

# Inhaltsverzeichnis

Vorwort .....	17
Einleitung .....	19
Business Intelligence und SQL Server .....	20
Data Mining .....	20
Ziel des Buchs und Voraussetzungen .....	21
Aufbau des Buchs .....	21
Kontakt zu den Autoren .....	22
Teil A – Data Mining-Konzepte .....	23
1 Einführung in Data Mining .....	25
Überblick über Data Mining .....	26
Beispiel 1: Vorhersage von Flugstornierungen .....	28
Beschreibung der Aufgabenstellung .....	28
Lösungsweg .....	28
Beispiel 2: Warenkorbanalyse .....	30
Beschreibung der Aufgabenstellung .....	30
Lösungsweg .....	31
Beispiel 3: Marktsegmentierung .....	34
Beschreibung der Aufgabenstellung .....	34
Lösungsweg .....	35
Zusammenfassung .....	36
2 Data Mining-Algorithmen .....	37
Daten- und Inhaltstypen .....	38
Der Inhaltstyp <i>Diskret</i> (Discrete) .....	38
Der Inhaltstyp <i>Fortlaufend</i> (Continuous) .....	39
Der Inhaltstyp <i>Diskretisiert</i> (Discretized) .....	39
Der Inhaltstyp <i>Schlüssel</i> (Key) .....	39
Der Inhaltstyp <i>Schlüssel-Sequenz</i> (Key Sequence) .....	39
Der Inhaltstyp <i>Zeit-Schlüssel</i> (Key Time) .....	39
Der Inhaltstyp <i>Geordnet</i> (Ordered) .....	39
Der Inhaltstyp <i>Zyklisch</i> (Cyclical) .....	40
Daten- und Inhaltstypen kombinieren .....	40

Algorithmengruppen .....	40
Klassifikationsalgorithmen .....	40
Regressionsalgorithmen .....	42
Zuordnungsalgorithmen .....	44
Segmentierungsalgorithmen .....	44
Sequenzanalysealgorithmen .....	44
Plug-In-Algorithmen .....	45
Zusammenfassung .....	45
<b>3 Überlegungen zur Datenvorbereitung .....</b>	<b>47</b>
Handhabung von Einträgen .....	48
Zusammenfassen von Einträgen .....	48
Auswählen von Einträgen .....	48
Handhabung von Attributen .....	49
Auswählen und Erzeugen von Attributen .....	49
Diskretisieren von Attributen .....	49
Fehlende und ungenaue Werte .....	50
Umgang mit fehlenden Werten .....	50
Umgang mit ungenauen Werten .....	50
Strukturelle Anpassungen .....	50
Zusammenfassung .....	51
<b>4 Überlegungen zur Validierung von Miningmodellen .....</b>	<b>53</b>
Validierungstechniken .....	54
Naiver Ansatz .....	54
HoldOut .....	55
Kreuzvalidierung .....	55
Leave One Out-Validierung .....	56
Darstellung von Validierungsergebnissen .....	56
Klassifikationsmatrix .....	56
Lift- und Cost-Charts .....	56
Zusammenfassung .....	58
<b>Teil B – Die Data Mining-Plattform von Microsoft .....</b>	<b>59</b>
<b>5 Anwendungen und Technologien im Überblick .....</b>	<b>61</b>
SQL Server 2005/2008 .....	62
Relationales Datenbankmodul .....	62
SQL Server Integration Services .....	63
SQL Server Analysis Services (OLAP) .....	63
SQL Server Analysis Services (Data Mining) .....	63
SQL Server Reporting Services .....	63

Verwaltungs- und Entwicklungstools .....	64
Beispieldatenbank Adventure Works .....	64
Skript- und Abfragesprachen .....	65
ASSL und XMLA .....	65
DMX .....	65
Office 2007 .....	66
SQL Server Data Mining-Add-Ins für Office 2007 .....	66
Excel 2007 .....	66
Visio 2007 .....	67
Visual Studio 2005/2008 .....	67
Objektbibliotheken und Steuerelemente .....	67
AMO .....	67
ADOMD.NET .....	68
Data Mining Viewer Controls .....	68
Zusammenfassung .....	68
 6 Unterschiede in Versionen und Editionen von SQL Server .....	 69
Übersicht der Unterschiede .....	70
Versionen von SQL Server .....	70
SQL Server 2005 .....	71
SQL Server 2008 .....	71
Editionen von SQL Server .....	73
SQL Server 2005 .....	73
SQL Server 2008 .....	74
Zusammenfassung .....	75
 7 Einrichten der Arbeitsumgebung .....	 77
Bestehende Arbeitsumgebung .....	78
Erforderliche Office-Komponenten .....	78
Visual Studio 2008 Professional Edition .....	79
SQL Server 2008 Developer Edition .....	79
Die Beispieldatenbank Adventure Works .....	84
Beispielprojekte für SQL Server .....	88
SQL Server Data Mining-Add-Ins für Office 2007 .....	89
Zusammenfassung .....	90
 Teil C – Data Mining mit Office 2007 .....	 91
 8 Einführung in die Data Mining-Add-Ins .....	 93
Technische Voraussetzungen .....	94
Verbindung zu Analysis Services herstellen .....	94
Beschreibung und Funktionen der Add-Ins .....	96
Tabellenanalysetools für Excel 2007 .....	96

Data Mining-Client für Excel 2007 .....	98
Excel-Funktionen für Data Mining .....	102
Data Mining-Vorlagen für Visio 2007 .....	102
Beispieldaten für Excel .....	102
Hilferessourcen .....	103
Eigene Daten in Excel verwenden .....	103
Externe Daten importieren .....	103
Daten als Excel-Tabelle formatieren .....	104
Zusammenfassung .....	104
<b>9 Tabellenanalysetools für Excel 2007 .....</b>	<b>105</b>
Wichtige Einflussfaktoren analysieren .....	106
Das Tool <i>Wichtige Einflussfaktoren analysieren</i> verwenden .....	106
Informationen zum Bericht der wichtigen Einflussfaktoren .....	108
Kategorien erkennen .....	109
Das Tool <i>Kategorien erkennen</i> verwenden .....	109
Informationen zum Kategoriebericht .....	110
Vorhersagen treffen .....	111
Fehlende Werte vorhersagen .....	111
Zeitreihen vorhersagen .....	112
Muster mit dem Vorhersagerechner erkennen .....	114
Ausnahmen hervorheben .....	116
Szenarienanalysen .....	117
Zielsuche durchführen .....	118
Was-wäre-wenn-Analyse durchführen .....	119
Warenkörbe analysieren .....	121
Das Tool <i>Warenkorbanalyse</i> verwenden .....	121
Informationen zu den Berichten der Warenkorbanalyse .....	123
Zusammenfassung .....	124
<b>10 Data Mining-Client für Excel 2007 .....</b>	<b>125</b>
Daten zur Auswertung vorbereiten .....	126
Daten durchsuchen und gruppieren .....	126
Ausreißer entfernen .....	128
Stichproben entnehmen .....	129
Miningmodelle erstellen .....	130
Miningmodelle mithilfe von Assistenten erstellen .....	131
Miningmodelle manuell erstellen .....	134
Externe Daten verwenden .....	141
Leistung von Miningmodellen bewerten .....	144
Genauigkeitsdiagramme zur Leistungsbewertung .....	144
Weitere Methoden zur Leistungsbewertung .....	146
Vorhersagen treffen .....	146
Vorhersagen mit dem Assistenten erstellen .....	147

Vorhersagen als DMX-Abfrage erstellen .....	149
Weitere Funktionen im Data Mining-Client .....	151
Miningmodelle durchsuchen und dokumentieren .....	151
Miningstrukturen und -modelle verwalten .....	152
Interaktion von Excel mit SQL Server überwachen .....	153
Zusammenfassung .....	153
11 Excel-Funktionen für Data Mining .....	155
Erstellen des Beispielminingmodells .....	156
Verwenden der Excel-Funktionen für Data Mining .....	156
Vorhersagen mit <i>DMPREDICT</i> .....	157
Vorhersagen mit Datenzeilen und <i>DMPREDICTTABLEROW</i> .....	158
Modellinhalt abfragen mit <i>DMCONTENTQUERY</i> .....	158
Zusammenfassung .....	160
12 Data Mining-Vorlagen für Visio 2007 .....	161
Erstellen der Beispielminingmodelle .....	162
Erstellen eines Klassifikationsmodells .....	162
Erstellen eines Clustermodells .....	163
Diagramme für Miningmodelle erstellen .....	163
Entscheidungsstrukturdiagramm erstellen .....	164
Weitere Diagrammtypen .....	166
Zusammenfassung .....	170
Teil D – Data Mining mit SQL Server 2008 .....	171
13 Miningmodelle im BIDS erstellen .....	173
Überblick über Strukturen und Modelle .....	174
Neues Projekt anlegen .....	174
Datenquelle und Datenquellensicht anlegen .....	175
Miningstruktur und -modell anlegen .....	176
Miningstruktur bereitstellen und verarbeiten .....	182
Miningmodell anzeigen .....	183
<i>Entscheidungsstruktur</i> anzeigen .....	183
Zusammenfassung .....	190
14 Miningmodelle im BIDS abfragen .....	191
Vorhersagen mit SINGLETON-Abfragen .....	192
Vorhersagen mit SQL-Abfragen .....	195
Zusammenfassung .....	197

15	Validieren von Miningmodellen .....	199
	Validierung nach der HoldOut-Methode .....	200
	Testmenge zurückhalten .....	200
	Rahmenbedingungen für Validierung angeben .....	201
	Prognosegütediagramm anzeigen .....	202
	Gewinndiagramm anzeigen .....	203
	Klassifikationsmatrix anzeigen .....	205
	Unterschiedliche Modelle vergleichen .....	206
	Durchführen einer Kreuzvalidierung .....	210
	Rahmenbedingungen für Kreuzvalidierung festlegen .....	210
	Ergebnisse einer Kreuzvalidierung anzeigen .....	212
	Zusammenfassung .....	213
16	Data Mining und SQL Server Integration Services .....	215
	Überblick über SQL Server Integration Services .....	216
	Ablaufsteuerung .....	216
	Datenfluss .....	223
	Ausgewählte Tasks für Data Mining-Szenarien .....	234
	Tasks der Ablaufsteuerung .....	234
	Tasks des Datenflusses .....	237
	Ausgewählte Tasks für Text Mining-Szenarien .....	238
	Ausdrucksextrahierungs-Task .....	238
	Ausdruckssuche-Task .....	239
	Datenstruktur für Text Mining-Szenarien .....	242
	Zusammenfassung .....	243
17	Data Mining und SQL Server Reporting Services .....	245
	Entwerfen eines Berichts .....	246
	Bericht bearbeiten .....	253
	Berichtsdaten definieren .....	253
	Berichtsparemeter definieren .....	254
	Datenquellen definieren .....	254
	Datasets definieren .....	255
	Bericht entwerfen .....	258
	Bereitstellen und Abrufen eines Berichts .....	260
	Zusammenfassung .....	261
	Teil E – Fallbeispiele .....	263
18	Warenkorbanalysen mit Zuordnungsregeln durchführen .....	265
	Geschachtelte Tabellen verstehen .....	266
	Miningmodell mit geschachtelten Tabellen erstellen .....	266
	Miningmodell anzeigen .....	272

	Ermittelte Regeln anzeigen .....	272
	Ermittelte Itemsets anzeigen .....	274
	Abhängigkeitsnetzwerk anzeigen .....	275
	Miningmodell abfragen .....	276
	Zusammenfassung .....	278
19	Kundensegmente mit Clustering erkennen .....	279
	Miningmodell erstellen .....	280
	Miningmodell anzeigen .....	287
	Clusterdiagramme anzeigen .....	288
	Clusterprofile anzeigen .....	289
	Clustermerkmale anzeigen .....	290
	Cluster vergleichen .....	291
	Miningmodell abfragen .....	291
	Zusammenfassung .....	293
20	Benutzerverhalten auf Websites mit Sequence Clustering analysieren .....	295
	Miningmodell erstellen .....	296
	Miningmodell anzeigen .....	302
	Clusterdiagramm anzeigen .....	302
	Clusterprofile anzeigen .....	303
	Clustermerkmale anzeigen .....	304
	Cluster vergleichen .....	305
	Miningmodell abfragen .....	306
	Zusammenfassung .....	309
21	Kennzahlen mit dem Time Series-Algorithmus vorhersagen .....	311
	Miningmodell erstellen .....	312
	Miningmodell anzeigen .....	318
	Zeitreihe als Diagramm anzeigen .....	318
	Modell anzeigen .....	319
	Miningmodell abfragen .....	319
	Zusammenfassung .....	321
22	Streuverluste mit neuronalen Netzwerken vermindern .....	323
	Miningmodell erstellen .....	324
	Miningmodell anzeigen .....	328
	Optionen für Anzeige festlegen .....	328
	Anzeige interpretieren .....	328
	Miningmodell abfragen .....	329
	Zusammenfassung .....	331

Teil F – Fortgeschrittene Techniken und Entwicklung .....	333
23 Einführung in ASSL und XMLA .....	335
Anwendungsgebiet von XMLA .....	336
Methoden und Erweiterungen von XMLA .....	337
Discover-Methode .....	337
Execute-Methode .....	337
XMLA-Erweiterungen (ASSL) .....	337
Einführungsbeispiel: Miningmodell verarbeiten .....	338
XMLA-Abfragen im Management Studio erstellen .....	338
XMLA-Vorlagen verwenden .....	339
Beispiel: Datenbank sichern .....	340
XMLA mit SQL Server Profiler weiterverwenden .....	341
DMX mit XMLA versenden .....	343
Zusammenfassung .....	345
24 Data Mining-Erweiterungen (DMX) .....	347
Sprachelemente von DMX .....	348
Operatoren .....	348
Datendefinitionsanweisungen .....	349
Datenbearbeitungsanweisungen .....	350
SELECT-Anweisung .....	350
DMX-Funktionen .....	351
Kommentare .....	353
DMX-Abfragen in Management Studio erstellen .....	354
Editor für DMX-Abfragen .....	354
DMX-Vorlagen verwenden .....	355
Generator für Vorhersageabfragen verwenden .....	356
Miningstrukturen und -modelle erstellen .....	357
Miningstrukturen erstellen .....	357
Miningmodelle erstellen .....	357
Miningmodelle trainieren .....	358
Miningstrukturen mit geschachtelten Tabellen erstellen .....	358
Miningmodelle für geschachtelte Tabellen erstellen .....	359
Miningmodelle für geschachtelte Tabellen trainieren .....	359
Vorhersageabfragen erstellen .....	360
Einfache Vorhersageabfragen erstellen .....	360
Batchvorhersageabfragen erstellen .....	360
Vorhersageabfragen mit geschachtelten Tabellen erstellen .....	361
Weitere Abfragen für Miningmodelle .....	362
Statistische Daten abfragen .....	362
Werte einer Spalte abfragen .....	363

Miningmodellinhalt abfragen .....	363
Miningstrukturen und -modelle löschen .....	365
Zusammenfassung .....	365
25 Analysis Management Objects (AMO) .....	367
Konzepte und Objektmodell von AMO .....	368
Grundlegende Klassen in AMO .....	369
Klassen für Data Mining in AMO .....	370
Weitere Klassen in AMO .....	370
Erstellen der Beispielanwendung .....	370
Oberfläche der Anwendung erstellen .....	371
Verweis auf die Klassenbibliothek von AMO hinzufügen .....	372
Datenbankverbindung herstellen .....	372
Mit Miningstrukturen und -modellen arbeiten .....	373
Miningstrukturen und -modelle auslesen .....	373
Miningstrukturen und -modelle erstellen .....	375
Miningstrukturen und -modelle verwalten .....	378
Datenbanken sichern und wiederherstellen .....	379
Datenbanken sichern .....	379
Datenbanken wiederherstellen .....	380
AMO verteilen .....	380
Zusammenfassung .....	380
26 ActiveX Data Objects MultiDimensional .NET (ADOMD.NET) .....	381
Konzepte und Objektmodell von ADOMD.NET .....	382
Klassen für den Datenzugriff in ADOMD.NET .....	382
Klassen für Data Mining in ADOMD.NET .....	383
Erstellen der Beispielanwendung .....	383
Oberfläche der Anwendung erstellen .....	384
Verweis auf die Klassenbibliothek von ADOMD.NET hinzufügen .....	385
Datenbankverbindung herstellen .....	385
Datenabfragen ausführen .....	385
Metadaten von Miningmodellen abfragen .....	386
Spaltenwerte von Miningmodellen abfragen .....	388
Vorhersagen mit Miningmodellen treffen .....	389
Verwaltungsaufgaben mit ADOMD.NET ausführen .....	390
ADOMD.NET verteilen .....	390
Zusammenfassung .....	390
27 Data Mining Viewer Controls .....	391
Überblick der Data Mining Viewer Controls .....	392
Erstellen der Beispielanwendung .....	393
Oberfläche der Anwendung erstellen .....	393

Verweise auf benötigte DLLs hinzufügen .....	394
Programmcode hinzufügen .....	394
Data Mining Viewer Controls verteilen .....	397
Zusammenfassung .....	397
<b>Anhang A – Algorithmus-Parameter .....</b>	<b>399</b>
Microsoft Association Rules-Algorithmus .....	400
MAXIMUM_ITEMSET_COUNT .....	400
MAXIMUM_ITEMSET_SIZE .....	400
MAXIMUM_SUPPORT .....	400
MINIMUM_ITEMSET_SIZE .....	401
MINIMUM_PROBABILITY .....	401
MINIMUM_SUPPORT .....	401
OPTIMIZED_PREDICTION_COUNT .....	401
Microsoft Clustering-Algorithmus .....	402
CLUSTERING_METHOD .....	402
CLUSTER_COUNT .....	402
CLUSTER_SEED .....	402
MINIMUM_SUPPORT .....	403
MODELLING_CARDINALITY .....	403
STOPPING_TOLERANCE .....	403
SAMPLE_SIZE .....	403
MAXIMUM_INPUT_ATTRIBUTES .....	403
MAXIMUM_STATES .....	404
Microsoft Decision Trees-Algorithmus .....	404
COMPLEXITY_PENALTY .....	404
FORCE_REGRESSOR .....	404
MAXIMUM_INPUT_ATTRIBUTES .....	404
MAXIMUM_OUTPUT_ATTRIBUTES .....	405
MINIMUM_SUPPORT .....	405
SCORE_METHOD .....	405
SPLIT_METHOD .....	405
Microsoft Linear Regression-Algorithmus .....	406
MAXIMUM_INPUT_ATTRIBUTES .....	406
MAXIMUM_OUTPUT_ATTRIBUTES .....	406
FORCED_REGRESSOR .....	406
Microsoft Logistic Regression-Algorithmus .....	406
HOLDOUT_PERCENTAGE .....	406
HOLDOUT_SEED .....	406
MAXIMUM_INPUT_ATTRIBUTES .....	407
MAXIMUM_OUTPUT_ATTRIBUTES .....	407
MAXIMUM_STATES .....	407
SAMPLE_SIZE .....	407

Microsoft Naive Bayes-Algorithmus .....	407
MAXIMUM_INPUT_ATTRIBUTES .....	407
MAXIMUM_OUTPUT_ATTRIBUTES .....	408
MINIMUM_DEPENDENCY_PROBABILITY .....	408
MAXIMUM_STATES .....	408
Microsoft Neural Network-Algorithmus .....	408
HIDDEN_NODE_RATIO .....	408
HOLDOUT_PERCENTAGE .....	409
HOLDOUT_SEED .....	409
MAXIMUM_INPUT_ATTRIBUTES .....	409
MAXIMUM_OUTPUT_ATTRIBUTES .....	409
MAXIMUM_STATES .....	409
SAMPLE_SIZE .....	409
Microsoft Sequence Clustering-Algorithmus .....	410
CLUSTER_COUNT .....	410
MINIMUM_SUPPORT .....	410
MAXIMUM_SEQUENCE_STATES .....	410
MAXIMUM_STATES .....	410
Microsoft Time Series-Algorithmus .....	410
AUTO_DETECT_PERIODICITY .....	411
COMPLEXITY_PENALTY .....	411
FORECAST_METHOD .....	411
HISTORIC_MODEL_COUNT .....	411
HISTORICAL_MODEL_GAP .....	411
INSTABILITY_SENSITIVITY .....	411
MAXIMUM_SERIES_VALUE .....	412
MINIMUM_SERIES_VALUE .....	412
MINIMUM_SUPPORT .....	412
MISSING_VALUE_SUBSTITUTION .....	412
PERIODICITY_HINT .....	413
PREDICTION_SMOOTHING .....	413
 Stichwortverzeichnis .....	 415
 Über die Autoren .....	 427