

Inhaltsverzeichnis

Vorwort	9
1 Einleitung	15
1.1 Machine Learning – ein Überblick	15
1.1.1 Anwendungsszenarien	15
1.1.2 Funktionen von SAP HANA	17
1.2 Zentrale Begriffe und Methoden	20
1.2.1 Daten für Machine Learning	21
1.2.2 Beaufsichtigtes Lernen	24
1.2.3 Unbeaufsichtigtes Lernen	29
1.3 Beispieldaten mit Eclipse importieren	29
1.4 Python-Tools verwenden	38
2 Predictive Analysis Library mit SQLScript	47
2.1 Entscheidungsbäume für Klassifikationsprobleme	49
2.1.1 Trainieren eines Entscheidungsbaums	51
2.1.2 Testen des trainierten Modells	63
2.2 Clusteranalysen mit K-Means	70
3 Automated Predictive Library mit SQLScript	85
3.1 Ad-hoc-Analysen mit Predictive Business Functions	86
3.1.1 Statistiken und Datenprofil berechnen	87
3.1.2 Einflussfaktoren ermitteln	94
3.2 Predictive Model Functions	102
3.2.1 Modelle trainieren	102
3.2.2 Modelle testen und anwenden	123
3.2.3 Modelle bewerten	131
3.2.4 Weiterführende Informationen	140
4 Der Python Machine Learning Client für SAP HANA	143
4.1 Datenexploration und -vorbereitung mit HANA-Techniken	143

4.2	Datenvisualisierung im Jupyter Notebook	158
4.2.1	Visualisierungsfunktionen des Python ML Client	158
4.2.2	Verwendung von Pandas und Plotly	172
4.3	PAL in Python aufrufen	181
4.3.1	Klassifikationsmodelle	181
4.3.2	Clusteranalysen	207
4.3.3	Modelle speichern und wiederverwenden	219
4.4	APL in Python aufrufen	226
5	Text Mining	231
5.1	Import der Beispieldaten	233
5.2	Verwendung des Volltextindex	237
5.2.1	Erstellen des Volltextindex	237
5.2.2	Dokumentensuche und Schlagwörter	243
5.3	Themenextraktion	246
5.3.1	Normalisieren der Texte	248
5.3.2	Der Algorithmus Latent Dirichlet Allocation	251
5.3.3	Visualisierung der Themenscores in Python	259
5.4	Textklassifikation mit dem Python ML Client	267
6	Integration in SAP-Anwendungen	275
6.1	Vergleich der SAP-Produkte und -Techniken	275
6.1.1	Native SAP-HANA-Entwicklung	276
6.1.2	SAP S/4HANA	276
6.1.3	SAP BW/4HANA oder BW on HANA	277
6.1.4	SAP Predictive Analytics	278
6.1.5	SAP Data Intelligence	278
6.1.6	SAP Analytics Cloud	279
6.1.7	SAP AI Business Services	280
6.1.8	Python, SQLScript oder ABAP	280
6.2	PAL-Prognosemodelle in Calculation Views	281
7	Aufbau einer Testumgebung	289
7.1	Einrichten der HANA-Instanz	289
7.2	Einrichten der Python-Umgebung	293

8	Fazit und Ausblick	297
A	Der Autor	304
B	Quellcodeverzeichnis	305
C	Literaturverzeichnis	307
D	Index	311
E	Disclaimer	316