of Research and Development, 3. Jg., Nr. 3, S. 210–229
- Scopus, 2019, Datenbank, <https://www.scopus.com/home.uri> [15.3.2019]
- Statistisches Bundesamt, 2018, Bruttoinlandsprodukt 2017 für Deutschland, <https://www.destatis.de/DE/Presse/Pressekonferenzen/2018/BIP2017/pressebroeschue-re-bip.pdf?blob=publicationFile&v=2> [26.6.2019]
- Statistisches Bundesamt, 2019, Bruttoinlandsprodukt 2018 für Deutschland, <https://www.destatis.de/DE/Presse/Pressekonferenzen/2019/BIP2018/pressebroeschue-re-bip.pdf?blob=publicationFile&v=3> [26.6.2019]
- Steier, Andreas / Duisberg, Alexander, 2018, Datenschutz und rechtlicher Rahmen, in: Digitale Welt, 2. Jg., Nr. 4, S. 45–48
- TechAmerica Foundation, 2012, Demystifying Big Data. A Practical Guide To Transforming The Business of Government, Washington D. C.
- The Economist, 2016, Economists are prone to fads, and the latest is machine learning, <https://www.economist.com/finance-and-economics/2016/11/24/economists-are-prone-to-fads-and-the-latest-is-machine-learning> [15.3.2019]
- Uckelmann, Dieter / Gimber, Nicolas / Krug, Thomas / Litzel, Nico, 2018, Logistikunternehmen sollten auf Big Data setzen, <https://www.bigdata-insider.de/logistikunternehmen-sollten-auf-big-data-setzen-a-729489/> [7.3.2019]
- vbw – Vereinigung der Bayerischen Wirtschaft, 2016, Big Data im Freistaat Bayern. Chancen und Herausforderungen, München
- ZEW – Zentrum für Europäische Wirtschaftsforschung, 2015, Industrie 4.0: Digitale (R)Evolution der Wirtschaft, IKT-Report, Mannheim

## Autoren

**Barbara Engels**, B. Sc. und M. Sc. in Economics, geboren 1988 in Erlangen; Studium der Volkswirtschaftslehre in Berlin, New York und Barcelona; seit 2015 im Institut der deutschen Wirtschaft, Economist im Kompetenzfeld „Strukturwandel und Wettbewerb“.

Dr. rer. pol. **Henry Goecke**, geboren 1982 in Dortmund; Studium der Volkswirtschaftslehre in Dortmund und Glasgow sowie Promotion in Dortmund; seit 2012 im Institut der deutschen Wirtschaft, seit 2017 Leiter der Forschungsgruppe „Big Data Analytics“.

# Abstract

## Big Data in Business and Research

### A Stocktaking

The growing quantity and diversity of data generated as digitalisation progresses is closely linked to the term “Big Data”. From an economic point of view, this has led to an increase in the importance accorded not only to data itself but also, and especially, to Big Data. This IW Analysis measures the impact Big Data is having through the two channels of commerce and scientific research. The steadily increasing significance of Big Data for the sciences and related fields is apparent from the fact that since 2015 its share of all academic publications has almost doubled. Research findings often lead to innovation and thus have an indirect economic effect. The direct economic effect, which is reflected in such things as the extent to which companies use Big Data technologies, is also growing, but remains small overall. Only just over 8 percent of the 1,235 German companies in manufacturing and related services surveyed exclusively for this study perform Big Data analyses. The same survey also revealed that many companies currently face numerous challenges impeding their exploitation of Big Data. The most significant of these are a lack of in-house expertise, the low quality and quantity of their data and inadequate budgets. Where know-how is lacking, the authors recommend collaboration with start-ups. The low quality and quantity of companies’ data can best be remedied by raising consciousness of the importance of data management. This can also be expected to increase awareness of the real value for money offered by Big Data technologies and consequently lead to greater investment in Big Data.