

Inhalt

Einleitung	15
------------------	----

1 Einführung in Predictive Analytics 21

1.1 Grundlagen des Data Minings	21
1.1.1 Der Begriff »Data Mining« und seine Historie	21
1.1.2 Typische Anwendungsfälle von Data Mining	23
1.2 Der Data-Mining-Prozess	25
1.2.1 Problemdefinition	27
1.2.2 Datenanalyse	28
1.2.3 Datenvorbereitung	28
1.2.4 Modellierung	29
1.2.5 Ergebnisvisualisierung und -bewertung	30
1.2.6 Deployment des Modells	32
1.3 Methoden der Datenvorbereitung	32
1.3.1 Variablen	32
1.3.2 Deskriptive Statistik	33
1.3.3 Integration und Transformation	36
1.3.4 Sampling	37
1.3.5 Ausreißerbehandlung	38
1.3.6 Binning	39
1.3.7 Missing Values	41
1.4 Algorithmen und Methoden des Data Minings	42
1.4.1 Regressionsanalyse	43
1.4.2 Zeitreihenanalyse	45
1.4.3 Klassifikations- und Clusterverfahren	47
1.4.4 Assoziationsanalyse	50
1.4.5 Entscheidungsstrukturen	51
1.4.6 Künstliche neuronale Netze	53
1.4.7 Weitere Algorithmen	54
1.5 Einordnung von Predictive Analytics in den Bereich Data Mining	56
1.5.1 Was ist Predictive Analytics?	57
1.5.2 Abgrenzung von Predictive Analytics	57

2 Mehrwert durch Predictive Analytics 61

2.1	Warum Predictive Analytics?	61
2.2	Warum Predictive Analytics mit SAP?	68
2.3	Anwendungsbeispiele	70
2.3.1	Anwendungsbeispiel 1: Preiselastizitäten im Einzelhandel	71
2.3.2	Anwendungsbeispiel 2: Absatzprognosen in der Musikindustrie	73

3 SAP Predictive Analytics 77

3.1	Einführung und Einordnung in das SAP-Produktportfolio	77
3.1.1	Einführung in SAP Predictive Analytics	77
3.1.2	Zielgruppen	79
3.1.3	Produktportfolio	81
3.2	Plattformintegration und Architektur	84
3.2.1	Plattformintegration	84
3.2.2	Architektur	86
3.3	Installation	88
3.3.1	SAP Download Center	88
3.3.2	Desktop-Installation	90
3.3.3	Serverinstallation	94
3.3.4	Predictive-Factory-Installation	98
3.3.5	Hardwarevoraussetzungen	101
3.4	Benutzeroberfläche und Navigation	102
3.5	Einstellungen	104

4 Mit dem Modus »Automated Analytics« arbeiten 113

4.1	Grundlagen	113
4.1.1	Unterteilungsstrategie	113
4.1.2	Modellerstellung und Evaluierung	115
4.1.3	SRM-Grundlagen	117

4.2	Datencodierung	118
4.2.1	Codierung nominaler Variablen	118
4.2.2	Codierung ordinaler Variablen	120
4.2.3	Codierung stetiger Variablen	122
4.2.4	Manuelle Variablencodierung	124
4.2.5	Behandlung fehlender Werte in Automated Analytics	124
4.3	Datenzugriff und -vorbereitung mit dem Data Manager	124
4.3.1	Datenvorbereitung mit dem Data Manager	125
4.3.2	Datenmanipulation mit dem Data Manager	126
4.4	Klassifikations-/Regressionsanalyse	127
4.4.1	Datenquelle auswählen	128
4.4.2	Datenbeschreibung	133
4.4.3	Filter	138
4.4.4	Auswahl von Variablen	139
4.4.5	Zusammenfassung der Modellierungsparameter	142
4.4.6	Erweiterte Modellparameter	143
4.4.7	Trainieren des Modells und Modellübersicht	148
4.4.8	Modellverwendung: Anzeigen	151
4.4.9	Modellverwendung: Ausführen	174
4.4.10	Modellverwendung: Sichern/exportieren	187
4.5	Zeitreihen	191
4.6	Clustering	196
4.7	Analyse sozialer Netzwerke	200
4.8	Weitere Algorithmen und Toolkit	207
4.8.1	Kollokationsanalyse	207
4.8.2	Analyse der häufigen Pfade	210
4.8.3	Recommendation	212
4.8.4	Assoziationsanalyse	214
4.8.5	Sequenzanalyse	217
4.8.6	Textanalyse	218
4.8.7	Toolkit	223

5 Mit dem Modus »Expert Analytics« arbeiten 225

5.1	Funktionen von Expert Analytics	225
5.2	Navigation und Einstellungen in Expert Analytics	226

5.2.1	Einstellungen und Navigation im Tool	226
5.2.2	Ansichten in Expert Analytics	230
5.3	Datenvorbereitung	234
5.3.1	Vorbereitungssicht	234
5.3.2	Data Type Definition	239
5.3.3	Filter	239
5.3.4	Formula	241
5.3.5	Normalization	242
5.3.6	Partition	243
5.3.7	Sample	243
5.3.8	Model Statistics	244
5.3.9	Model Compare	246
5.3.10	SAP-HANA-Datenvorbereitungskomponenten	248
5.4	Assoziationsanalyse	251
5.4.1	Grundlegende Begriffe	251
5.4.2	R-Apriori	252
5.4.3	Anwendungsbeispiele	257
5.5	Clustering und Klassifikation	257
5.5.1	Auto Clustering	258
5.5.2	R-K-Means	261
5.5.3	SAP-HANA-Clustering-Algorithmen	264
5.5.4	Auto Classification	265
5.5.5	R-Bagging Classification	266
5.5.6	R-Boosting Classification	268
5.5.7	R-Random Forest Classification	269
5.5.8	SAP-HANA-Klassifikationsalgorithmen	270
5.6	Regressionsalgorithmen	273
5.6.1	Auto Regression	273
5.6.2	Exponentielle Regression	275
5.6.3	Geometrische Regression	276
5.6.4	Lineare Regression	277
5.6.5	Logarithmische Regression	278
5.6.6	R-exponentielle Regression	279
5.6.7	R-geometrische Regression	281
5.6.8	R-lineare Regression	281
5.6.9	R-logarithmische Regression	282
5.6.10	R-multilineare Regression	282
5.6.11	R-Random Forest Regression	284
5.6.12	SAP-HANA-Regressionsalgorithmen	285

5.7	Zeitreihen	287
5.7.1	R-Single Exponential Smoothing	287
5.7.2	R-Double Exponential Smoothing	289
5.7.3	R-Triple Exponential Smoothing	291
5.7.4	Triple Exponential Smoothing	292
5.7.5	SAP-HANA-Zeitreihenalgorithmen	292
5.8	Weitere Algorithmen	292
5.8.1	Entscheidungsstrukturen: R-CNR Tree	293
5.8.2	SAP-HANA-Entscheidungsstrukturen	297
5.8.3	Neuronale Netze	298
5.8.4	Ausreißeranalyse	303
5.8.5	SAP-HANA-Ausreißeralgorithmen	304

6 Integration von R im Modus »Expert Analytics« 307

6.1	Grundlagen von R	308
6.2	R-Integration	315
6.3	Beispiel: ABC-Analyse	317

7 Visualisierungen 325

7.1	Visualisierungen in Automated Analytics	325
7.2	Visualisierungen in Expert Analytics	326
7.2.1	Überblick der Darstellungsmöglichkeiten in Expert Analytics	326
7.2.2	Schlagwortwolke in Expert Analytics	329
7.2.3	Liniendiagramm in Expert Analytics	330
7.2.4	Säulendiagramm in Expert Analytics	330
7.2.5	Streudiagramm in Expert Analytics	331
7.2.6	Geoblasendiagramm	332
7.3	Visualisierungserweiterung durch VizPacker	333
7.3.1	Überblick zum VizPacker	333
7.3.2	Externe Visualisierungen einsetzen	334

8 Model Management mit der Predictive Factory 341

8.1	Einführung in die Predictive Factory	341
8.1.1	Einstellungen	342
8.1.2	Benutzer	344
8.1.3	Modellierungsserver	344
8.1.4	Externe Befehle	345
8.1.5	Projekte	346
8.1.6	Variablenstatistik	349
8.2	Deployment von Automated-Analytics-Modellen	350
8.2.1	Zeitreihen	351
8.2.2	Klassifikation	354
8.2.3	Clustering	355
8.3	Deployment von Expert-Analytics-Modellen	357
8.3.1	Schritt 1: Model Chain in Expert Analytics erstellen	357
8.3.2	Schritt 2: Model Chain aus Expert Analytics exportieren	357
8.3.3	Schritt 3: Model Chain in die Predictive Factory importieren	359

9 SAP-HANA-integriertes Data Mining 361

9.1	Einführung in SAP HANA Native	362
9.2	Application Function Library (AFL)	365
9.2.1	PAL-Algorithmen im Modus Expert Analytics einsetzen	366
9.2.2	PAL-Komponenten im Modus Expert Analytics erstellen	370
9.2.3	Automated Predictive Library (APL)	372
9.2.4	OFL und die SAP-HANA-Optimierungsfunktion	383
9.3	Weitere integrierte Szenarien	386
9.3.1	Modellexport im Modus Automated Analytics	386
9.3.2	Modellexport im Modus Expert Analytics	395

10 Integration von R in SAP HANA 407

10.1 Eigene Algorithmen für SAP HANA entwickeln	407
10.1.1 Serverarchitektur	408
10.1.2 RLANG-Prozedur	409
10.2 Beispiel: Netzwerkoptimierung	409
10.2.1 Mathematische Modellierung	410
10.2.2 Formulierung des Optimierungsproblems	411
10.2.3 Optimierung des Modells	411
10.2.4 Technische Implementierung	412
10.2.5 Ergebnis	418

11 Zusammenfassung und Ausblick 419

11.1 Zusammenfassung	419
11.2 Ausblick	420
Die Autoren	423
Index	427