

STUDIE
EINSATZ UND NUTZENPOTENZIALE VON DATA
MINING IN PRODUKTIONSUNTERNEHMEN

ERGEBNISSE

Autoren

Dipl.-Wirt.-Ing.
Markus Weskamp

Dipl.-Wirt.-Ing.
Andrei Tamas

Dipl.-Wirt.-Ing.
Thomas Wochinger

Dipl. oec. soc.
Anja Schatz

Studie

**Einsatz und Nutzenpotenziale von
Data Mining in Produktionsunternehmen**

Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA

März 2014

Inhaltsverzeichnis

1	Einleitung	1
2	Eckdaten der Studie	2
3	Kenntnisstand der Industrie	7
4	Softwareeinsatz	18
5	Rahmenbedingungen in den Unternehmen	22
6	Einschätzung zu Data Mining Potenzialen	25
7	Zusammenfassung und Fazit	34
8	Unterstützer der Studie	36

Abbildungsverzeichnis

Abbildung 1: Branchenzuordnung der Befragten	2
Abbildung 2: Anzahl der Mitarbeiter der befragten Unternehmen	3
Abbildung 3: Fertigungsarten der befragten Unternehmen	4
Abbildung 4: Positionen der Befragten im Unternehmen	5
Abbildung 5: Aufgabenbereich der Teilnehmer	6
Abbildung 6: Kenntnisstand zum Thema Data Mining	7
Abbildung 7: Kenntnisstand nach Unternehmensgröße (Mitarbeiterzahl)	8
Abbildung 8: Kenntnisstand nach Aufgabenbereich	9
Abbildung 9: Bekanntheit und Anwendung von Data Mining-Methoden	10
Abbildung 10: Informationsquellen zum Thema Data Mining	15
Abbildung 11: Anwendung von Data Mining im Unternehmen	16
Abbildung 12: Anwendung von Data Mining nach Fertigungsart	17
Abbildung 13: Genutzte Lizenzart für Data Mining-Software	18
Abbildung 14: Eingesetzte Softwareprodukte	19
Abbildung 15: Entscheidungskriterien bei der Auswahl von Data Mining Software	20
Abbildung 16: Gründe gegen die Anwendung von Data Mining	21
Abbildung 17: Bewertung von Aussagen zu Data Mining	22
Abbildung 18: Nutzungsart von Data Mining	24
Abbildung 19: Empfohlene Anwendungsgebiete von Data Mining	25
Abbildung 20: Bewertung der Aussagen zu Data Mining	26
Abbildung 21: Aussagen zu Bewertung von Data Mining	27
Abbildung 22: Potenziale von Data Mining in verschiedenen Anwendungsgebieten	28
Abbildung 23: Heutiges und zukünftiges Potenzial von Data Mining	29
Abbildung 24: Herausforderungen bei der Anwendung von Data Mining	30
Abbildung 25: Prozessmodell zum CRISP-DM	31
Abbildung 26: Bedarf zur Weiterentwicklung von Data Mining	32

1 Einleitung

Data Mining ermöglicht das Erkennen von Mustern und Wirkzusammenhängen in großen Datenbeständen. Es unterstützt die Analyse von linearen, nicht-linearen und zeitverzögerten Abhängigkeiten. Data Mining wird als Sammlung von Methoden für die Produktion bereits seit über 10 Jahren in der Wissenschaft diskutiert. Auch in der praktischen Anwendung sind Methoden des Data Mining seit vielen Jahren bekannt und mit zunehmender Leistungsfähigkeit der IT immer besser für den Einsatz geeignet. Die Methoden des Data Mining bedienen sich bislang hauptsächlich Unternehmen aus Branchen mit starkem direktem Endkundenbezug wie beispielsweise Banken, Versicherungen und der Handel.

Die Verwendung von Data Mining in produzierenden Unternehmen ist bisher weniger verbreitet. Lediglich im Bereich der Qualitätsdatenanalyse und -sicherung etablieren sich nach und nach einzelne Anwendungen. Das Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA stellte sich die Aufgabe, mehr über den Einsatz und die Potenziale von Data Mining in Produktionsunternehmen in Erfahrung zu bringen. Hierzu wurde eine Online-Umfrage durchgeführt, bei der sich knapp 90 Teilnehmer aus dem produzierenden Umfeld zu Fragestellungen bezüglich Data Mining äußerten. Der vorliegende Studienbericht stellt die Ergebnisse der Befragung zusammen.

Die Studie verfolgt das Ziel, die Themen Data Mining und Produktion stärker miteinander zu vernetzen. Hierzu wurde zunächst ermittelt, wie stark das Thema bereits heute in produzierenden Unternehmen verankert ist. Darüber hinaus stellt die Studie bereits etablierte Anwendungsbereiche für Data Mining dar und verdeutlicht Potenziale für neue Anwendungsfelder. Sie zeigt ebenso auf, wie die Studienteilnehmer den Nutzen und den Aufwand von Data Mining Projekten bewerten. Allgemein werden die folgenden Fragestellungen beantwortet:

- Wie stark ist die Verbreitung von Data Mining Methoden in der produzierenden Industrie?
- Welche Methoden werden genutzt und was sind die Anwendungsgebiete?
- Wie bekannt und verbreitet sind entsprechende Softwarelösungen?
- Wie werden Aufwand und Nutzen von Data Mining bewertet?
- Welche Rahmenbedingungen finden sich in den Unternehmen?
- Welche weiteren Potenziale birgt Data Mining für produzierende Unternehmen?

2 Eckdaten der Studie

Einleitend wird ein Überblick über die Struktur der Umfrageteilnehmer gegeben. Die Zielgruppe war entsprechend der Aufgabenstellung vorwiegend im produzierenden Gewerbe zu finden. Dies spiegelt sich in den Unternehmensbranchen der Befragten wider (vgl. Abbildung 1).

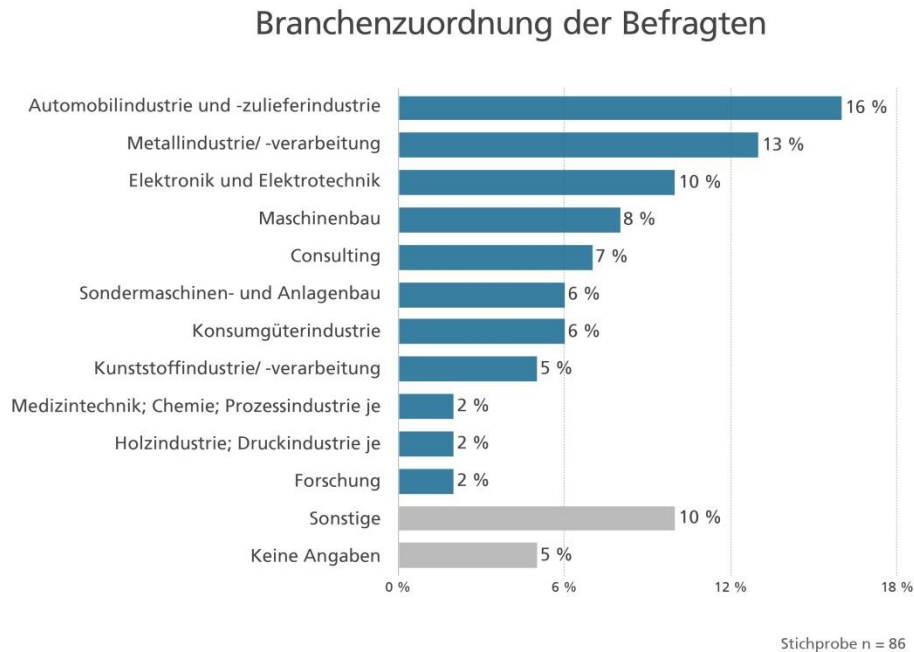


Abbildung 1: Branchenzuordnung der Befragten

Etwa ein Sechstel der Befragten kommt aus der Automobil- oder Zulieferindustrie. Diese Branche stellt damit die stärksten Teilnehmervorteiler dar. 13% der Befragten kommen aus der metallverarbeitenden Industrie. Jeder Zehnte ordnet sein Unternehmen dem Bereich Elektronik oder Elektrotechnik zu. Der Maschinenbau stellt mit 8% der Nennungen die viertgrößte Teilnehmergruppe. Daher lässt sich einerseits davon ausgehen, dass die Thematik des Data Mining in diesen Bereichen von besonderem Interesse ist. Andererseits spiegeln sich im Teilnehmerkreis ebenfalls die Geschäftsfelder des Fraunhofer IPA wieder, dessen Industriekunden bei der Befragung vorrangig adressiert wurden. Mit 7% sind Consultingunternehmen die stärksten Vertreter der nicht-produzierenden Branchen. Neben den aufgeführten Kategorien nahmen unter anderem auch Vertreter der Textilindustrie, der Messtechnik sowie der Lebensmittelindustrie teil.

Insgesamt waren an der Umfrage vorwiegend Mitarbeiter aus Großunternehmen beteiligt. Jeder dritte Teilnehmer stammt aus einem Unternehmen mit mehr als tausend Mitarbeitern.

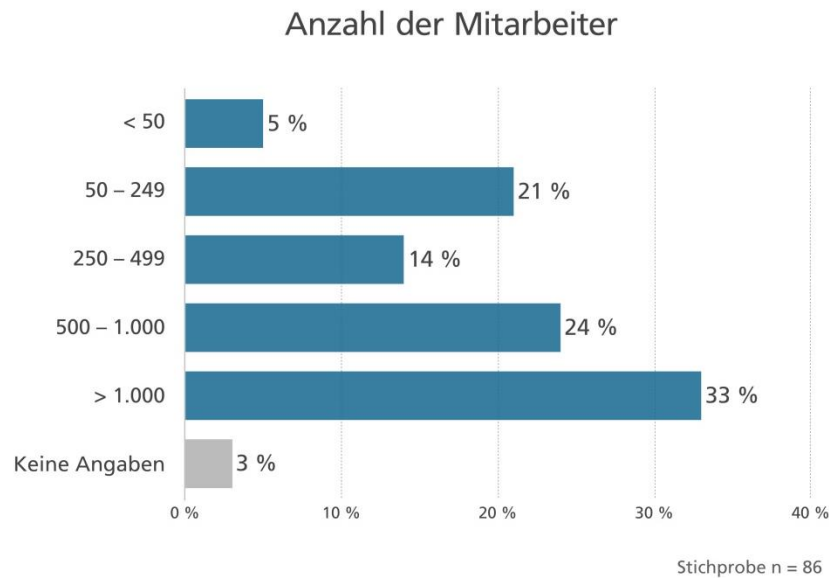


Abbildung 2: Anzahl der Mitarbeiter der befragten Unternehmen

Die Auswertung der Unternehmensgröße basiert auf der Anzahl der Beschäftigten am Standort des Befragten. Sie zeigt, dass 26% der Befragten an einem Unternehmensstandort mit weniger als 250 Mitarbeitern beschäftigt sind. Die Studie bildet daher in Summe einen Querschnitt über alle Unternehmensgrößen ab. Interessant ist die Angabe der Unternehmensgröße unter anderem, da überprüft werden soll, ob sich Großunternehmen tendenziell stärker mit der Thematik des Data Mining auseinandersetzen als kleine Unternehmen. Neben der Unternehmensgröße wurden von den Teilnehmern ebenfalls die in den Unternehmen relevanten Fertigungsarten bewertet (vgl. Abbildung 3).

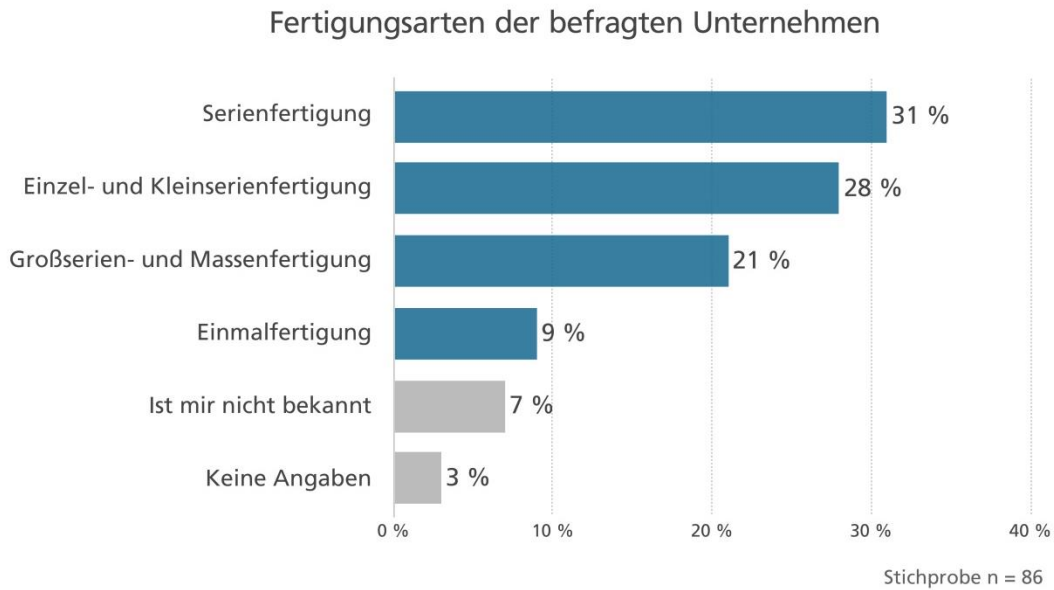


Abbildung 3: Fertigungsarten der befragten Unternehmen

Die Produktionsstruktur zeigt, dass es ein Großteil der Befragten mit Serienfertigung (ca. 20 bis 1.000 Stück pro Monat) oder Einzel- und Kleinserienfertigung zu tun hat. Knapp ein Drittel der Teilnehmer nennt die Serienfertigung als für ihr Unternehmen wesentliche Fertigungsart. Neben dieser stellt die Einzel- und Kleinserienfertigung mit 28% die zweithäufigste Fertigungsart dar. Ein Fünftel der Befragten hat es in seinem Unternehmen mit Großserien- und Massenfertigung zu tun. Jedes zehnte befragte Unternehmen betreibt Einmalfertigung. Diesbezüglich soll überprüft werden, ob die vorliegende Fertigungsstruktur Einfluss auf die Nutzungsmöglichkeiten von Data Mining aufweist, da eine komplexe und kundenindividuelle Einmalfertigung andere Anforderungen an Data Mining stellen wird als eine Massenfertigung, bei der ein höherer Standardisierungsgrad vorliegt und erwartungsgemäß eine enorme Datenmenge zu verarbeiten sein wird.

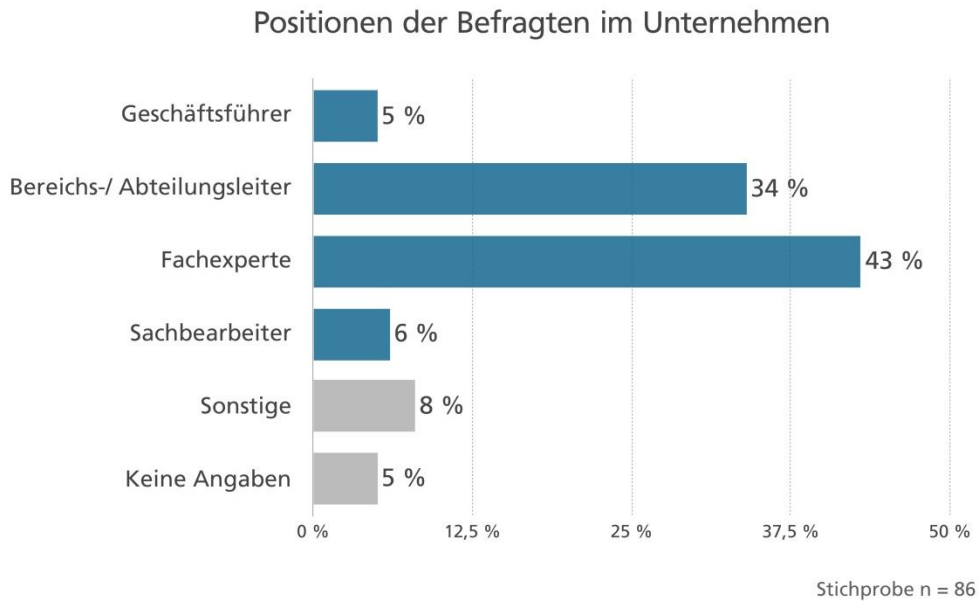


Abbildung 4: Positionen der Befragten im Unternehmen

Mit 43% stellen die Fachexperten die am stärksten vertretene Gruppe der Befragten dar. Unter der Kategorie der Fachexperten finden sich dabei beispielsweise Produktionsplaner, Prozessoptimierer und Qualitätsmanager. In der Gruppe der Fachexperten finden sich vermutlich ebenfalls vorrangig diejenigen Mitarbeiter, die tendenziell am stärksten in ihrer Arbeit mit Data Mining zu tun haben. Die zweitstärkste Gruppe stellen mit gut einem Drittel der Teilnehmer die Bereichs- oder Abteilungsleiter dar. 5% der Befragten bekleiden eine Position als Geschäftsführer und 6 % sind Sachbearbeiter. Neben den genannten Positionen gaben einige Teilnehmer an, beispielsweise als Projekt- oder Produktmanager tätig zu sein.

Zielsetzung der Studie war es, einen Überblick über die Einsatzbereiche und die Nutzenpotenziale von Data Mining speziell im produzierenden Umfeld zu erlangen. Dementsprechend kommen die meisten Befragten aus den Bereichen der Fertigung und Produktion. Gut jeder Fünfte stammt direkt aus einem Fertigungsbereich. Es wird zu überprüfen sein, ob Teilnehmer aus den direkten Fertigungsbereichen einen anderen Bezug zum Thema Data Mining haben als Mitarbeiter aus produktionsunterstützenden Bereichen wie beispielsweise dem Controlling oder dem Marketing.



Abbildung 5: Aufgabenbereich der Teilnehmer

Mit 20% stellen die Vertreter aus Forschung und Entwicklung die zweitstärkste Teilnehmergruppe dar. 7% der Teilnehmer kommen aus dem Bereich der Geschäftsführung und jeweils 6% stammen aus einem der Bereiche IT, Qualität oder Logistik. Je 5% kommen aus dem Bereich Controlling oder Vertrieb. Die Bereiche Disposition, Einkauf, Marketing und Produktmanagement sind mit jeweils 2% vertreten. Neben den genannten Gruppen beteiligten sich Mitarbeiter beispielsweise aus den Bereichen Prozessoptimierung, technischer Service oder dem Technologiemanagement.

3 Kennntnisstand der Industrie

Wie stark das Thema Data Mining im Bereich des produzierenden Gewerbes bisher vertreten ist, wurde in einem ersten Schritt über den Bekanntheitsgrad von Data Mining abgefragt. Abbildung 6 stellt die Ergebnisse dieser Abfrage dar.

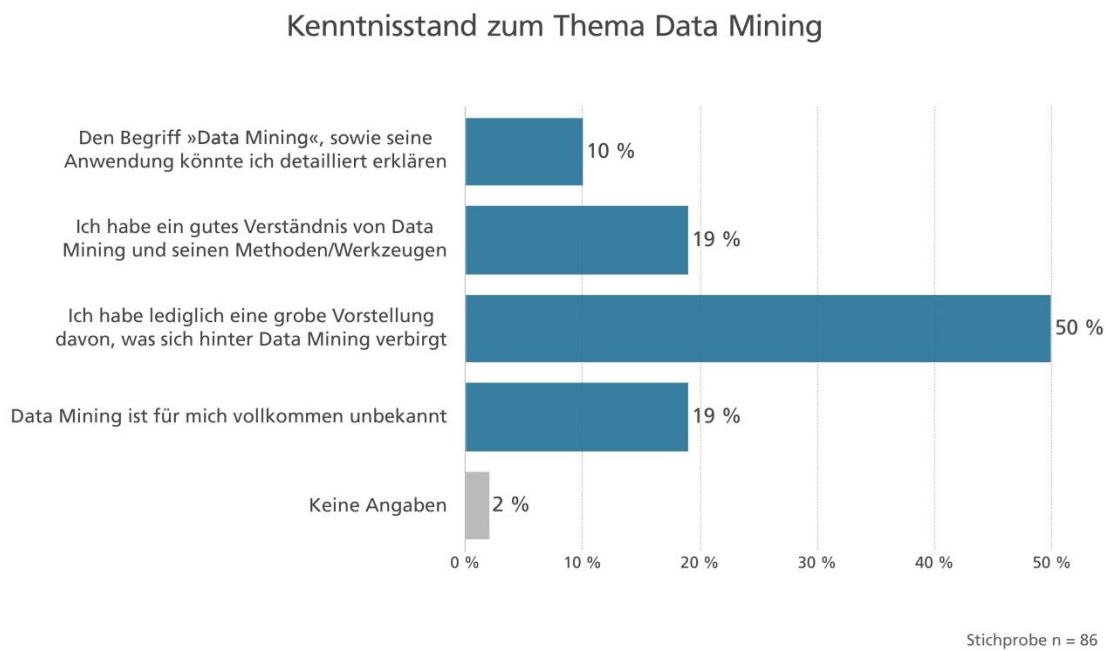


Abbildung 6: Kennntnisstand zum Thema Data Mining

Unter den Teilnehmern der Umfrage zeigt sich ein stark unterschiedliches Kenntnissniveau bezüglich der übergeordneten Thematik des Data Mining. Lediglich jeder zehnte Befragte traut sich aktuell zu, den Begriff sowie die entsprechenden Methoden und Anwendungen detailliert erklären zu können. Weitere knapp 20% der Umfrageteilnehmer verfügen immerhin über ein gutes Verständnis über die Thematik sowie über die Werkzeuge des Data Mining. Die Hälfte der Befragten hat zumindest eine grobe Vorstellung davon, was sich hinter Data Mining verbirgt. Allerdings ist für fast jeden Fünften der Begriff Data Mining vollkommen unbekannt.

Ausgehend von der Hypothese, dass sich vor allem größere Unternehmen stärker mit dem Thema Data Mining beschäftigen, überrascht die Auswertung des Bekanntheitsgrades nach Unternehmensgröße (vgl. Abbildung 7). Die Detailanalyse zeigt, dass sich in Großunternehmen mit mehr als tausend Mitarbeitern (insgesamt 28 Befragte) unter den Teilnehmern kein Spezialist mit detailliertem Data Mining Fachwissen befand. Hingegen gaben 18% (4 Teilnehmer) der Unternehmen mit weniger als 250 Mitarbeitern an, über spezielles Fachwissen zu verfügen. Der größte Anteil (64%) der Vertreter von Großunternehmen traut sich lediglich ein grobes Verständnis bezüglich Data Mining zu. Vertreter des Mittelstandes haben mit einem Anteil von 42% überwiegend ein gutes Verständnis von der Thematik. Insgesamt lässt sich die Vermutung der Bekanntheitsgrad sei in Großunternehmen tendenziell besser als in kleineren Unternehmen somit nicht bestätigen. Anzumerken sei hierbei noch einmal die unterschiedliche Bewertungsbasis, da wie in Abbildung 2 zu sehen war, die verschiedenen Unternehmensgrößen unterschiedlich stark vertreten waren.

Kennntnisstand nach Unternehmensgröße

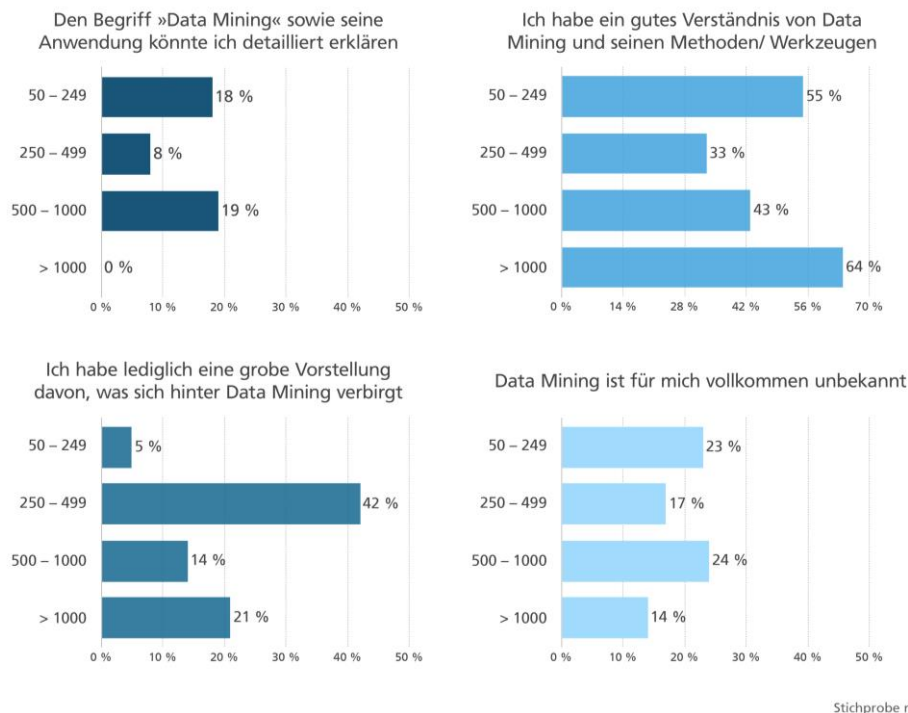


Abbildung 7: Kennntnisstand nach Unternehmensgröße (Mitarbeiterzahl)

Die befragten Sachbearbeiter (5 Teilnehmer) kannten Data Mining gar nicht oder hatten bestenfalls eine grobe Vorstellung davon. Von insgesamt vier Umfrageteilnehmern, die angaben als Geschäftsführer tätig zu sein, trauten sich drei ein gutes oder sogar sehr detailliertes Fachwissen zu, während für einen Befragten das Thema völlig fremd war. Am größten war, wie bereits dargestellt, der Anteil der Fachexperten (37 Befragte) sowie der Abteilungsleiter (29 Befragte). Vor allem bei den Abteilungsleitern ist der Anteil derjenigen, denen der Begriff Data Mining unbekannt ist, mit lediglich 10% unterdurchschnittlich repräsentiert. Andererseits gaben mit knapp 40% besonders viele Fachexperten mit spezifischem Aufgabenbereich (z.B. Produktionsplaner, Qualitätsmanager) an, sich sehr gut mit der Thematik Data Mining auszukennen.

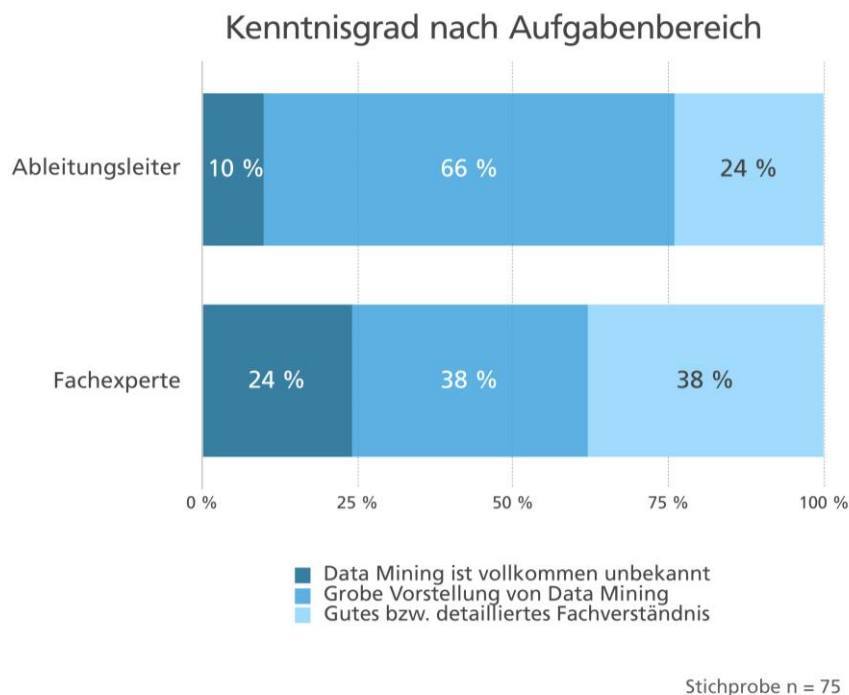


Abbildung 8: Kenntnisstand nach Aufgabenbereich

Befragt nach den Methoden des Data Mining zeigt sich, dass vor allem Entscheidungsbäume, Regressionsanalysen und Clusteranalysen einem Großteil der Umfrageteilnehmer bekannt sind bzw. eingesetzt werden. Zwischen 10 und 15% der Teilnehmer gab an, eine dieser Methoden beruflich anzuwenden. Neuronale Netze sowie Text Mining sind zwar einigen Teilnehmern bekannt, aber nur 3% der Befragten verwenden eine der Methoden beruflich. Nur gut jedem Dritten sind Assoziati-

onsregeln als Methode geläufig. Entscheidungsbäume und Regressionsanalysen werden dabei vor allem von Mitarbeitern aus der Fertigung und Produktion, dem Qualitätsmanagement sowie in der Forschung und Entwicklung verwendet. Clusteranalysen finden sich neben den Anwendungsgebieten in der Produktion auch im Bereich Logistik und Supply Chain Management.

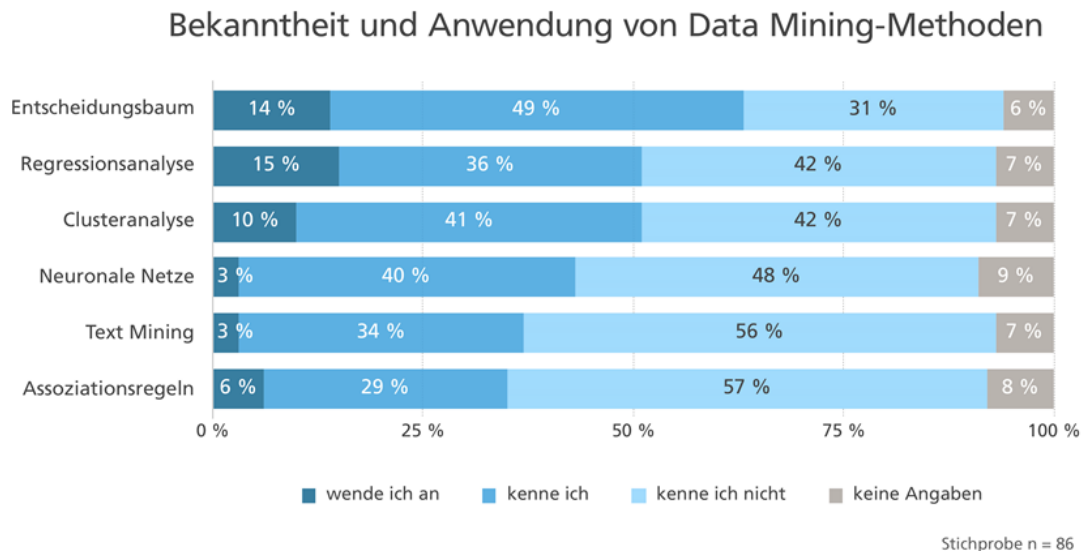


Abbildung 9: Bekanntheit und Anwendung von Data Mining-Methoden



Entscheidungsbäume gehören zum Bereich der klassifizierenden Verfahren und werden, wie bereits aus dem Namen ersichtlich, durch eine baumartige Struktur beschrieben. Sie dienen zur Unterscheidung einer betrachteten Menge von Datensätzen in zwei oder mehr Gruppen. Dabei versucht die Klassifikation jeden Datensatz (z.B. ein Bauteil, ein Prüfergebnis oder einen Transportauftrag) zu bewerten: Entweder in gute oder schlechte (z.B. Bauteile), einen kleinen, mittleren oder großen Aufwand (z.B. Prüfungen) oder jede beliebige vom Anwender gewünschte Unterteilung. Das Verfahren benötigt hierfür historische Daten und zusätzlich die Bewertung durch den Anwender, ob die betrachteten historischen Datensätze jeweils „gut“, „mittel“ oder „schlecht“ klassifiziert werden. Anschließend sucht die Methode Ähnlichkeiten und Unterschiede zwischen den Datensätzen,

um für neue Datensätze die richtige Klasse voraussagen zu können. Die Genauigkeit der Vorhersagen hängt von der Vielzahl und Qualität der verfügbaren Messwerte ab. In klassischen Produktionsprozessen lassen sich Genauigkeiten teilweise so exakt voraussagen, dass sogar auf Gutteileprüfungen verzichtet wird.

Aus der Praxis

MI226U

In der Praxis werden Klassifikationsmethoden verwendet, um beispielsweise die Qualität von Bauteilen zu prognostizieren. Wenn ein Bauteil als fehlerbehaftet klassifiziert wird, kann auf eine Prüfung verzichtet werden und das Teil verworfen bzw. nachgearbeitet werden, wenn die Wahrscheinlichkeit für ein defektes Bauteil bei gleichzeitig hohen Kosten für die Prüfung ausreichend hoch ist.

Die Klassifikation wird bereits eingesetzt, um nicht nur die Qualität der produzierten Teile, sondern auch den Zustand von Fertigungsanlagen und deren Verschleißteilen zu bewerten. So ist es möglich, die Anlage bis zu einem gerade noch vertretbaren Zustand auszureizen und Ersatzteile so spät wie möglich auszutauschen. Neben der verbesserten Auslastung der Maschinen kann in der Praxisanwendung eine höhere Produktionsmenge je Verschleißteil erzeugt werden, wodurch die Herstellkosten je Teil sinken.

Durch die ständige Überwachung der generierten Kennzahlen kann dadurch die Wartung zusätzlich von einer periodischen auf eine zustandsbasierte Wartung umgestellt werden. Hierdurch ist es ebenfalls möglich, die Wartungskosten zu senken.



Regressionsanalysen bieten ähnliche Funktionen wie Entscheidungsbäume, führen dabei aber keine eindeutige Einordnung in eine Gruppe durch, sondern sagen Wahrscheinlichkeiten für das Eintreten eines Ergebnisses oder die Zugehörigkeit zu einer Gruppe voraus. Durch die Berechnung quantitativer Werte kann die Stärke eines Zusammenhangs, im Gegensatz zu Entscheidungsbäumen, aufgezeigt werden. So wird eine falsche Einordnung unsicherer Datensätze vermieden. Eingee-

setzt wird die Methode entweder für Prognosen über mögliche Eintritte von Ereignissen oder für die Überprüfung der Stärke eines Zusammenhangs von zwei oder mehr Merkmalen.

Während bei der Überprüfung von Zusammenhängen mehrerer Merkmale ähnliche Aspekte wie auch bei Entscheidungsbäumen untersucht werden können, bietet die Prognosefunktion Aussagen über mögliche zukünftige Ereignisse und Entwicklungen.

Aus der Praxis

▲▲▲

Regressionsanalysen bieten insbesondere die Möglichkeit, Klassifikationsmethoden um quantitative Aussagen zu erweitern. Insbesondere bei Klassifikationsmodellen, die nicht den erwarteten Genauigkeitsgrad liefern, können stattdessen Regressionsanalysen eine bessere Vorauswahl erzielen. Obwohl im Vergleich zur Gutteile-/Schlechtteileprüfung lediglich Wahrscheinlichkeitswerte mit der Zuordnung zu einer bestimmten Gruppe angegeben werden, reicht dies häufig aus, um kostenintensive Überprüfungen bei bestimmten Produkten einzusparen. Es ist zudem möglich, durch eine Echtzeitverfolgung der Maschinenparameter vom Sollzustand abweichende Kennwerte zu erfassen und somit prognostizierte Schlechtteile direkt aus dem Prozess zu schleusen.



Clusteranalysen gehören zum Bereich der segmentierenden Verfahren und gruppieren Datensätze in eine definierte Anzahl von Klassen, in denen die eingeteilten Datensätze möglichst ähnlich zueinander und möglichst unterschiedlich im Vergleich zu anderen Klassen sind. Auch Clusteranalysen sind klassifizierende Verfahren, für die im Gegensatz zu Entscheidungsbäumen jedoch keine Klassen im Voraus festgelegt sind und vorab keine Unterteilung mithilfe historischer Trainingsdaten durchgeführt wurde. Die Kriterien zur Bildung der Zusammenhänge sind durch die Methode festgelegt und eine vorausgehende Einordnung einzelner Datensätze in eine bestimmte Klasse ist nicht möglich. Nach erfolgter Gruppierung obliegt es dem Anwender, die erhaltenen Klassen zu bewerten und entsprechende Schlussfolgerungen zu ziehen.

Aus der Praxis

ANZ DER PRAXIS

Clusteranalysen eignen sich gut für eine erste Analyse von Datensätzen bei überschaubarem Aufwand. Durch die Klassifikation von Datensätzen ist es z.B. möglich, schnell eine Gruppe verspäteter Aufträge zu finden und basierend auf diesen Daten die Merkmale zu identifizieren, die für die Verspätungen von Bedeutung und möglicherweise ursächlich sind. Um anschließend die Anzahl verspäteter Aufträge zu senken, können Kennzahlen festgelegt werden, ab denen ein Eingreifen durch die Planung notwendig wird, um drohenden Verspätungen frühzeitig entgegenzuwirken.

In kundennahen Anwendungsbereichen wie dem Handel werden Clusteranalysen beispielsweise dazu verwendet, Kunden in Gruppen aufzuteilen und anschließend gezielte Produkte und Vertriebskanäle für diese Kundengruppen zu entwickeln. Die Anwendung von Clusteranalysen ist mit überschaubarem Aufwand auch in der Produktion möglich. So eignen sich Clusteranalysen beispielsweise zur Rüstoptimierung, indem Klassen ähnlicher Produkte gebildet werden, die während eines Produktionsslots gefertigt werden und untereinander keine oder nur geringe Rüstzeiten aufweisen.

Wissen

Neuronale Netze beruhen auf der Bestrebung die Nervenstrukturen von Lebewesen in vereinfachter Form nachzuahmen, um auf dieser Basis Entscheidungen abzuleiten. Bei Neuronalen Netzen wird die Signalverarbeitung durch eine große Anzahl gleichartiger Elemente (Neuronen) die miteinander in Verbindung stehen, durchgeführt. Informationen werden in Form von Signalen über Verbindungskanäle (Synapsen) weitergeleitet. Die Signale werden dabei verstärkt oder vermindert, wodurch eine Gewichtung erzeugt wird. Die Einheiten eines Neuronales Netzes sind dabei schichtweise aufgebaut, wobei das Netz aus einer Input-Schicht, diversen inneren oder verborgenen Schichten sowie einer Output-Schicht besteht. Somit wird bei einem Neuronales Netz durch

eine Menge an Input-Signalen ein Output-Signal erzeugt. Neuronale Netze können beispielsweise zur Entscheidungsunterstützung im Rahmen der Produktionsplanung genutzt werden.



Text Mining bezeichnet die Suche nach Mustern in Texten mit dem Ziel, relevante Informationen daraus zu gewinnen. Ein Anwendungsgebiet ist das Web Mining, in dem Informationen aus dem Internet gewonnen werden, beispielsweise zur automatisierten Gewinnung von (bereits bestehenden) Informationen oder der Analyse von Nutzerverhalten mit dem Ziel, darauf aufbauende Analysen zu starten. Die auf die Daten angewandten Analysen sind vielfältig. Hauptsächlich werden Cluster gebildet oder Zusammenhänge beschrieben.

Im Gegensatz zu anderen vorgestellten Verfahren des Data Mining unterscheidet sich das Text Mining insbesondere dadurch, dass es auf weitgehend unstrukturierten Daten angewandt wird. Um trotzdem relevante Informationen aus den Daten zu erhalten, werden die Daten zu Beginn stärker strukturiert, um anschließend klassische Methoden des Data Mining anwenden zu können. In produzierenden Unternehmen ergeben sich überall dort Potenziale, wo längere Textblöcke verwendet werden, wie z.B. im Bereich Angebotswesen im kundenindividuellen Maschinen- und Anlagenbau.



Assoziationsregeln sind insbesondere im Rahmen der Warenkorbanalyse bekannt. Dabei wird geprüft, welche Produkte vom Kunden vermehrt zusammen gekauft werden und aus diesen Erkenntnissen versucht, Produkte im Supermarkt nah beisammen zu platzieren oder, im Falle von Online-Käufen, den (potenziellen) Kunden weitere Produkte zu empfehlen, die auch interessant für ihn sind. Je besser diese Prognosen sind, desto eher werden zusätzliche Verkäufe angeregt.

Aus der Praxis
VIRUS

Warenkorbanalysen können analog zum typischen Anwendungsfall bei Käufen auch im Produktionsumfeld angewandt werden. Eine Möglichkeit bietet die Optimierung des Produktionsprogramms. Durch die Erkenntnis, welche Produkte häufig gemeinsam oder zeitnah herzustellen sind, können so Produktionsreihenfolgen angepasst und in der Folge Rüstzeiten und Lagerbestände gesenkt werden. Je nach Situation in der Fertigung kann auch die Anzahl von Umplanungen und der Bedarf an Überstunden kontinuierlich gesenkt werden.

Ergänzend zur Frage nach der Bekanntheit einzelner Methoden wurde nach der Herkunftsquelle für die Informationen zum Thema Data Mining gefragt (vgl. Abbildung 10).

Informationsquellen zum Thema Data Mining

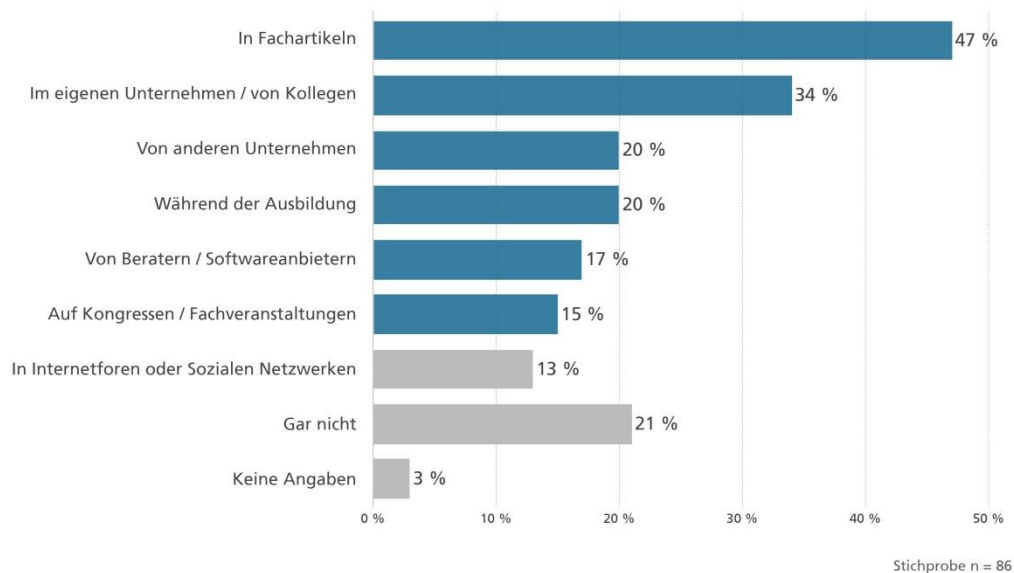


Abbildung 10: Informationsquellen zum Thema Data Mining

Fachartikel stellen für die meisten Befragten die vorrangige Informationsquelle dar. Fast jeder Zweite gab an, auf diesem Wege Informationen über Data Mining zu bekommen. Ein Drittel der Teilnehmer bezieht das Wissen außerdem aus dem eigenen Unternehmen sowie von Kollegen. Gespräche mit anderen Unternehmen oder Beratern stellen, ebenso wie das Erlernte während der Ausbildung, bei rund 20% eine Informationsbasis dar. Des Weiteren zählen Kongresse, Fachveranstaltungen, Internetforen oder Soziale Netzwerke für 13 bis 15% zu den wesentlichen Informationsquellen. Fast jeder Fünfte Befragte gab hingegen an, sich gar nicht über Data Mining zu informieren. Dies spiegelt sich in dem gemäß Abbildung 6 erläuterten Wissensstand wieder.

Weiterhin wurde in Erfahrung gebracht, welcher Anteil der Befragten Data Mining bereits im Unternehmen einsetzt, wer den Einsatz in naher Zukunft plant und wie groß der Anteil derjenigen ist, die auch in Zukunft auf den Einsatz von Data Mining verzichten werden.

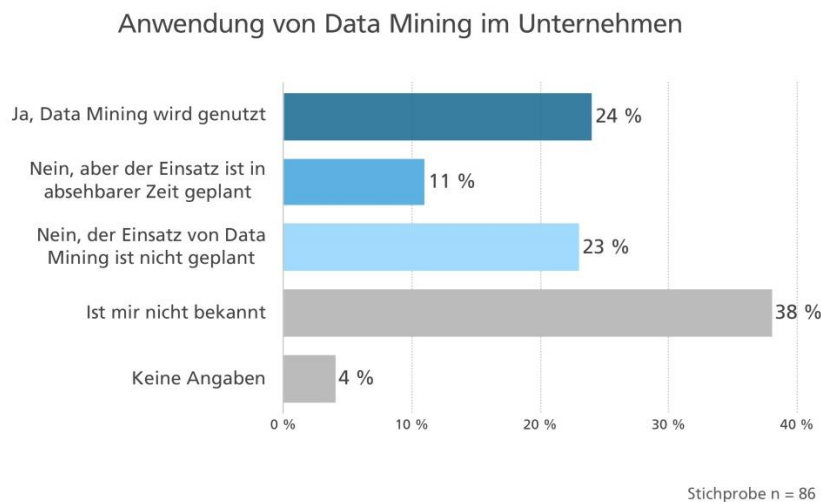


Abbildung 11: Anwendung von Data Mining im Unternehmen

Lediglich jeder vierte Teilnehmer der Online-Umfrage gab an, dass in seinem Unternehmen Data Mining genutzt werde. 10% der Befragten planen den Einsatz von Data Mining in naher Zukunft. Bei 23% der Befragten ist auf absehbare Zeit kein Einsatz von Data Mining geplant. Ein großer Anteil von knapp 40% konnte keine Aussagen bezüglich des aktuellen oder geplanten Data Mining

Einsatzes machen, was wiederum zeigt, dass die Beschäftigung mit diesem Thema noch nicht durchgängig verbreitet ist.

In einer Detailanalyse sollte die Hypothese überprüft werden, ob sich Vertreter von Unternehmen unterschiedlicher Fertigungsarten oder Branchen, hinsichtlich des Nutzungsverhaltens von Data Mining unterscheiden. Hierzu wurde die Frage nach der Anwendung von Data Mining den jeweiligen Branchen und Fertigungsarten gegenübergestellt. In Bezug auf die Fertigungsarten lässt sich jedoch kein eindeutiger Trend zur Anwendung von Data Mining erkennen (vgl. Abbildung 12). So ist Data Mining auf Grundlage der befragten Teilnehmer am stärksten im Bereich der Einmalfertigung vertreten. Hier gaben 67% derer, die sich zum Einsatz von Data Mining äußerten an, es bereits zu nutzen oder den Einsatz in absehbarer Zeit zu planen. Zweitstärkste Gruppe bilden die Vertreter der Massenfertigung und drittstärkste Gruppe die Serienfertiger. Die geringe Stichprobengröße von lediglich 47 Befragten ist dabei vor allem auf die geringe Aussagefähigkeit bezüglich des Einsatzes von Data Mining sowie der Unsicherheit bei der Klassifizierung der eigenen Fertigungsart zurückzuführen. Die Hypothese, die Fertigungsstruktur habe prinzipiell Einfluss auf die Nutzungswahrscheinlichkeit von Data Mining, lässt sich somit nicht bestätigen.

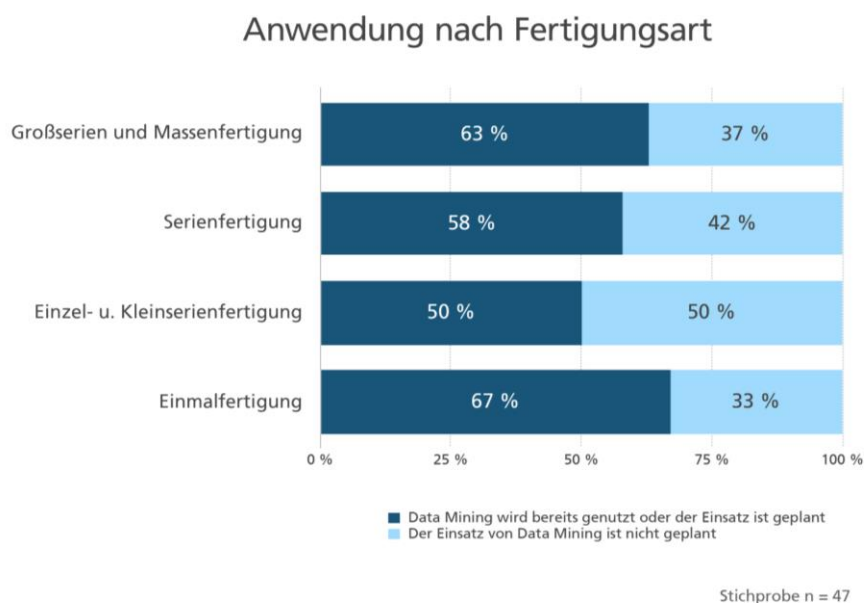


Abbildung 12: Anwendung von Data Mining nach Fertigungsart

4 Softwareeinsatz

Die Anwendung von Data Mining ist aufgrund der großen benötigten Datenmengen sowie der teilweise komplexen Berechnungen in der Regel nur unter Zuhilfenahme von entsprechenden Softwarelösungen wirtschaftlich möglich, sodass diejenigen, die bereits Data Mining nutzen, ebenfalls Aussagen zur verwendeten Software machen konnten. Hierbei sollte geklärt werden, welche Arten von Lizenzen genutzt werden (vgl. Abbildung 13).

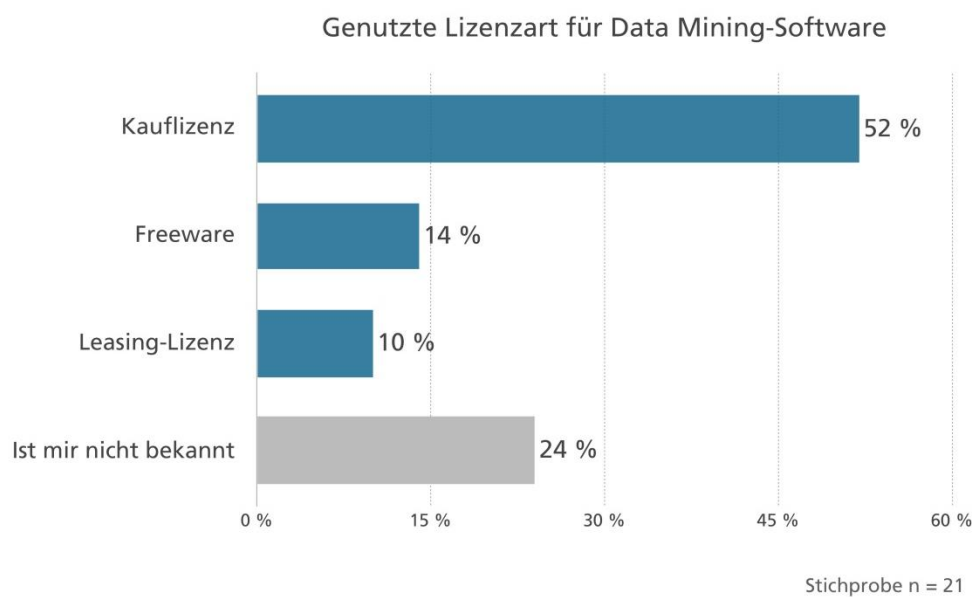


Abbildung 13: Genutzte Lizenzart für Data Mining-Software

Mehr als die Hälfte derer, die bereits Data Mining im Unternehmen einsetzen, verwendet eine Software in Form einer Kauflizenz. Kauflizenzen stellen unter den gängigen Softwareprodukten somit mit Abstand den größten Anteil dar. 14% der Befragten greifen auf Freeware zurück und 10% setzen auf den Einsatz von Software mit Leasing-Lizenz. Knapp einem Viertel ist nicht bekannt, welcher Art von Lizenzierung die verwendete Software unterliegt. Neben der Lizenzierungsart wurde gefragt, welches Softwareprodukt eingesetzt wird. Abbildung 14 zeigt die häufigsten Nennungen.

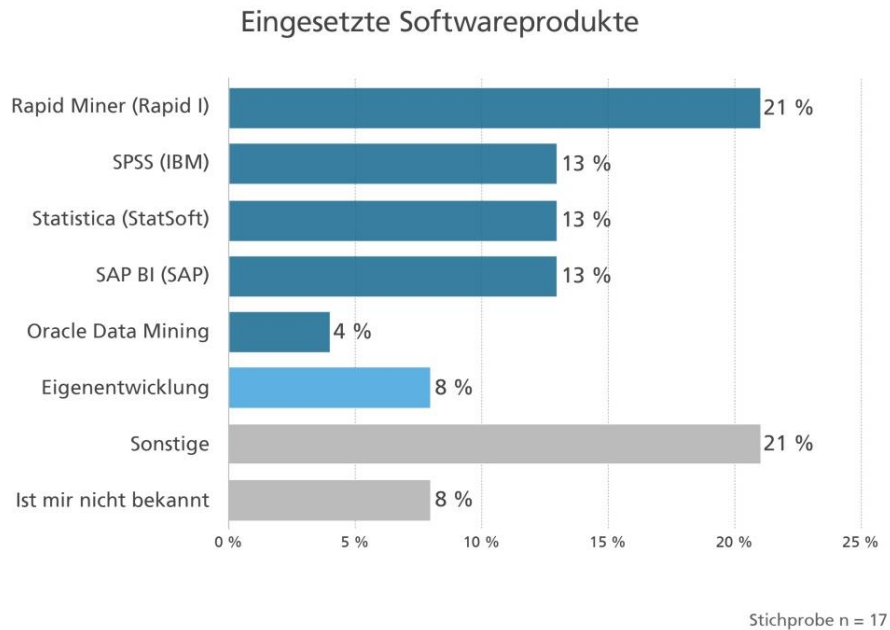


Abbildung 14: Eingesetzte Softwareprodukte

Mit gut 20% stellt die Software Rapid Miner den größten Anteil an eingesetzten Softwarelösungen unter den Befragten, gefolgt von SPSS, Statistica und SAP BI, die mit jeweils 13% am zweithäufigsten eingesetzt wurden. 4% der Teilnehmer verwenden Oracle Data Mining und die restlichen 8% setzen bei der Anwendung auf selbst entwickelte Lösungen.

Bei der Frage nach Kriterien für die Softwareauswahl zeigt sich ein eindeutiges Bild hinsichtlich der wichtigsten Entscheidungsfaktoren. Die Softwarefunktionalität, Anwenderfreundlichkeit und Kostenkriterien haben bei der Investitionsentscheidung den größten Einfluss (vgl. Abbildung 15).

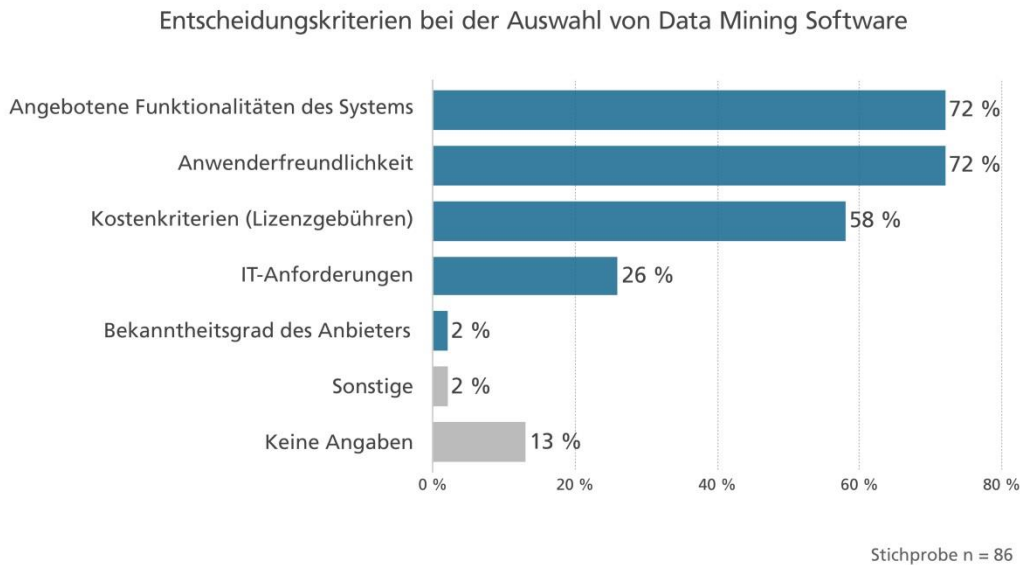


Abbildung 15: Entscheidungskriterien bei der Auswahl von Data Mining Software

Nahezu drei Viertel der Teilnehmer sind überzeugt, dass die angebotene Funktionalität bei der Entscheidung für eine Data Mining Software wichtigstes Kriterium sei. Ebenso viele messen dem Kriterium der Anwenderfreundlichkeit des Systems höchste Priorität bei. Damit überwiegen die beiden Kriterien den Kostenkriterien beispielsweise in Form von Lizenzgebühren. Immerhin beinahe 60% der Befragten nennen Kostenkriterien als wesentlich für die Investitionsentscheidung. Lediglich rund ein Viertel der Befragten sieht die Anforderungen an die IT-Systeme als entscheidenden Punkt bei der Softwareauswahl. Die Verfügbarkeit immer höherer Rechenleistungen lässt die IT-Anforderungen scheinbar als Entscheidungskriterium an Bedeutung verlieren. Erfahrungen aus der Praxis zeigen hingegen, dass vor allem bei großen Datenmengen und anspruchsvollen Berechnungen die IT-Anforderungen nicht zu vernachlässigen sind.

Der Bekanntheitsgrad des Softwareanbieters hat mit lediglich 2% so gut wie keinen Einfluss auf die Kaufentscheidung. Dies lässt erwarten, dass im Bereich Data Mining auch Softwarelösungen kleinerer Systemhäuser gute Chancen am Markt haben, sofern diese durch die angebotene Funktionalität und Anwenderfreundlichkeit überzeugen.

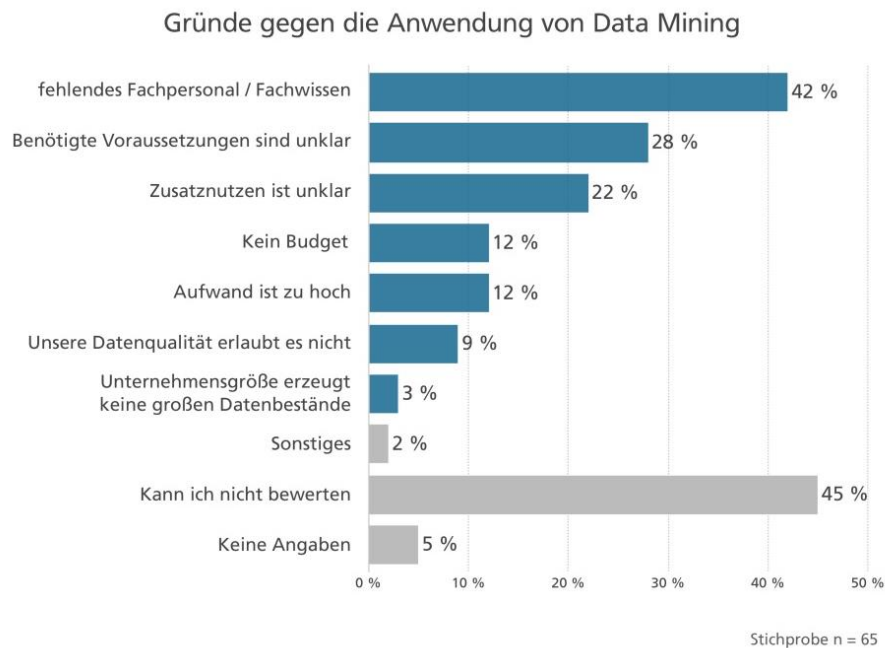


Abbildung 16: Gründe gegen die Anwendung von Data Mining

Gut 40% der Teilnehmer nannten den Mangel an Fachpersonal und Fachwissen als Grund, weshalb sie kein Data Mining verwenden. Knapp 30% begründeten den Verzicht auf die Nutzung von Data Mining dadurch, dass die benötigten Voraussetzungen unklar seien und nahezu einem Viertel der Befragten war der Zusatznutzen durch Data Mining unklar. Fehlendes Budget wurde von 12% der Befragten als einer der Gründe gegen Data Mining genannt. Für ebenso viele Teilnehmer ist der Aufwand zur Einarbeitung und Anwendung zu hoch. Nur 9% sehen in mangelnder Datenqualität einen Grund, der gegen die Nutzung von Data Mining in ihrem Unternehmen spricht. Praxiserfahrungen zeigen hingegen, dass im Rahmen von Data Mining Projekten die mangelnde Datenqualität eine der wesentlichen Hürden darstellt. Andererseits zeigt sich auch immer wieder, dass Unternehmen oftmals zunächst für das Thema Datenqualität sensibilisiert werden müssen und selbst gravierende Schwachstellen oft erst im Rahmen von entsprechenden Analyseprojekten aufgedeckt werden. Die Methoden des Data Mining erlauben es zudem, schnell entsprechende Qualitätsprobleme in den vorliegenden Daten zu identifizieren (bspw. über Ausreißerererkennung). Lediglich 3% der Teilnehmer gehen davon aus, ihr Unternehmen generiere für die Anwendung von Data Mining keine ausreichenden Datenbestände. Alle Teilnehmer, die diese Aussage tätigten, kamen aus Unternehmen mit weniger als 250 Mitarbeitern.

5 Rahmenbedingungen in den Unternehmen

Im nächsten Teil der Befragung sollte geklärt werden, welche für die Nutzung von Data Mining relevanten Rahmenbedingungen in den Unternehmen vorzufinden sind. Hierzu wurden Fragen zur Verfügbarkeit und Qualität von Daten, zur Einbindung von Data Mining in die Arbeitsabläufe sowie zur Dynamik der Datenbestände gestellt (vgl. Abbildung 17).

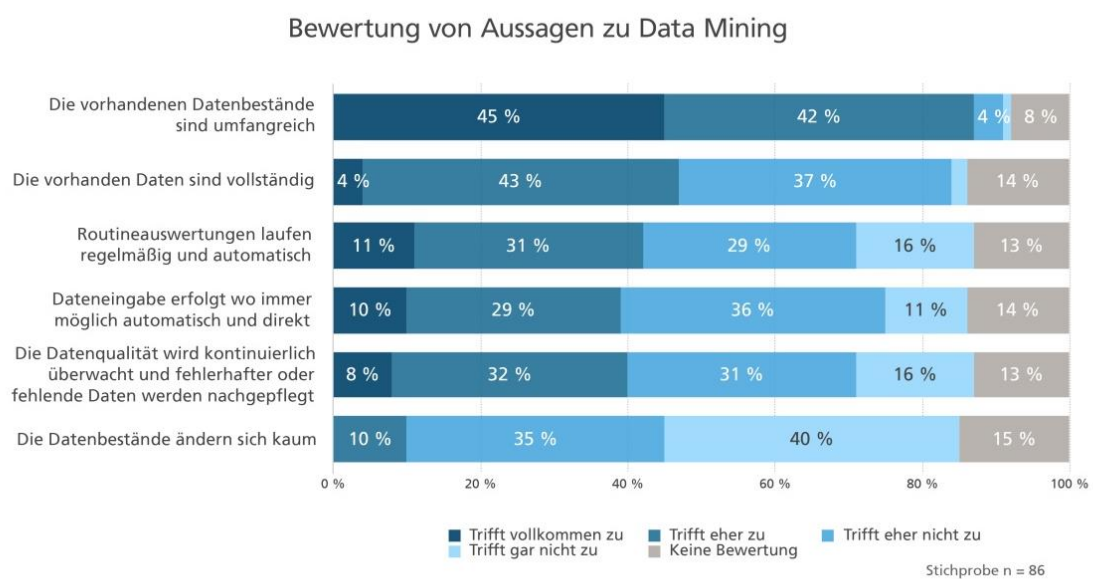


Abbildung 17: Bewertung von Aussagen zu Data Mining

Wie bereits die Auswertung zu den Hinderungsgründen bestätigt, stellt ein ausreichender Datenbestand für die wenigsten Teilnehmer ein Argument gegen die Anwendung von Data Mining dar. Insgesamt stimmen 87% der Befragten zu, ihr Unternehmen verfüge über umfangreiche Datenbestände. Lediglich 5% stimmen dieser Aussage nicht zu. Bezüglich der Datenqualität stimmen mit 47% weniger als die Hälfte der Befragten zumindest teilweise zu, dass die Ihnen zur Verfügung stehenden Daten vollständig seien. Die Befragung unterstreicht demnach die Aussage, dass die in der Industrie verfügbaren Daten oftmals eine schlechte Qualität aufweisen. Zwar stellt dies keinen prinzipiellen Hinderungsgrund für die Anwendung des Data Mining dar, sie macht die Datenaufbereitung jedoch zu einem wesentlichen Bestandteil eines jeden Data Mining Projektes, da die zielfüh-

rende Anwendung von Data Mining in der Regel recht hohe Anforderungen an die Datenqualität voraussetzt. Diese Datenaufbereitung wird in der Praxis nicht selten mit einem hohen manuellen Aufwand in Verbindung gebracht und von vielen potenziellen Anwendern gescheut. Wie bereits erläutert werden im Rahmen von Data Mining auch Werkzeuge angeboten, die aufwandsarm eine Prüfung und Verbesserung der Datenqualität ermöglichen.

Die unbefriedigende Datenqualität ist dabei nicht selten auf die Art der Datengenerierung zurückzuführen. Nur 39% der Befragten erklären, dass Dateneingaben zu großen Teilen automatisiert und direkt erfolgen. 46% widersprechen dieser Aussage zumindest tendenziell. Ein Großteil der heute in der Industrie verwendeten Daten beruht auf manuellen Dateneingaben. Hierzu zählen im Bereich des Produktionsmanagements beispielsweise die Generierung und Rückmeldung von Fertigungsaufträgen sowie das Anlegen von Produkt- und Produktionsstammdaten.

Eine noch größere Herausforderung als die einmalige Erzeugung der benötigten Datenqualität liegt in deren langfristigen Sicherung. Eine kontinuierliche Überwachung sowie die permanente Pflege fehlerhafter Daten erfolgt bei lediglich 39% der Befragten. Fast die Hälfte der Umfrageteilnehmer gab hingegen an, dass eine derartige Überwachung und Datenpflege nicht regelmäßig stattfindet.

Bezüglich der Änderungsdynamik der Daten erklärten ca. drei Viertel der Befragten, dass sich ihre Datenbestände kontinuierlich verändern. Vor allem bei den dynamischen Daten stellt die langfristige Sicherung der Datenqualität eine besondere Herausforderung dar. Die Änderungshäufigkeit der Daten kann ebenfalls einen Einfluss auf die Nutzungsart von Data Mining ausüben.

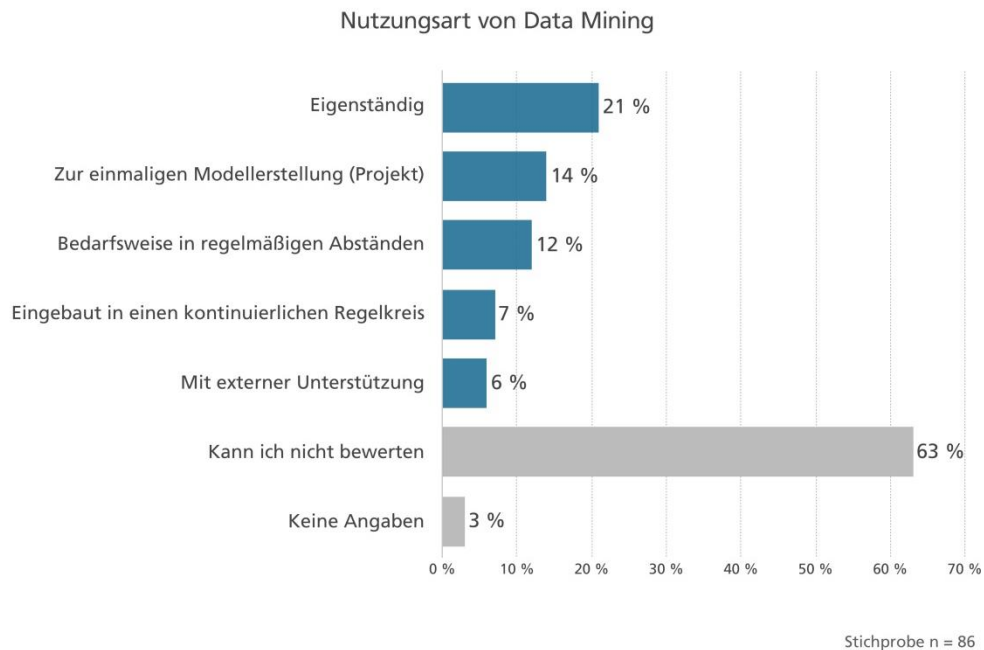


Abbildung 18: Nutzungsart von Data Mining

Lediglich 7% der Umfrageteilnehmer gaben an, dass sie Data Mining kontinuierlich und eingebaut in einen entsprechenden Regelkreis verwenden. Doppelt so viele Befragte gaben an, dass sie Data Mining lediglich projektbezogen verwenden. Die kontinuierliche Nutzung stellt daher tendenziell eher eine Ausnahme dar. 21% der Teilnehmer nutzt Data Mining eigenständig und ohne externe Unterstützung. Lediglich 6% der Befragten bauen bei der Anwendung auf Unterstützung durch externe Berater. Insgesamt ist mit 63% der Anteil derer, die sich diesbezüglich keine Aussage zutrauen, allerdings enorm hoch.

6 Einschätzung zu Data Mining Potenzialen

Ziel des letzten Teils der Befragung war es, einen Überblick über die noch vorhandenen Potenziale für Data Mining Anwendungen im produzierenden Umfeld zu erlangen. Hierzu werden allgemeine Aussagen über Data Mining bewertet und konkrete Anwendungsgebiete für Data Mining abgeleitet. Zunächst wurde gefragt, wo aus heutiger Sicht bereits geeignete Anwendungsfelder für Data Mining gesehen werden. Diese Anwendungsgebiete empfehlen sich aus Sicht der Studienteilnehmer bereits unter den heutigen Rahmenbedingungen für den Einsatz von Data Mining. Im späteren Verlauf sollte ergänzend bewertet werden, wohin sich laut Teilnehmer zukünftig die Potenziale entwickeln werden.

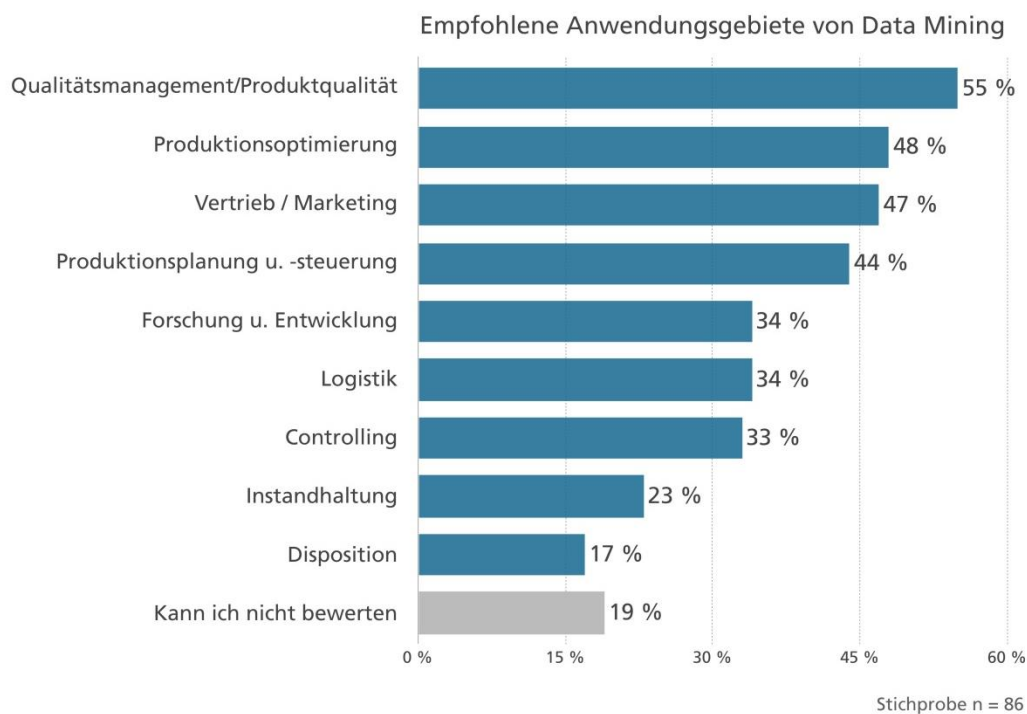


Abbildung 19: Empfohlene Anwendungsgebiete von Data Mining

Ein Drittel der Befragten nahm die Möglichkeit, fünf oder mehr Antworten auszuwählen, in Anspruch. Des Weiteren haben die restlichen zwei Drittel jeweils ein bis zwei oder drei bis vier Auswahlmöglichkeiten genutzt.

Die Mehrheit der Teilnehmer (55%) hält das Qualitätsmanagement sowie die Sicherung der Produktqualität für ein geeignetes Anwendungsfeld für Data Mining. Nahezu die Hälfte der Befragten sieht ebenso in der Produktionsoptimierung, im Vertrieb und Marketing aber auch in der Produktionsplanung und -steuerung Potenziale. Im Gebiet der Forschung und Entwicklung, in der Logistik aber auch im Controlling, ist immerhin ein Drittel der Befragten davon überzeugt Data Mining sinnvoll anwenden zu können. Die Bereiche der Instandhaltung sowie der Disposition werden hingegen als eher untergeordnet bewertet.

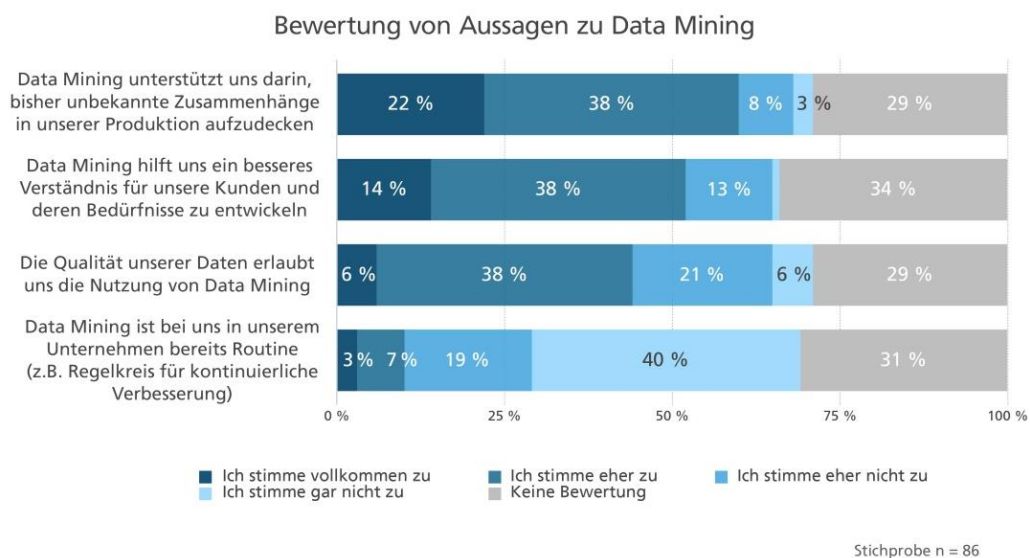


Abbildung 20: Bewertung der Aussagen zu Data Mining

Mit 60% ist über die Hälfte der Befragten davon überzeugt, mit Data Mining bisher noch unbekannte Zusammenhänge in der Produktion aufdecken zu können. Nur jeder Zehnte steht dieser Aussage skeptisch gegenüber.

Durch Data Mining ein besseres Verständnis für Kunden und deren Bedürfnisse zu entwickeln sehen 52% der Befragten als Argument für die Anwendung. Lediglich 1% der Teilnehmer stimmt dieser Aussage gar nicht zu.

Nahezu die Hälfte der Teilnehmer vermutet, die vorhandene Qualität der Daten erlaube die Anwendung von Data Mining im Unternehmen. Rund 27% der Befragten sehen im Hinblick auf die

Datenqualität weiteren Handlungsbedarf. Diese Einschätzung steht damit im Einklang mit der Aussage, dass ca. die Hälfte der Befragten die vorhandenen Daten tendenziell als vollständig charakterisierten.

Nur für 10% der Befragten stellt die Anwendung von Data Mining im eigenen Unternehmen bereits Routine dar. Knapp 60% der Teilnehmer können dieser Aussage noch nicht zustimmen.

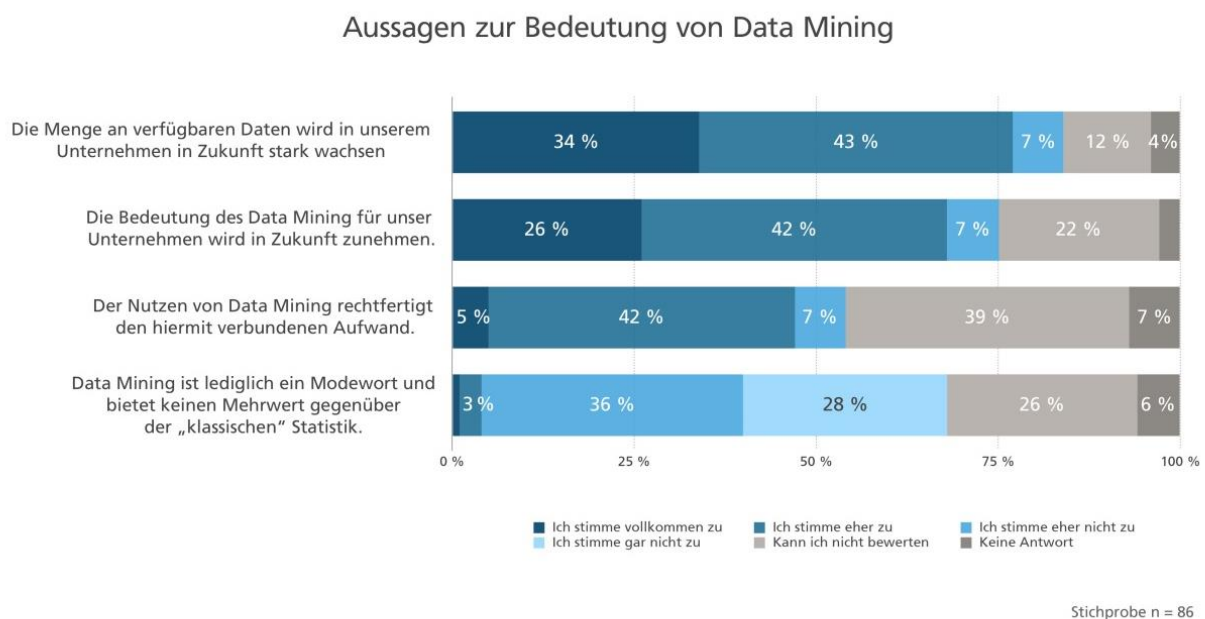


Abbildung 21: Aussagen zu Bewertung von Data Mining

Im Hinblick auf die zukünftige Entwicklung der Datenbestände sind sich die Teilnehmer der Studie größtenteils einig, dass der Datenbestand der Unternehmen stark zunehmen wird. Die dadurch entstehenden Datenmengen mit Data Mining auszuwerten ist für beinahe 70% eine denkbare Alternative zu den manuellen Analysetechniken, die zumeist noch verwendet werden. Des Weiteren stimmen 47% der Befragten zu, dass sich der durch Data Mining verbundene Aufwand durch den damit gewonnenen Nutzen rechtfertigen lässt.

Der Aussage Data Mining sei lediglich ein Modewort und habe keinen Mehrwert im Vergleich zur „klassischen“ Statistik können gerade einmal 4% zustimmen, was verdeutlicht, dass die Potenziale in den Unternehmen bereits deutlich erkannt werden. Konkrete Einsatzpotenziale müssen jedoch

für eine weitere Verbreitung noch klarer dargestellt und individuelle Anwendungsszenarien aufgezeigt werden.

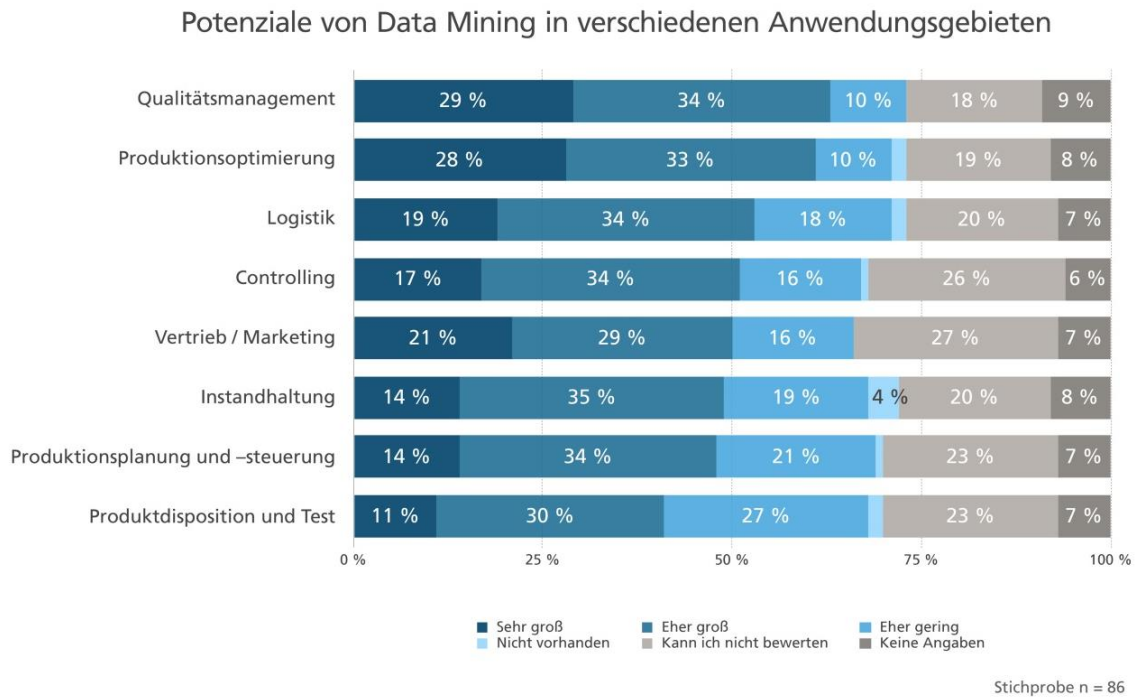


Abbildung 22: Potenziale von Data Mining in verschiedenen Anwendungsgebieten

Zu Beginn des Kapitels wurde bereits beschrieben, welche Anwendungsgebiete aus Sicht der Studienteilnehmer bereits heute empfehlenswert für die Nutzung von Data Mining erscheinen. Dem gegenüber wurde ergänzend nach den zukünftigen Entwicklungspotenzialen von Data Mining in den jeweiligen Anwendungsgebieten gefragt. Demnach wird das Qualitätsmanagement nicht nur als eines der bedeutendsten Anwendungsfelder für Data Mining bewertet, auch das in diesem Bereich zukünftig noch erzielbare Potenzial wird von 63% der Teilnehmer als hoch eingeschätzt. Dicht gefolgt bieten Einsatzfelder der Produktionsoptimierung für 61% der Befragten großes Potenzial für zukünftige Anwendungen des Data Mining. Logistik, Controlling sowie Vertrieb und Marketing weisen nach Meinung der Befragten mit rund 50% ebenfalls eher große Wachstumspotenziale auf. Fast ebenso viele Befragte sehen im Bereich Instandhaltung wie auch bei der Produktionsplanung und -steuerung Chancen für Data Mining. Mit 41% Zustimmung schneidet das Anwendungsfeld der Disposition am schlechtesten ab. Insgesamt zeigt sich jedoch eine deutlich positive Erwartungs-

haltung hinsichtlich zukünftiger Potenziale für Data Mining in den genannten Einsatzfeldern. Abbildung 23 zeigt die Einschätzung des heutigen und des künftigen Potenzials im Vergleich. Auffällig ist, dass das bereits in vielen Teilen der Industrie etablierte Data Mining Themenfeld des Qualitätsmanagements sowie der Bereich der Produktionsoptimierung sowohl in der heutigen als auch in der künftigen Potenzialbewertung deutlich positiv herausstechen. Den Bereichen Vertrieb und Marketing sowie der Produktionsplanung und -steuerung wird bereits heute ein hohes Potenzial beigemessen. Im direkten Vergleich zu den anderen Anwendungsgebieten wird das Entwicklungspotenzial allerdings eher geringer bewertet. Die Bereiche Logistik und Controlling werden sowohl aus heutiger als auch aus künftiger Sicht vergleichsweise mit mittleren Potenzialen bewertet. Überraschend ist, dass dem ebenfalls bereits stark etablierten Feld der Instandhaltung eher geringe Potenziale zugesprochen werden. Herausgestellt werden soll ebenfalls noch einmal, dass trotz der nachrangigen Bewertung der Disposition im direkten Vergleich, immerhin noch 41% der Befragten das Wachstumspotenzial als positiv bewerten.

Potenzialbewertung nach Anwendungsgebiet



Abbildung 23: Heutiges und zukünftiges Potenzial von Data Mining

Um die erwarteten Potenziale auch heben zu können, sind jedoch wie teilweise bereits aufgezeigt, noch einige Hürden zu überwinden. Abbildung 24 zeigt die Einschätzung der Studienteilnehmer bezüglich der größten Herausforderungen, die es für eine erfolgreiche Anwendung von Data Mining im produktionsnahen Umfeld zu meistern gilt.

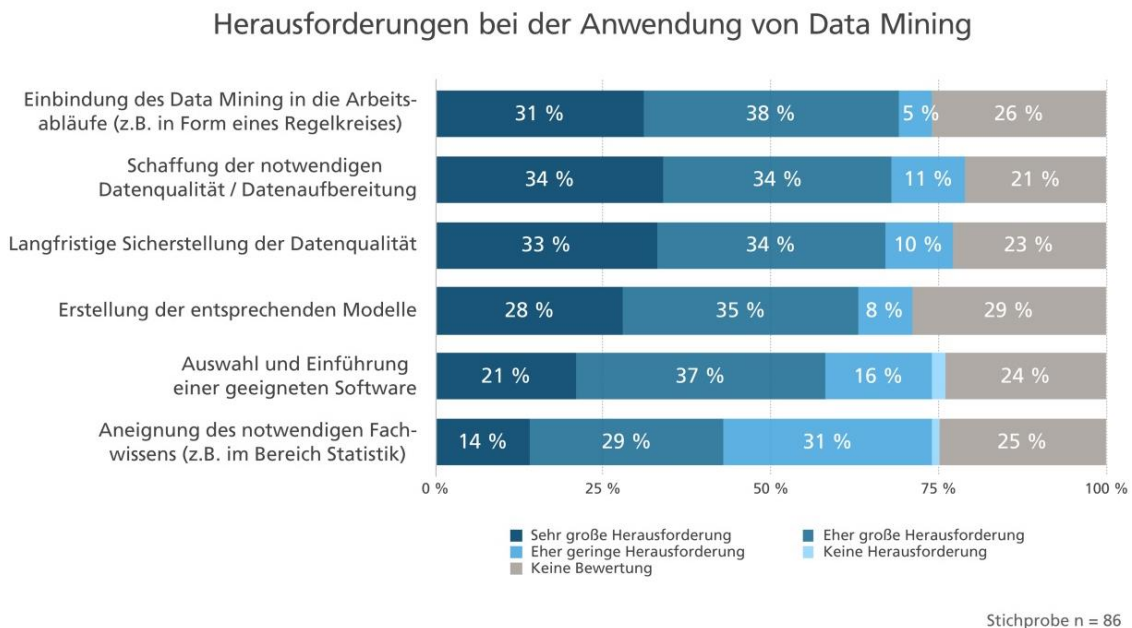


Abbildung 24: Herausforderungen bei der Anwendung von Data Mining

Als größte Herausforderung bei der Anwendung von Data Mining bewerten die Befragten dessen Einbindung in die Arbeitsabläufe. Den Weg bis zur kontinuierlichen Verbesserung in Form eines Regelkreises sehen 69% als eher kritisch an. Die Schaffung der notwendigen Datenqualität sowie die Datenaufbereitung werden beinahe von genauso vielen als herausfordernd bewertet. Die Datenaufbereitung ist somit Bestandteil eines jeden Data Mining Projektes. Auf längere Sicht betrachtet vermuten 67% der Teilnehmer, dass die Sicherstellung der Datenqualität problematisch sei. Gut 60% sehen in der Bildung von Data Mining Modellen eine große Herausforderung. Lediglich 2% der an der Umfrage beteiligten sehen bei der Auswahl und Einführung der geeigneten Software kein Problem, wohingegen 21% dies als sehr große Herausforderung erachten. Letztlich sieht nur ein Drittel der Befragten das Aneignen des notwendigen Fachwissen als schwierig an.

Ein verbreitetes Modell zur Beschreibung eines Data Mining Prozesses stellt das CRISP-DM (Cross Industry Standard Process for Data Mining) dar. CRISP-DM wurde auf Basis von Projekterfahrungen erstellt und zeichnet sich durch seinen hohen Praxisbezug aus. CRISP-DM besteht aus sechs untereinander in Beziehung stehenden Phasen (vgl. Abbildung 25).

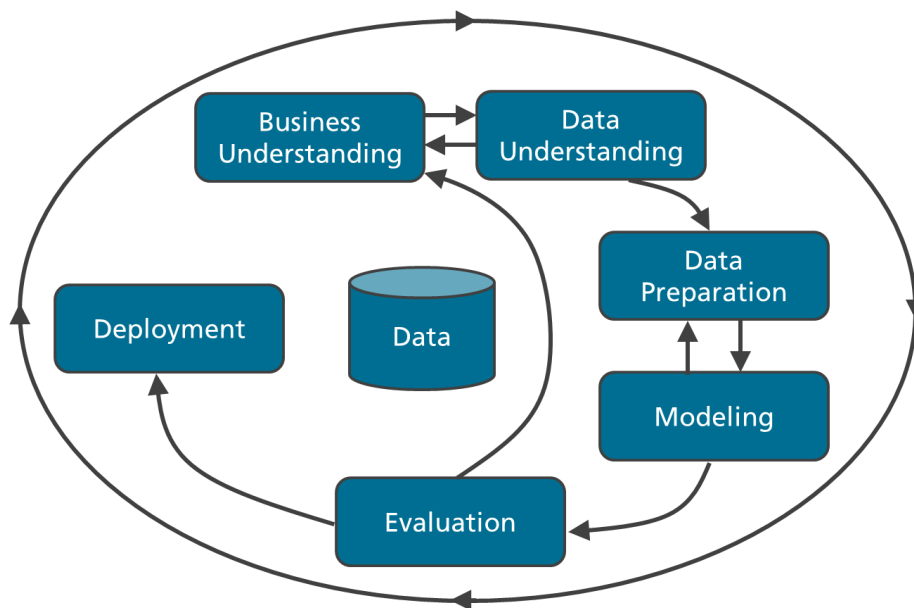


Abbildung 25: Prozessmodell zum CRISP-DM¹

Den Ausgangspunkt bildet der Datenbestand. Die verfügbaren Daten durchlaufen die sechs Bearbeitungsschritte des Modells, wobei die Einzelschritte jedoch keine starre Reihenfolge bilden, sondern ebenfalls Rückschleifen zu vorgelagerten Prozessschritten gestattet oder gar zwingend vorsieht. Der äußere Kreis symbolisiert den kontinuierlichen Data Mining Prozess. Gewonnene Erkenntnisse aus dem Data Mining Prozess können somit beispielsweise ein weiteres Verständnis über die Qualität und Charakteristik der eigenen Geschäftsabläufe schaffen. Das „Business Understanding“ sorgt für ein Verständnis der Geschäftsabläufe sowie für die mit dem Data Mining verfolgten Projektziele. Das „Data Understanding“ schafft Verständnis über den vorliegenden Datenbestand und ist wesentliche Voraussetzung, um Fehlinterpretationen zu vermeiden. „Data Prepara-

¹ Quelle: Chapman et al. (2000): CRISP-DM 1.0 step-by-step data mining guide.

tion“ bezeichnet die Datenaufbereitung selbst. Die Methoden des Data Mining ermöglichen es hierbei, die Datenqualität der vorhandenen Daten signifikant zu verbessern. Hierzu werden verschiedene Methoden angewendet. Das Erkennen von Dubletten, fehlerhaften und unvollständigen Datensätzen sowie das Erkennen von Ausreißern und logischen Fehlern ist Bestandteil der Datenaufbereitung. Somit lassen sich mangelhafte Daten erkennen und die Datenqualität systematisch verbessern. Im Rahmen des „Modeling“ erfolgt die Erstellung des eigentlichen Data Mining Modells, auf dessen Basis die neuen Erkenntnisse erzielt werden sollen. Hier kommen die bereits beschriebenen Methoden des Data Mining zum Einsatz. „Evaluation“ bezeichnet die Bewertung der Modelle. Hierbei wird geklärt, ob die Modelle geeignet sind, die zuvor definierten Anforderungen zu erfüllen und ob sie sich für die Anwendung eignen. „Deployment“ beschreibt die Verwendung der neu gewonnen Erkenntnisse beispielsweise in Form von Verbesserungsmaßnahmen.

Ergänzend zu den Herausforderungen bei der Anwendung von Data Mining wurden die Umfrageteilnehmer bezüglich des künftigen Entwicklungsbedarfs befragt. Hierbei stellen sich aus Sicht der Teilnehmer vor allem Anforderungen an die Softwaresysteme (vgl. Abbildung 26).

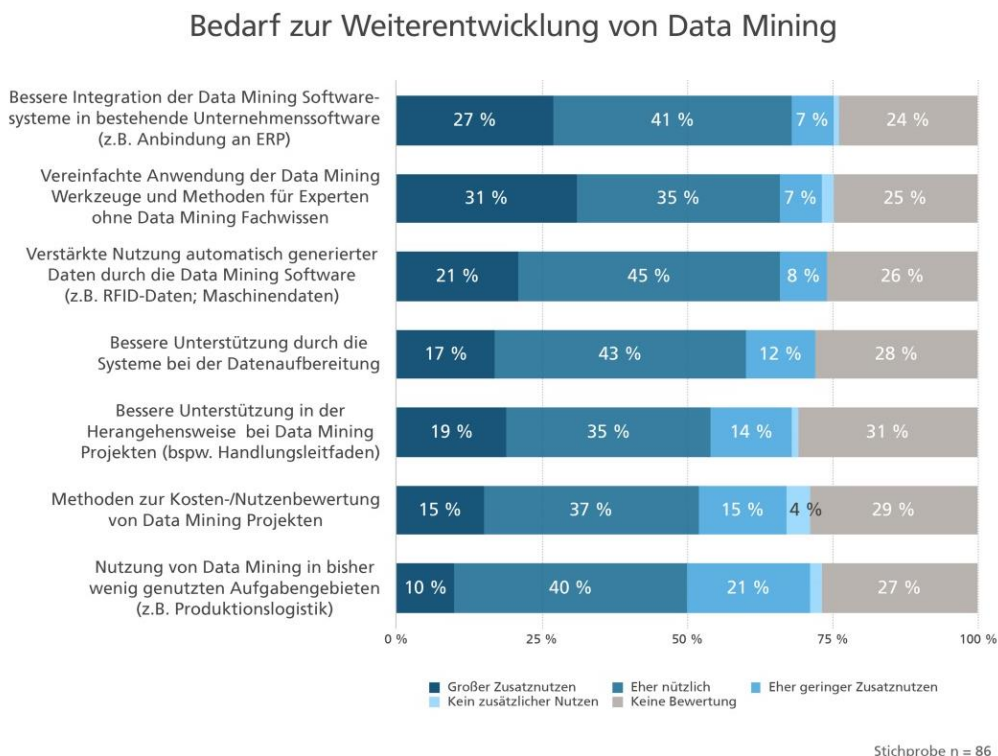


Abbildung 26: Bedarf zur Weiterentwicklung von Data Mining

Knapp 70% der Befragten erhofft sich in Zukunft eine bessere Integration der Data Mining Softwaresysteme in die bestehende Unternehmenslandschaft und würden beispielsweise in der besseren Anbindung an ein ERP einen nützlichen Zusatzeffekt sehen. Knapp ein Drittel der Befragten sieht einen großen Zusatznutzen in einer vereinfachten Anwendung der Data Mining Werkzeuge und Methoden, sodass Mitarbeiter aus anderen Fachgebieten wie beispielsweise der Auftragsplanung auch ohne Expertenwissen im Bereich Data Mining die Methoden und Werkzeuge verwenden können. Die Integration automatisch generierter Daten in die Data Mining Software erscheint für zwei Drittel der Befragten tendenziell nützlich. Derartige Daten könnten beispielsweise in Form von Maschinendaten oder über RFID-Tags in die Systeme gelangen.

Für 60% der Teilnehmer wäre eine bessere Unterstützung der Systeme bei der Datenaufbereitung wünschenswert. Ein Handlungsleitfaden, der die Herangehensweise bei Data Mining Projekten besser unterstützt, wird von ca. der Hälfte der Befragten als denkbare Möglichkeit gesehen, um unerfahrenen Anwendern die Chance zur Nutzung von Data Mining zu erleichtern. In den Methoden zur Kosten- und Nutzenbewertung von Data Mining Projekten sieht immerhin gut die Hälfte der Befragten einen Zusatznutzen. Die Hälfte der Teilnehmer gab an, dass darüber hinaus weiterer Entwicklungsbedarf darin besteht, neue Anwendungsgebiete prinzipiell für Data Mining zu erschließen. Hierbei werden vor allem Erfahrungsberichte aus erfolgreich durchgeführten Projekten dabei helfen, neue konkrete Anwendungsbeispiele aufzuzeigen.

7 Zusammenfassung und Fazit

In einigen produktionsnahen Anwendungsbereichen ist Data Mining als Sammlung von Methoden zur Datenanalyse bereits erfolgreich angekommen und unterstützt bei der Erzeugung von Informationen zur Verbesserung von Produkten und Prozessen. Insgesamt besteht jedoch noch ein großes Potenzial für die Ausweitung auf neue Anwendungsfelder. Oftmals stellt Data Mining dabei noch immer eher ein exotisches Instrument dar, welches vor allem Fachexperten vorbehalten ist. Ein Großteil der in produktionsnahen Bereichen tätigen Mitarbeiter hat bestenfalls lediglich ein grobes Verständnis über die Methoden und Werkzeuge des Data Mining. Dennoch wird den Methoden des Data Mining zukünftig Potenzial für weitere Anwendungen zugetraut. Diese Potenziale werden vor allem in Bereichen des Qualitätsmanagements, der Produktionsoptimierung sowie in Logistik und Controlling gesehen.




Bei der Umsetzung der beschriebenen Potenziale stellen aus unternehmensinterner Sicht vor allem eine ungenügende Datenqualität, mangelndes Fachwissen sowie eine unzureichende organisatorische Verankerung von Data Mining Hürden dar, die es zu überwinden gilt. Für Softwareanbieter sollten vor allem Themenstellungen einer besseren Integration in die bestehende Unternehmenssoftware (wie beispielsweise ERP und MES) sowie die Verbesserung der Bedienbarkeit auch für Anwender ohne spezifische Data Mining Fachkenntnisse auf der Agenda stehen. Nicht zuletzt entscheiden vor allem die angebotenen Funktionalitäten über den Erfolg der jeweiligen Softwarelösung.

Die Relevanz des Umgangs mit den erfassten Daten in der Produktion wird in Zukunft weiter zunehmen. Aktuell wird in der Industrie und in den Medien das Thema „Industrie 4.0“ stark diskutiert, unter der die zunehmende informationstechnische Vernetzung innerhalb der Produktion zusammenfassend betrachtet wird. Die stärkere Einbeziehung der Informationstechnik in die Produktion führt bereits heute zur Erzeugung riesiger Datenmengen. Moderne, sensorreiche Produktionsmaschinen oder Produkte mit entsprechenden heterogenen Sensoren tragen in den kommenden Jahren zu einer großen Menge an automatisch erzeugten Informationen mit hoher Datenqualität bei. Die Entwicklungen im Bereich Industrie 4.0 werden somit die Relevanz und die Verbreitung von Data Mining in produktionsnahen Bereichen weiter steigern. Die Ausweitung der automatisierten

Datenerfassung in Form von Bewegungsdaten auf Ebene einzelner Produktionsaufträge wird beispielsweise völlig neuartige Potenziale und Anwendungsfelder für Data Mining ermöglichen. So werden sich für Unternehmen, welche dem Thema aufgeschlossen gegenüberstehen, gute Möglichkeiten zur datengestützten Optimierung von Produkten und Produktionsabläufen ergeben.

8 Unterstützer der Studie

Besonderer Dank gilt der Firma IBM, welche durch die finanzielle Unterstützung die Durchführung der Studie ermöglicht hat. IBM ist einer der weltweit größten Anbieter von Informationstechnologie (Hardware, Software und Services) und B2B-Lösungen.

1	<ul style="list-style-type: none">▪ Größtes statistisch-mathematisches Forschungsteam weltweit seit 1960▪ #1 US-Patent-Rang; >500 Analytics-Patente pro Jahr	» Modernste Analytics-Forschung & Innovation	
2	<ul style="list-style-type: none">▪ > 16 Mrd. \$ Investitionen in Software-Akquisitionen zu Analytics seit 2005▪ Integriertes Portfolio Analytics-BI-IM	» Umfassendes Portfolio von Analytics-Technologie & Software	
3	<ul style="list-style-type: none">▪ Über 9.000 erfahrene Berater/Spezialisten für Analytics-Strategie & -Technologie▪ Spezifische Lösungen für unterschiedliche Anwendungsfälle & Branchen	» Erfahrung & Kompetenz in der Analytics-Beratung	

Mit der Akquisition von SPSS im Jahr 2010 integrierte die IBM das marktführende Unternehmen im Bereich der statistischen Datenanalyse in ihr Business-Analytics-Portfolio. IBM SPSS ist der Spezialist für Predictive Analytics und seit 46 Jahren der weltweit führende Anbieter in den Bereichen Statistik, Analyse, Data und Text Mining.

Der IBM SPSS Modeler ist die Data-Mining-Workbench von IBM, die den gesamten Analyseprozess durch eine einzigartige grafische Benutzeroberfläche darstellt.

Neben umfangreichen Datenimport- und Aufbereitungsverfahren bietet der SPSS Modeler zudem eine breite Auswahl an leistungsfähigen Analyse- und Modellierungstechniken zur Klassifikation, Anomaliendetektion, Segmentierung und Assoziation. Der IBM SPSS Modeler zeichnet sich durch seine leistungsfähigen Datenaufbereitungsfunktionen aus, mit denen Anwender mühelos alle gesammelten Daten erfassen, transformieren, bereinigen und validieren können. Einen besonderen Vorteil stellt der hohe Bedienkomfort der IBM Data-Mining-Lösung dar. Durch umfassende grafische Möglichkeiten und intuitiv gestaltete Abläufe unterstützt die Software sowohl eine effiziente Arbeitsweise im täglichen Einsatz als auch die rasche Einarbeitung von neuen Mitarbeitern. Auf-

grund der Benutzerfreundlichkeit und der kurzen Implementierungszeit liefert die hochperformante Data-Mining-Workbench von SPSS schnell einen positiven ROI.

Vor allem im Bereich Operational Excellence – also der Anwendung von Data Mining Verfahren auf technische Daten – hat IBM SPSS in zahlreich erfolgreich durchgeführten Projekten seine Marktführerschaft untermauert.

So konnte ein Automobilhersteller mit dem Einsatz der SPSS Data-Mining- Lösung den Ausschuss bei der Produktion von Kurbelgehäusen innerhalb von 12 Wochen um 80 Prozent reduzieren. In einem weiteren Projekt konnte in einer Gießerei die Ausschussrate von 9 auf 1,5 Prozent gesenkt und die Geradeaus-Produktion um 20 Prozent verbessert werden – bei einem ROI von 8 Wochen.

Des Weiteren bedanken wir uns bei Herrn Jochen Koch von der Firma Technology Consulting Koch für seine fachliche Unterstützung der Studie.



KONTAKT

Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA

Ansprechpartner

Markus Weskamp

Projektleiter Auftragsmanagement und Wertschöpfungsnetze

Telefon +49 711 970-1149

markus.weskamp@ipa.fraunhofer.de

Anja Schatz

Abteilungsleiterin Auftragsmanagement und Wertschöpfungsnetze

Telefon +49 711 970-1076

anja.schatz@ipa.fraunhofer.de

www.ipa.fraunhofer.de