

## Vorwort

Die Fertigungsindustrie hat gerade begonnen, Analytics – deutsch als Analytik bezeichnet – auf der Grundlage von gespeicherten und aktuellen Prozess- und Produktionsdaten zu nutzen. Diese Daten können dann zur besseren und schnelleren Entscheidungsfindung in Echtzeit und zur Verbesserung von Prozessen und Abläufen genutzt werden. Ohne Zweifel ist die Analytik mit ihren mathematischen und statistischen Methoden der wichtigste Beschleuniger der Digitalisierungs- und Transformationsbemühungen von Industrieunternehmen.

In diesem Fachbuch wird aufgezeigt, was Analytik in der Industrie ist, warum diese Schlüsseltechnologie für die digitale Transformation wichtig ist und wie an Anwendungsfälle herangegangen werden kann. Dazu habe ich das Buch in sechs aufeinander aufbauende und sich ergänzende Kapitel gegliedert.

Bei der Analytik werden Daten gesammelt, neu strukturiert und analysiert, um Erkenntnisse zu gewinnen, welche Entscheidungsfindungen unterstützen und bei der Durchführung von Aktionen helfen. Je nach Aufgabe und Ziel gibt es verschiedene Methoden und Techniken zur Umsetzung. Um ein breiteres Verständnis zu erreichen, werden im **Kapitel 1** zunächst die Methoden der Analytik allgemein beschrieben und klassifiziert sowie notwendige Grundlagen erläutert.

Im darauffolgenden **Kapitel 2** wird die Verarbeitungskette der Datenanalyse vorgestellt. Diese Verarbeitungskette beschreibt den Prozess der Datensammlung und -nutzung von der Ermittlung des Datenbedarfs, über die Datenerstellung bis zur endgültigen Verwendung und möglichen Wiederverwendung sowie die Umwandlung in verwertbare Informationen. Hierbei werden die einzelnen Schritte beschrieben und im Detail erläutert. Diese Erläuterungen helfen dabei, einen besseren Einblick in die notwendigen Datenanalyseschritte zu bekommen.

Machine Learning/ML befasst sich mit dem Erkennen von Gesetzmäßigkeiten und Mustern aus Daten sowie der eigenständigen Identifizierung und Entwicklung von Lösungen. Im **Kapitel 3** werden dazu gängige Algorithmen am Referenzbeispiel einer virtuellen Smart Factory vorgestellt und erläutert. Algorithmen lassen sich in verschiedene Lernkategorien einteilen und werden oft nach Ähnlichkeit in Bezug auf ihre Funktion gruppiert. Diese repräsentative Auflistung verschafft einen guten ersten Einblick.

Im **Kapitel 4** erfolgt die Beschreibung der Rollenprofile mit entsprechenden Qualifikationen der Mitarbeiter. Qualifikationen sind allgemein Kenntnisse und Fähigkeiten eines Mitarbeiters, die er im Laufe seines Berufslebens erworben hat. Aus den verschiedenen Rollen lassen sich dann zugewiesene Verantwortlichkeiten ableiten. Die Beschreibung der Rollen kann Unternehmen unterstützen, im Projektvorfeld eventuelle Lücken zu entdecken und die Kompetenzentwicklung ihrer Mitarbeiter stärker auszubauen.

Das im **Kapitel 5** beschriebene Vorgehensmodell ist eine idealtypische Abbildung und stellt mit seinen Phasen eine Arbeitshilfe dar. Diese ist an die Gegebenheiten des jeweiligen Anwendungsfalles anzupassen. Die Phasen können durch einzelne Prozesspakete ergänzt werden. Vor der technischen Validierung des Modells sollte überprüft werden, welche Fähigkeit der Algorithmus selbst hat, Daten zu verarbeiten. Es wird beschrieben, mit welchen Mitteln sich die Qualität von Modellen messen lässt und was getan werden kann, um sicherzustellen, dass

Modelle nicht nur für wenige Trainingsdaten, sondern möglichst robust auch bei neuen Daten funktionieren.

Die Erfolgsfaktoren von Anwendungsfällen werden im **Kapitel 6** präsentiert. Grundlage ist die Fachkompetenz von Mitarbeitern, ebenso wie die Bereitschaft, die Handlungssituation zu erfassen und das Problem zu lösen. Prozessingenieuren mit ihrer ingenieurtechnischen Denkweise und einem tiefen Prozessverständnis kommt hierbei eine Schlüsselrolle zu. Weiterhin werden Auswahlkriterien für Anwendungs-Software zur industriellen Analytik beschrieben sowie Proficy CSense von GE Digital als flexibles Software-Framework mit seinen Grundbausteinen vorgestellt.

Proficy CSense gewinnt Wissen aus historischen Daten und liefert umsetzbare Erkenntnisse. Optimierungs- sowie adaptive Steuerungslösungen lassen sich damit schnell entwickeln, testen und zur Anwendung bringen. **Kapitel 7** beschreibt die Durchführung von drei Anwendungsfällen in industriellen Einsatzgebieten, in denen die beschriebenen Grundlagen und Methoden validiert wurden. Dazu zählt die Verbesserung der Produktqualität und Abfallreduzierung in der Papier- und Kartonproduktion, das Regelkreismanagement zur Minimierung der Prozessschwankungen in der Chromerzverarbeitung und die Verbesserung der Energieeffizienz und Optimierung eines Heizsystems von Fertigungshallen in der Auftragsfertigung eines Zulieferunternehmens.

Hiermit möchte ich meinen Dank all jenen aussprechen, die am Entstehen dieses Fachbuches mitgewirkt haben. Den Autoren aus Deutschland, Österreich und der Schweiz gilt meine besondere Anerkennung, weil sie den qualifizierten Wissenstransfer mit Ihrer Kompetenz und Erkenntnissen erst ermöglicht haben.

Wir wünschen allen Leserinnen und Lesern viel Inspiration beim Durchlesen und vor allem viel Erfolg bei der Umsetzung in der Praxis.

Thomas Schulz  
Herausgeber

# Inhaltsverzeichnis

<b>Vorwort</b> .....	5
<b>Abkürzungen</b> .....	9
<b>1 Methoden der Analytik</b> .....	11
THOMAS SCHULZ	
1.1 Operational Intelligence .....	13
1.1.1 Deskriptive Analytik .....	13
1.1.2 Diagnostische Analytik .....	14
1.2 Advanced Analytics .....	15
1.2.1 Prädiktive Analytik .....	15
1.2.2 Präskriptive Analytik .....	16
<b>2 Die Verarbeitungskette der Datenanalyse</b> .....	19
Prof. Dr.-Ing. MARCO HUBER, HENRIK OPPERMANN	
2.1 Datenerfassung .....	19
2.2 Datenvorverarbeitung .....	21
2.2.1 Datenbereinigung .....	22
2.2.2 Datentransformation .....	22
2.2.3 Datenreduktion .....	23
2.2.4 Datenintegration .....	23
2.3 Datenanalyse .....	23
2.3.1 Data Mining und Maschinelles Lernen .....	23
2.3.2 Explorative Datenanalyse .....	25
2.3.3 Korrelation vs. Kausalität .....	25
2.3.4 Stromdaten-Analyse (Complex Event Processing) .....	27
2.4 Ergebnisdarstellung und -bewertung .....	27
<b>3 Maschinelles Lernen</b> .....	29
BARIS AYAZ	
3.1 Maschinelles Lernen in der Smart Factory .....	30
3.2 Klassifikation .....	32
3.3 Modellvalidierung .....	35
3.4 Regression .....	38
3.5 Clustering .....	41
<b>4 Rollenprofile</b> .....	45
JOHANNES KRÖCKEL, PH.D.	
4.1 Datenanalyst / Data Scientist .....	46
4.2 Data Engineer / Data Architect .....	48
4.3 Business Analyst / Fachbereich .....	49
4.4 Software-Entwickler / Systemadministrator .....	49

<b>5</b>	<b>Vorgehensmodell</b> .....	53
	<b>JOHANNES KRÖCKEL, PH.D.</b>	
5.1	Business-/Use-Case-Verständnis .....	54
5.2	Evaluation und Optimierung von Modellen .....	56
5.2.1	Qualitätskennzahlen .....	56
5.2.2	Validierung der Qualität .....	61
5.2.3	Optimierungsmöglichkeiten .....	63
<b>6</b>	<b>Erfolgsfaktoren für Anwendungen</b> .....	69
	<b>DENIS REIMANN</b>	
6.1	Regelkreise mit Sensorsystemen .....	69
6.2	Prozesswissen in den Fachbereichen .....	71
6.3	Auswahlkriterien für Anwendungssoftware .....	73
6.4	Proficy CSense – ein Software-Framework für industrielle Analysen .....	74
6.4.1	Troubleshooter .....	75
6.4.2	Architect .....	76
6.4.3	Action Object Manager .....	78
<b>7</b>	<b>Anwendungsfälle Analytik</b> .....	81
	<b>HARALD TASCHKE, REMO INGOLD, THOMAS SCHULZ</b>	
7.1	Skjern Paper nutzt KI zur Verbesserung Produktqualität und Abfallreduzierung .....	82
7.2	Chromerzverarbeiter optimiert seinen komplexen Produktionsprozess .....	87
7.3	Optimierung des Heizsystems von Fertigungshallen im Schichtbetrieb .....	91
	<b>Anhang: Kategorien und Verfahren Advanced Analytics</b> .....	99
	<b>THOMAS SCHULZ</b>	
A1	Übersicht und Kategorien .....	99
A2	Verfahren prädiktive Analytik .....	100
A3	Verfahren präskriptive Analytik .....	102
A4	Integrierte Verfahren Proficy CSense .....	105
	<b>Lebensläufe der Autoren</b> .....	107
	<b>Stichwortverzeichnis</b> .....	109

## Abkürzungen

AI	Artificial Intelligence
AND	logischer Operator Konjunktion
APC	Advanced Process Control
CEP	Complex Event Processing
CRISP-DM	Cross Industry Standard Process for Data Mining
CSV	Comma-separated values
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DCS	Distributed Control System
DevOps	Development and IT Operations
DL	Deep Learning
EM	Erwartungs-Maximierung
ETL	Extract, Transform, Load
F	Flüssigkeitsdurchfluss
FN	falsch negativ
FP	falsch positiv
GUI	Graphical User Interface
I	Stromaufnahme
IIoT	Industrial Internet of Things
IoT	Internet of Things
IT	Informationstechnik
KI	Künstliche Intelligenz
KN	korrekt negativ
KNN	künstliche neuronale Netze
KP	korrekt positiv
KPI	Key Performance Indicator
M1	Maschine 1
M2	Maschine 2
M3	Maschine 3
ML	Machine Learning
MMC	Microsoft Management Console
MQF	mittlerer quadratischer Fehler
MSE	Mean Squared Error
MWh	Megawattstunde
NOT	logischer Operator Negation
OEE	Overall Equipment Effectiveness
OEM	Original Equipment Manufacturer
ONNX	Open Neural Network Exchange
OR	logischer Operator Disjunktion

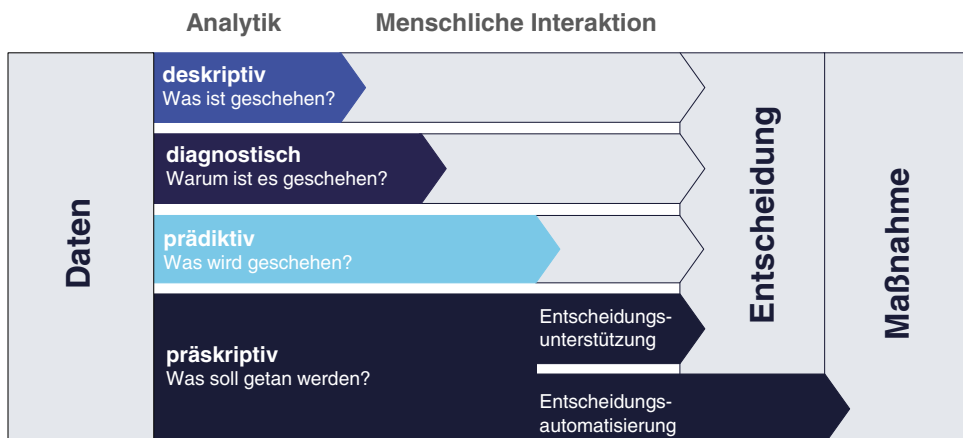
OSI	Open Systems Interconnection
OT	Operation Technology
PCA	Principal Component Analysis
Pi	Eringangsdruck
PID	Proportional Integral Derivative
PLS	Partial Least Squares
Po	Ausgangsdruck
ppm	parts per million
QS	Qualitätssicherung
R <sup>2</sup>	Bestimmtheitsmaß
RN	richtig negativ
RP	richtig positiv
SCADA	Supervisory Control and Data Acquisition
SPS	speicherprogrammierbare Steuerung
SRCUM	Vorgehensmodell des agilen Projektmanagements
TPOT	Tree-based Pipeline Optimization Tool)

# 1 Methoden der Analytik

Das verarbeitende Gewerbe erzeugt mehr Daten als jeder andere Wirtschaftszweig. Es wird erwartet, dass die grenzüberschreitenden Datenströme schneller wachsen werden als der Welt-handel [1.1]. Datenanalyse und Künstliche Intelligenz sind dabei Schlüsseltechnologien in den zukünftigen Digitalisierungsbemühungen der Industrie. Für Unternehmen ist es jedoch immer noch oft eine enorme Herausforderung, den Nutzen dieser Technologien in ihrem Produktionssystem zu erkennen. Immer ausgereifere Algorithmen werden eingesetzt, um die menschliche Tätigkeit zu unterstützen, und dies nicht nur, wenn Entwicklungen analysiert und vorhergesagt werden sollen, sondern auch, wenn tatsächlich Entscheidungen getroffen werden, die sich auf die Fabrik, Produktion und Mitarbeiter auswirken.

Auch wenn die Begriffe Analyse und Analytik manchmal synonym verwendet werden, unterscheidet sich ihre Bedeutung.

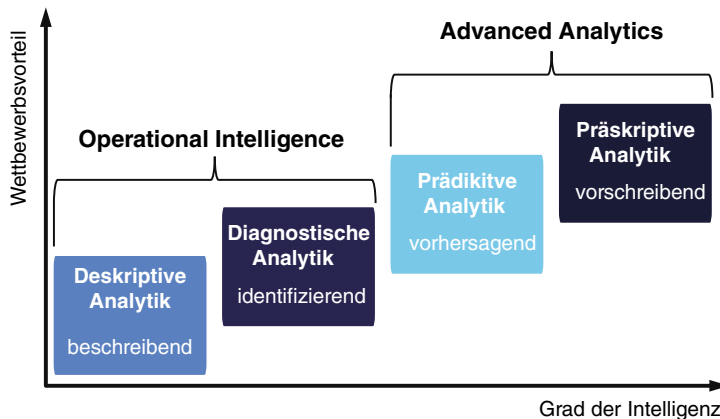
- **Analyse** bezieht sich auf den Prozess der Zerlegung eines Gesamtproblems in seine Teile, oftmals einhergehend mit der Umwandlung von Daten in kontextbezogene und zuverlässige Informationen. Diese Teile werden auf der granularen Ebene untersucht, um anschließend Entscheidungen zu treffen. Analyse wird häufig bei komplexen Systemen eingesetzt, bei denen die Untersuchung des Gesamtsystems nicht möglich oder unpraktikabel ist und daher vereinfacht werden muss, indem es in verständlichere Komponenten zerlegt wird. Sobald die Entscheidungen auf der granularen Ebene realisiert sind und die Untersuchung der Teile abgeschlossen ist, wird das gesamte System mit Hilfe einer Synthese wieder zusammengesetzt.
- **Analytik** ist die Vielfalt an Methoden, Technologien und zugehörigen Werkzeugen zur Gewinnung von neuem Wissen und Einblicken zur Lösung komplexer Probleme. Analytik ist ein vielschichtiger und multidisziplinärer Ansatz. Diese Herangehensweise ermöglicht es einerseits, Verständnis für komplexe Situationen zu entwickeln, andererseits soll sie dazu beitragen, diese anschließend auch zu bewältigen. Analytik macht sich Daten und mathematische Modelle zunutze, um der immer komplizierter werdenden Welt einen Sinn zu geben. Obwohl die Analytik den Vorgang der Analyse in verschiedenen Stadien des Entdeckungsprozesses beinhaltet, ist sie nicht nur eine Analyse, sondern umfasst auch die Synthese und die anschließende Umsetzung. Mehr als alles andere ist sie eine Methodik, welche wiederum eine Vielzahl von Methoden und Praktiken umfasst.



**Bild 1.1** Möglichkeiten durch Analytik [1.2]

Dank des stetig voranschreitenden technologischen Fortschritts sammeln Industrieunternehmen eine enorme Menge an Daten. Es gibt nun mehrere Möglichkeiten, diese Daten sinnvoll zu nutzen. Welche Methode Sie wählen, hängt von der aktuellen Aufgabenstellung ab sowie von den Informationen, die Sie aus Ihrem Datensatz gewinnen möchten. Insgesamt können Anwendungen durch vier Arten der Analytik charakterisiert werden: deskriptiv, diagnostisch, prädiktiv und präskriptiv. Bild 1.1 enthält eine Übersicht dieser vier Arten der Analytik [1.2]. Sollen beispielsweise bestehende Abläufe analysiert und Ursachen untersucht werden – was ist passiert und warum – so eignen sich hierfür insbesondere die deskriptive und diagnostische Analytik. Fragestellungen, die sich auf zukünftige Ereignisse und/oder in der Zukunft liegende Handlungsabläufe beziehen, können hingegen mit Hilfe der prädikativen und/oder präskriptiven Analytik beleuchtet werden.

Die Bandbreite von Datenanalysetechniken und -prozessen reicht von beschreibenden Diagnose-Methoden zur reaktiven Entscheidungsfindung bis hin zu anspruchsvollen Vorhersagen und Optimierungen zur proaktiven Entscheidungsfindung. Wie in Bild 1.2 dargestellt, sind die einfacheren Prozesse noch unter dem Namen Operational Intelligence bekannt und der Begriff Advanced Analytics hat sich für prädiktive und präskriptive Analytik durchgesetzt [1.3], [1.4], [1.5].



**Bild 1.2** Reifegradmodell Analytik (modifiziert nach [1.6] und [1.7])

Die Möglichkeiten der Analytik sollen Industrieunternehmen in die Lage versetzen, schnellere, bessere und intelligenter Entscheidungen zu treffen. Bislang liegt der Schwerpunkt im akademischen und industriellen Bereich auf der diagnostischen und prädiktiven Analytik. Die präskriptive Analytik, die darauf abzielt, die beste Vorgehensweise für die Zukunft zu finden, hat in den letzten Jahren zunehmend das Interesse der Wissenschaft geweckt. Einen tieferen Einblick in dieses Forschungsfeld bietet der Artikel «Prescriptive analytics: Literature review and research challenges» [1.8]. Dabei werden aktuelle Herausforderungen identifiziert und zukünftige Richtungen und Entwicklungen skizziert. Letztendlich kann präskriptive Analytik als der nächste Schritt zur Erhöhung der Reife der Datenanalyse angesehen werden und dient damit nachhaltigen Verbesserung der Unternehmensleistung.



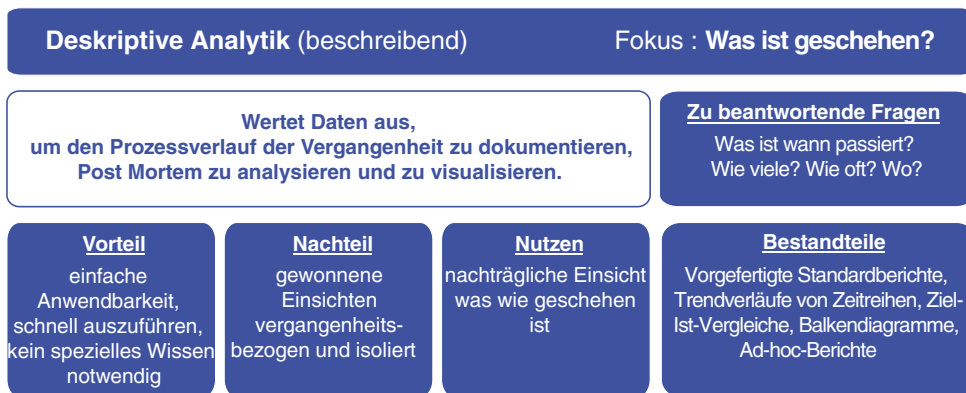
## 1.1 Operational Intelligence

Operational Intelligence ist eine Form der dynamischen Echtzeit-Datenanalyse, die Transparenz und Einblicke in Prozesse und Abläufe liefert. Es ist ein Ansatz zur Analyse maschinengenerierter Daten, der es ermöglicht, Entscheidungen und Maßnahmen auf der Grundlage von erzeugten und gesammelten Echtzeitdaten zu treffen [1.9]. Zu ihr gehören die Verfahren der deskriptiven und der diagnostischen Analytik.

Im überwachten Modus arbeiten Systeme auf der Grundlage von Daten, die von Menschen aufgrund ihrer Eigenschaften und ihrer bekannten Auswirkungen auf das Ergebnis ausgewählt wurden. Die Intelligenz wird in diesem Fall hauptsächlich durch den Menschen bereitgestellt. Systeme werden dann vor allem für Berechnungsleistungen auf der Grundlage mehrerer Datenreihen herangezogen. Anwendungen in diesem Bereich richten sich in erster Linie an Mitarbeiter, welche nah am Prozess arbeiten und die so durch den Zugang zu aktuellen Daten besser informierte Entscheidungen treffen oder schneller auf Probleme reagieren können.

### 1.1.1 Deskriptive Analytik

Deskriptive Analytik ist die Untersuchung von eingehenden Daten zur Steuerung von Maßnahmen und Warnungen (Bild 1.3). Mit Hilfe der deskriptiven Analytik kann ein Unternehmen aus seinen Daten Erkenntnisse gewinnen und feststellen, was im vorangegangenen Zeitraum geschehen ist. Hauptbestandteile sind Standardberichte für kurzfristige Zielsetzungen (Was ist passiert? Wann ist es passiert?) sowie Ad-hoc-Berichte für Problemlösungen (Wie viele? Wie oft? Wo?).



**Bild 1.3** Deskriptive Analytik

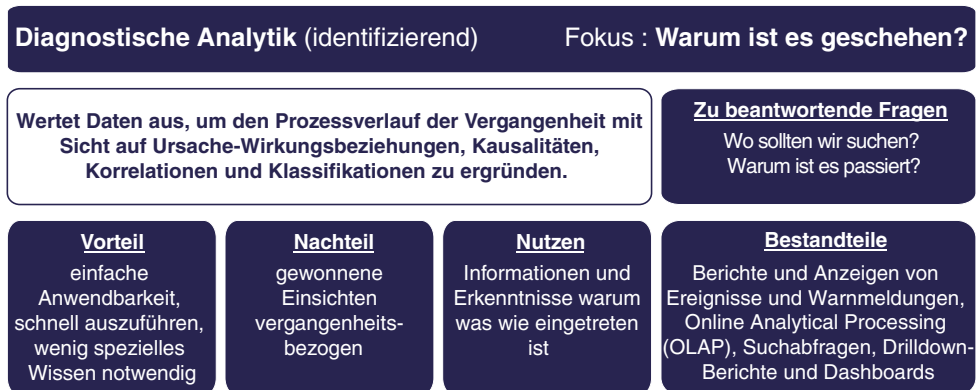
Die deskriptive Analytik ist die Einstiegsebene in die Analysetaxonomie und beinhaltet die Rückwärtsbeobachtung zum Auffinden von Vorkommnissen [1.10]. Sie erfordert die Sammlung relevanter Daten aus der Vergangenheit, die Verarbeitung der Daten und ihre Darstellung in visualisierter Form. Sie stellt dar, was in der Vergangenheit geschehen ist. In industriellen Anwendungen ist die Anzahl der Attribute in der Regel größer bis hin zu Dutzenden, Hunderten oder sogar noch mehr. Wenn bei der Analyse eines Datensatzes mehr als zwei Attribute untersucht werden, spricht man von einer multivariaten Analyse.

Sie wird oft auch als Berichterstattung bezeichnet, da die meisten Analyseaktivitäten auf dieser Ebene die Erstellung von Berichten zur Zusammenfassung von Aktivitäten und Abläufen betref-

fen, um die Frage «Was ist wann passiert?» zu beantworten. Das Spektrum dieser Berichte umfasst statische Momentaufnahmen nach einem festen Zeitplan ebenso wie dynamische Ansichten von Indikatoren. Die einfachste und schnellste Art Daten zu untersuchen, ist die Verwendung von Häufigkeitstabellen und Balkendiagrammen. Mit Hilfe der Pivot-Tabellenfunktion in Excel können beispielsweise eine Vielzahl von Informationen dargestellt werden. Bei Ad-hoc-Berichten hat der Anwender die Möglichkeit, seinen eigenen spezifischen Bericht zu erstellen, um eine individuelle Entscheidungssituation darzustellen.

### 1.1.2 Diagnostische Analytik

Diagnostische Analytik ist die intelligente Analyse von Daten zum Erkennen von Mustern und Ursachen, sie befasst sich mit der Frage nach dem «Warum». Sie bietet eine tiefere Analytik als die deskriptive Analytik, indem sie Ursachen von und Reaktionen auf eine Situation identifiziert (Bild 1.4). Mit ihr sind komplexere Analyseabfragen für Ursachenanalyse und Stratifikation möglich (Wo genau ist das Problem? Wie finde ich die Antwort? Wie lautet die Ursache für den vorliegenden Sachverhalt?).



**Bild 1.4** Diagnostische Analytik

Das Erkennen von Anomalien wird für Anwendungen als erster Schritt verwendet, bei dem es wichtig ist, eine Abweichung von einem erwarteten Muster zu erkennen. Das Erkennen von Anomalien bedeutet dabei, Unterbrechungen oder unerwartete Veränderungen in einer Aktivität aufzufinden. Dies kann in einer Vielzahl von Szenarien der Fall sein. Unüberwachte Verfahren erkennen Anomalien in einem unmarkierten Testdatensatz unter der Annahme, dass die Mehrheit der Instanzen im Datensatz normal ist, sie geben dann die Instanzen als Anomalien an, die am wenigsten zum Rest des Datensatzes zu passen scheinen [1.11].

Der zweite Schritt ist die Untersuchung der Anomalie. Dabei geht es darum herauszufinden, wie solche technischen Fehlerzustände zustande kamen. Bei diesem Vorgang werden statistisch signifikante Datenanomalien während eines bestimmten Zeitraumes erkannt.

Der dritte Schritt ist Untersuchung von Beziehungen zwischen den Anomalien. Hierbei helfen statistische Analysen wie Korrelations-, Diskriminanz- und Regressionsanalyse. Die Korrelationsanalyse untersucht die Abhängigkeit zweier Variablen oder Merkmale und gibt Auskunft über deren Zusammenhang. Die Diskriminanzanalyse ist eine Methode multivariater Verfahren, die dazu dient, zwei oder mehr Gruppen zu unterscheiden, die durch mehrere Merkmale (auch

Variablen) beschrieben werden. Regressionsanalysen berechnen, ob eine oder mehrere unabhängige Variable eine abhängige Variable beeinflussen. Mit diesen Analytikmethoden ist es möglich, Antworten auf die Fragen zu geben, warum etwas passierte und welche Möglichkeiten der Fehlerabwehr versäumt wurden.

## 1.2 Advanced Analytics

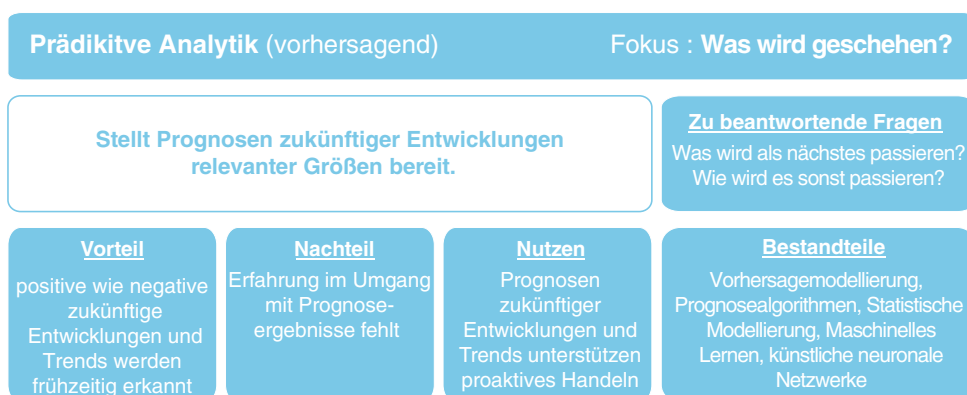
Advanced Analytics ist eine Methode der Datenanalytik, die Modellierung, maschinelle Lernalgorithmen, Deep Learning, Prozessautomatisierung und andere statistische Methoden zur Analyse aus einer Vielzahl von Datenquellen verwendet [1.12]. Zu ihr gehören die Verfahren der prädiktiven und der präskriptiven Analytik.

Advanced Analytics nutzt die Datenwissenschaft über die traditionellen Operational Intelligence-Methoden hinaus, um Muster vorherzusagen und die Wahrscheinlichkeit zukünftiger Ereignisse abzuschätzen. Dies kann sich vorteilhaft auf die Produktionsprozesse auswirken. Es kann nicht nur eine reaktionsschnellere Leistung erreicht, sondern auch die Genauigkeit bei der Entscheidungsfindung potenziert werden.

Breitere Anwendungen von Softwarelösungen erfolgten mit der Weiterentwicklung der Steuerungstechnik und dem Einzug der Computertechnik in die Industrie. Beispielhaft für Erstellung und Einführung von frühen Softwaresystemen ist an dieser Stelle die Arbeit von János REITH [1.13].

### 1.2.1 Prädiktive Analytik

Mit Hilfe der prädiktiven Analytik können Entwicklungen vorhergesagt und dargestellt werden. Es erfolgt eine Vorhersage von Ereignissen auf Basis von empirischen Daten über zurückliegendes Verhalten (Bild 1.5). Mit den Erkenntnissen aus der deskriptiven und diagnostischen Analytik kann die prädiktive Analytik genutzt werden, um Cluster, Tendenzen oder vielleicht Ausnahmen zu erkennen, und damit eine bestimmte Vorhersage zu machen. Dabei wird versucht, die Frage zu beantworten: Was wird wahrscheinlich passieren?



**Bild 1.5** Prädiktive Analytik

Das Ziel von Prognosemodellen ist es, zukünftige Ereignisse frühzeitig und besser einschätzen zu können. Der Kern der prädiktiven Analytik ist der Prädiktor, das sind Variablen und/oder Kenn-

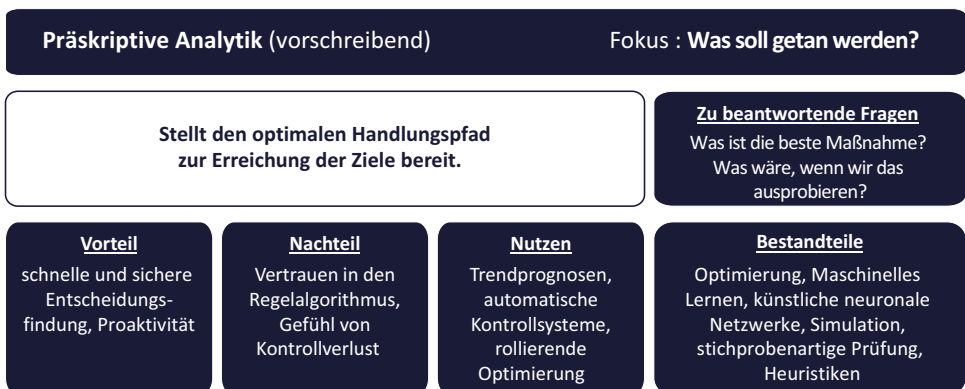
werte, die für eine Einheit oder eine Person gemessen werden, um zukünftiges Verhalten vorherzusagen. Wenn es sich bei der vorhergesagten Variablen um eine kategoriale Variable handelt, wird der Vorgang der Vorhersage als Klassifizierung bezeichnet, andernfalls als Regression. Wenn die vorhergesagte Variable zeitabhängig ist, wird der Vorhersageprozess oft als Zeitreihenprognose bezeichnet.

Insbesondere bei immer komplexeren Abläufen und Prozessen ist ein solches Wissen von entscheidendem Vorteil. Je differenzierter ein Prognosemodell ist, desto genauer wird es Vorhersagen treffen. Um verlässliche Vorhersagen zu erhalten, ist es nicht immer notwendig, möglichst viele Daten auszuwerten – mehr bedeutet nicht automatisch eine bessere Qualität der Vorhersagen. Prognosen und datengesteuerte Entscheidungen beinhalten mögliche zukünftige Entwicklungen (Was wird als nächstes passieren? Wie wirkt sich das auf meine Prozesse aus? Wann ist sehr wahrscheinlich womit zu rechnen?).

Die Mustererkennung analysiert die eingehenden Daten und versucht, Muster zu erkennen. Während die explorative Mustererkennung darauf abzielt, Datenmuster im Allgemeinen zu erkennen, beginnt die deskriptive Mustererkennung mit der Kategorisierung der erkannten Muster. Die Mustererkennung umfasst diese beiden Szenarien, je nach Anwendungsfall und Form der Daten werden unterschiedliche Mustererkennungsmethoden angewandt. Mustererkennung und Entscheidungsfähigkeit ist die Grundlage für Vorhersagen (Was, wenn diese Trends anhalten? Wie viel ist erforderlich? Wann ist es erforderlich?).

## 1.2.2 Präskriptive Analytik

Präskriptive Analytik ist die höchste Stufe der Analytik und ermittelt mögliche Lösungen für Probleme [1.8]. Hier wird mit Hilfe ausgefeilter mathematischer Modelle die beste von vielen Handlungsalternativen bestimmt, die in der Regel durch prädiktive und/oder deskriptive Analytik ermittelt werden (Bild 1.6). All die bisherigen Beobachtungen und Vorhersagen sind großartig, aber wichtig ist vor allem, wie wir daraus konkrete Maßnahmen ableiten können. Es bleibt nicht nur bei Vorhersagen. Wir würden vorzugsweise direkt auf diese Signale reagieren. Mit präskriptiver Analytik erstellen wir einen konkreten Ratschlag, was passieren soll, bis hin zur direkten Umsetzung.



**Bild 1.6** Präskriptive Analytik

Die präskriptive Analytik verwendet Optimierungs-, Simulations- und Heuristik-basierte Modellierungstechniken für Entscheidungen. Die Online-Optimierung dient der kontinuierlichen Überwachung des Prozesses und der Vorhersage eines optimalen Ablaufes durch Modellbezug.

Bei der präskriptiven Analytik gibt es wie in Bild 1.1. dargestellt zwei Ebenen menschlicher Intervention: Entscheidungsunterstützung, z. B. die Bereitstellung von Empfehlungen, und Entscheidungsautomatisierung, z. B. die Umsetzung der vorgeschriebenen Maßnahme. Die Ergebnisse von Optimierungssoftware können dem Bediener in Form von Entscheidungsunterstützungen präsentiert oder als neue Zielvorgabe für eine untergeordnete Regelstrategie automatisch festgelegt werden. Letzteres umfasst modellbasierte Software, die eine vollautomatisierte adaptive Rückkopplung mit intelligenten Systemen zur Regelung des Prozesses verwendet.

Die Optimierung wird zur Herausforderung, wenn sie mit Unsicherheit konfrontiert ist. Einige der Schlüsselfaktoren, die zur Unsicherheit beitragen, sind die Komplexität des Systems, die Länge des zukünftigen Zeithorizonts und das irrationale menschliche Verhalten. Deshalb sind Systeme für präskriptive Analytik nicht immer perfekt und müssen genau überwacht und ständig angepasst werden.

## Quellen

- [1.1] The Digitalisation of Science, Technology and Innovation: Key Developments and Policies. Paris: OECD Publishing, 2020.
- [1.2] LINDEN, Alexander; KART, Lisa; SCHULTE, W. Roy: Extend Your Portfolio of Analytics Capabilities. Gartner research note G00254653, 23. September 2013. <https://www.gartner.com/doc/2594822/extend-portfolio-analytics-capabilities> (abgerufen am 15.01.2022).
- [1.3] SAS Institute: Eight Level of Analytics, sascom magazine, Fourth quarter 2008.
- [1.4] BAUER, Per: IT-Serviceoptimierung nach dem Reifegradmodell. Otterfing: IT Verlag für Informationstechnik GmbH, 05. Oktober 2015, <https://www.it-daily.net/it-management/business-software/11532-it-serviceoptimierung-nach-dem-reifegradmodell> (abgerufen am 11.10.2021).
- [1.5] SCHULZ, Thomas: Datenanalyse: Schlüsseltechnologie für die digitale Transformation. In: etz Elektrotechnik + Automation, 142 (2021), Nr. S4, S. 34-39.
- [1.6] SCHLEGEL, Kurt; SALLAM, Rita; YUEN, Daniel; TAPADINHAS, Joao: Magic Quadrant for Business Intelligence and Analytics Platforms. Gartner research note G00239854, 5 Feb 2013. <https://www.gartner.com/en/documents/2326815/magic-quadrant-for-business-intelligence-and-analytics-p> (abgerufen am 15.01.2022).
- [1.7] DAVENPORT, Thomas. H.; HARRIS, Jeanne. G.: Competing on Analytics: The New Science of Winning. Boston, MA: Harvard Business Review, 2007.
- [1.8] LEPENIOTI, Katerina; BOUSDEKIS, Alexandros; APOSTOLU, Dimitris; MENTZAS, Gregoris: Prescriptive analytics: Literature review and research challenges. In: International Journal of Information Management, Ausgabe 50, Februar 2020, S. 57-70.
- [1.9] The Splunk Guide to Operational Intelligence. Turn Machine-Generated Data Into Real-Time Visibility, Insight and Intelligence. White Paper. San Francisco, CA: Splunk Inc. <https://www.splunk.com/pdfs/solution-guides/splunk-guide-to-operational-intelligence.pdf> (abgerufen am 11.10.2021).
- [1.10] GANDOM, Amir; HAIDER, Murtaza: Beyond the hype: Big data concepts, methods, and analytics. In: International Journal of Information Management, 35 (2015). Nr. 2, S. 137-144.
- [1.11] GROVER, Purvar; KAR, Kumar: Big data analytics: a review on theoretical contributions and tools used in literature. In: Global Journal of Flexible Systems Management, 18 (2017), Nr. 3, S. 203-229.
- [1.12] CHAE, Bongsug (Kevin); YANG, Chenlung; OLSON, David; SHEU, Chwen: The impact of advanced analytics and data accuracy on operational performance: A contingent resource based theory (RBT) perspective. In: Decision Support Systems, Ausgabe 59, März 2014, Seite 119-126.
- [1.13] REITH, János: Qualitätsüberwachung für die Feinbearbeitung von Bohrungen mit nachstellbaren Ausbohrwerkzeugen. Dissertation. Technische Universität Dresden, 1989.