

# Inhalt

<b>Vorwort</b> .....	<b>V</b>
<b>1 Einführung</b> .....	<b>1</b>
<b>2 Warum Data Mining? Wozu Big Data?</b> .....	<b>3</b>
2.1 Definition und Einordnung der Begriffe .....	6
2.1.1 Was ist Data Mining? .....	6
2.1.2 Was ist Big Data? .....	14
2.1.3 Data Mining im Kontext anderer Datenanalyseverfahren .....	15
2.2 Spezielle Anforderungen der Industrie an die Datenanalyse .....	22
2.3 Gibt es einen Handlungsbedarf für die Industrie? .....	29
<b>3 Das theoretische und mathematische Konzept der technischen Datenauswertung</b> .....	<b>33</b>
3.1 Einführung .....	33
3.2 Datenselektion und Datenzusammenführung .....	35
3.2.1 Aufbau einer Datentabelle .....	35
3.2.2 Denormalisierung von Datentabellen .....	36
3.2.3 Synchronisierung von Datentabellen .....	37
3.3 Datenvorverarbeitung .....	39
3.3.1 Festlegung der Datentypen .....	39
3.3.2 Diskretisierung von metrischen Daten .....	41
3.3.3 Statistiken und Tests für metrische Daten .....	43
3.3.4 Das Problem ungenauer Messungen .....	48
3.3.5 Behandlung von Datenlücken .....	51
3.3.6 Behandlung von Ausreißern .....	53
3.3.7 Behandlung von Mehrdeutigkeiten .....	55
3.4 Datentransformation .....	60

3.5	Datenanalyse .....	64
3.5.1	Visuelle explorative Analysen .....	64
3.5.2	Überblick über multivariate Verfahren zur Datenanalyse .....	67
3.5.2.1	Regressionsanalysen .....	67
3.5.2.2	Varianzanalyse .....	73
3.5.2.3	Diskriminanzanalyse .....	76
3.5.2.4	Korrelationsanalyse .....	79
3.5.2.5	Faktoranalyse .....	83
3.5.2.6	Clusteranalyse .....	86
3.5.3	Einführung in Data-Mining-Methoden .....	93
3.5.4	Data Mining zum Auffinden von Zusammenhängen ....	97
3.5.4.1	Neuronale Netze .....	99
3.5.4.2	Support-Vektor-Maschinen .....	114
3.5.4.3	Gütemaße für Modelle und Klassifikatoren ...	119
3.5.5	Data Mining zum Auffinden von Strukturen .....	127
3.5.5.1	Fuzzy-Clusterverfahren .....	128
3.5.5.2	Demographisches Clustern .....	130
3.5.5.3	Selbstorganisierende Merkmalskarten .....	131
3.5.5.4	Gütemaße für Clusterverfahren .....	143
3.5.6	Data Mining zum Generieren von Regeln .....	145
3.5.6.1	Bayessche Netze .....	146
3.5.6.2	Entscheidungsbäume .....	152
3.5.6.3	Assoziationsregeln .....	162
3.5.6.4	Gütemaße für Regeln .....	165
3.5.7	Data Mining zum Visualisieren hochdimensionaler Datenräume .....	166
3.5.7.1	Selbstorganisierende Merkmalskarten für topologieerhaltende Projektionen .....	166
3.5.7.2	Gütemaße für Projektionen .....	173
3.5.8	Zusammenfassung der Data-Mining-Verfahren .....	177
3.6	Interpretation der Ergebnisse .....	180
3.6.1	Fehlinterpretationen .....	181
3.6.2	Strittige Interpretationen .....	187
3.6.3	Konsequenzen .....	189
<b>4</b>	<b>Hilfreiche Auswertemöglichkeiten für praktische Anwendungsfälle .....</b>	<b>191</b>
4.1	Text Mining – das Auswerten unstrukturierter Daten .....	191
4.2	Versuchsplanungen zur Erzeugung von Prozessdaten .....	197
4.3	Automatische Diskretisierungen .....	202

4.4	Güte und Sicherheit von Regressionsschätzungen .....	204
4.5	Auffinden der sensitiven Einflussgrößen .....	208
4.6	Ausschluss von zufälligen Zusammenhängen .....	212
4.7	Datenbasierte Optimierungen .....	216
<b>5</b>	<b>Big Data – die Datenhaltungs- und Verarbeitungskonzepte der Gegenwart .....</b>	<b>229</b>
5.1	Digitale Transformation und Big Data .....	230
5.2	Grundprinzipien eines Paradigmenwandels .....	232
5.2.1	Die drei Vs – und der Wert .....	232
5.2.2	Scale-up und Scale-out .....	232
5.2.3	Unabhängige Verarbeitung direkt auf den Daten .....	233
5.2.4	Schema on Read versus Schema on Write .....	234
5.2.5	Hardwarevirtualisierung und Containermanagement ..	234
5.2.6	Datenvirtualisierung .....	235
5.2.7	Entkoppelte Systeme .....	236
<b>6</b>	<b>Technische Big-Data-Lösungen zur industriellen und kommerziellen Datenanalyse .....</b>	<b>237</b>
6.1	Datenmanagement im Big-Data-Umfeld .....	237
6.1.1	Hadoop machte den Anfang .....	237
6.1.2	Apache Spark – die nächste Evolutionsstufe .....	240
6.1.3	Abstrahierte Datenverarbeitung und -speicherung .....	241
6.1.4	Komplexe Eventverarbeitung mit Kafka & Co. ....	245
6.1.5	Das beste beider Welten – von Lambda und Kappa .....	246
6.1.6	Big-Data-Plattformen .....	247
6.1.7	NoSQL-Datenbanken .....	248
6.1.8	Anwendungsfälle für NoSQL-Datenbanken .....	249
6.1.9	Technologiestacks .....	250
6.2	Datenzentrische Architekturen .....	251
6.2.1	AI-basierte Systeme brauchen IA-basierte Plattformen ..	251
6.2.2	Die logische Architektur .....	252
6.2.3	Die Softwarearchitektur .....	252
6.2.4	Die technische Architektur .....	252
6.3	Der Supervised Data Lake (SDL) .....	253
6.3.1	Ein Data Lake braucht ein Konzept, damit der See nicht zum Sumpf wird .....	253
6.3.2	Die unterschiedlichen Bereiche eines SDL .....	255
6.3.3	Quellen und Ladearten .....	255

6.3.4	Raw Zone	256
6.3.5	Ingestion Zone	256
6.3.6	Discovery und Sandbox	256
6.3.7	Integration	257
6.3.8	Serving	258
6.3.9	Associated Processes	258
6.3.10	Access und Application	259
6.4	Aufbau eines Data Lakes	259
6.4.1	Think Big – Start Small – Act Now	259
6.4.2	Vision, Ziele und Standortbestimmung	260
6.4.3	Konzeption des Data Lakes	260
6.4.4	Implementierung der Basisumgebung	261
6.4.5	Data Lake Ramp-up – Use Case Driven	261
6.4.6	Industrialisierung – die betriebsfokussierte Datenfabrik	262
6.5	Cloud-Computing und Services	263
6.5.1	Die Cloud-Ausbaustufen – Everything as a Service	264
6.5.2	Offene Ökosysteme	265
6.5.3	Der Data Lake in der Cloud	266
6.6	Big Data, Data Mining und Artificial Intelligence	268
6.6.1	Analytic Data Hub	269
6.6.2	Data-Science- und Data-Mining-Plattformen	270

## **7 Die Anwendersicht – Systematik für industrielle**

<b>Anwendungen</b>	<b>279</b>	
7.1	Aufgabenstellung und Zielsetzung	279
7.1.1	Datengetriebene Identifikation von Aufgabenstellungen	279
7.1.2	„Produktgetriebene“ Identifikation	280
7.1.3	Geschäftsorientierte Identifikation von Aufgabenstellungen	280
7.1.3.1	Reduktion von Kosten, Verlusten, Verschwendungen	283
7.1.3.2	Erhöhung operativer Performance	284
7.1.3.3	Ergebnisverbesserung funktionaler Prozesse	285
7.2	Vorgehensmethodik	286
7.2.1	Workshop zur Ideenfindung und Datenanalyse	289
7.2.1.1	Design-Thinking-Workshop	289
7.2.1.2	Wertschöpfungsschritte	290
7.2.1.3	Perspektiven	291
7.2.1.4	Schmerzpunkte und Mehrwerte	292
7.2.1.5	Erzeugen des Mehrwertes	292

7.2.1.6	Geschäftsmodell .....	294
7.2.1.7	Anwendungen und Lösungsansätze identifizieren .....	296
7.2.2	Hackathons als alternative Möglichkeit der Lösungsfindung und Pilotierung .....	297
7.2.3	Aufsetzen konkreter Aufgabenstellungen .....	299
7.2.3.1	Definition der Aufgabenstellung .....	299
7.2.3.2	Modellauswahl .....	300
7.2.3.3	Beauftragung von Dienstleistern .....	301
7.2.4	Explorations- und Umsetzungsphase eines Use Case ...	302
7.2.4.1	Sichtung der Daten .....	302
7.2.4.2	Bestimmung der sensitiven Eingangsgrößen ..	308
7.2.4.3	Modellierung und Ergebnisbewertung .....	315
7.2.4.4	Die Königsklasse: Vektorielle Optimierung eines Use Case .....	316
7.2.5	Auswertung und Detailkonzept, Applikationserstellung und Implementierung .....	321

<b>8</b>	<b>Die Anwendersicht – typische Anwendungsfelder am konkreten Beispiel .....</b>	<b>327</b>
8.1	Anwendungen in den Geschäftsfunktionen .....	330
8.1.1	Forschung und Entwicklung .....	330
8.1.2	Engineering .....	333
8.1.3	Produktmanagement .....	334
8.1.4	Einkauf, Supply Chain Management, Logistik .....	336
8.1.5	Fertigung und Produktion .....	338
8.1.6	Qualitätsmanagement .....	340
8.1.7	Service und Instandhaltung .....	342
8.1.8	Service und After Market .....	344
8.1.9	Marketing und Vertrieb .....	347
8.2	Ausgewählte Data-Mining- und Big-Data-Beispiele .....	350
8.2.1	Forschung, Entwicklung und Engineering .....	351
8.2.1.1	Beschleunigung einer Produktentwicklung ...	351
8.2.2	Einkauf .....	358
8.2.2.1	Spend Cube .....	360
8.2.2.2	Bündelung .....	363
8.2.2.3	Spezifikations- und Kostenhebel .....	366
8.2.3	Produktion, Fertigung und Service .....	370
8.2.3.1	Störungsanalysen .....	370
8.2.3.2	Instabilitätsanalysen in einem Klärwerk .....	372
8.2.3.3	Fehlerdetektion in einem Kraftwerk .....	381

8.2.3.4	Analyse der Dynamik von chemischen Batchprozessen .....	390
8.2.4	Instandhaltung und Service .....	394
8.2.4.1	Aufbau einer Datenbasis für erweiterte Analysen und Monitoring von Industrieanlagen .....	394
8.2.4.2	Erweiterung eines digitalen Zwillings um Maschinendaten und Strompreisdaten im Bereich Windenergie .....	396
8.2.5	Marketing und Vertrieb .....	398
8.2.5.1	Cross-Selling-Effekte mit Data Mining finden ..	398
8.2.5.2	Cross-Selling-Analysen mit Big-Data- Technologien beschleunigen .....	405
8.2.5.3	Optimale Preisschwellen mit Data Mining aufspüren .....	407
8.2.6	Data Mining für die strategische Unternehmensführung	412
<b>9</b>	<b>Small Data gehört die Zukunft .....</b>	<b>421</b>
9.1	Einführung in die Thematik .....	421
9.2	Charakteristik von Small Data .....	423
9.3	Machine Learning versus menschlicher Geist – die Mind-Data- Hypothese .....	428
9.4	Bewusstsein als übergeordnete Ordnungsstruktur neuronaler Systeme .....	431
9.5	Mind-Data-Auswertungen mit maschinellem Bewusstsein .....	442
<b>10</b>	<b>Ausblick und mögliche Weiterentwicklungen von Data Mining und Big Data .....</b>	<b>451</b>
<b>11</b>	<b>Liste der häufig verwendeten Formelzeichen und Symbole ..</b>	<b>457</b>
<b>12</b>	<b>Literaturverzeichnis .....</b>	<b>461</b>
<b>13</b>	<b>Autoren .....</b>	<b>471</b>
<b>Index</b>	<b>.....</b>	<b>473</b>