Bernd Heesen

Künstliche Intelligenz im Business

Erstellung eigener Anwendungen mit Python

Springer Gabler

Inhaltsverzeichnis

1 Nutzen von Machine Learning und KI 1
1.1 Künstliche Intelligenz 3
1.2 Menschliche Intelligenz 4
1.3 Rationale Entscheidungen 5
1.4 Entscheidungsautomation 7
1.5 Entscheidungsprozess bei Mensch und Maschine 9
1.6 Geschwindigkeit der Entscheidungsfindung 10
1.7 Datenbasis für Entscheidungen 12
1.8 Business Intelligence 14
1.9 Innovation 15
Literatur 16

2 Machine Learning 19
2.1 Überwachtes Lernen 22
2.2 Unüberwachtes Lernen 22
2.3 Bestärkendes Lernen 23
Literatur 24

3 Best Practices 25
3.1 Datenqualität 27
3.2 Visualisierung 30
3.3 Entwicklungsumgebung 31
3.4 Programmierung 32
3.5 Künstliche Intelligenz 33
Literatur 35

4 Entwicklungsumgebung 37
4.1 Anaconda Navigator 40
4.2 Jupyter Notebook Editor 41
4.3 Spyder Editor 45
4.4 R 58

4.5 Rtools . . . 63
4.6 RStudio . . . 67
4.7 Pakete in R installieren . . . 74
4.8 Pakete in Python installieren . . . 77
4.9 Python in Anaconda Prompt ausführen . . . 80
4.10 Python als Executable in Windows Prompt ausführen . . . 82
4.11 Python in RStudio ausführen . . . 83
4.12 Paket pythonforbusiness . . . 85
Literatur . . . 87

5 Python-Grundlagen . . . 89
5.1 Elementare Datentypen . . . 91
5.2 Datenstrukturen . . . 96
5.2.1 Liste und Tupel . . . 96
5.2.2 Set . . . 101
5.2.3 Dictionary . . . 102
5.2.4 Array . . . 104
5.2.5 Dataframe und Serie . . . 107
5.3 Operationen . . . 114
5.3.1 Arithmetische Operationen . . . 114
5.3.2 Logische Operationen . . . 115
5.3.3 Statistische Operationen . . . 117
5.4 Kontrollstrukturen . . . 121
5.5 Funktionen, Methoden und Klassen . . . 125
5.5.1 Funktionen und Methoden aufrufen . . . 125
5.5.2 Eigene Funktionen . . . 127
5.5.3 Eigene Klassen . . . 131
5.6 Standardfunktionen . . . 132
5.6.1 Formatierte Ausgabe von Zeichenketten . . . 132
5.6.2 Zufallszahlen . . . 133
5.6.3 Performancemessung . . . 136
5.7 Eigenes Modul/Paket . . . 136
Literatur . . . 137

6 Erstellung von KI-Anwendungen . . . 139
6.1 Import und Export von Daten . . . 140
6.2 Datenvorverarbeitung . . . 152
6.2.1 Fehlende Werte . . . 152
6.2.2 Doppelte Datensätze . . . 154
6.2.3 Ausreißer . . . 156
6.2.4 Datenintegration . . . 158
6.2.5 Encoding . . . 161

6.2.6 Normalverteilung . . . 164
6.2.7 Selektion relevanter Merkmale . . . 172
6.2.8 Skalierung . . . 179
6.3 Explorative Datenanalyse . . . 180
6.3.1 Analysemethoden für Basistabellen . . . 182
6.3.2 Statistische Analysemethoden . . . 194
6.3.3 Datenvisualisierung . . . 214
6.4 Datenbereitstellung . . . 242
6.4.1 Stichproben . . . 243
6.4.2 Split in Trainings- und Testdaten . . . 249
6.4.3 Resampling . . . 254
6.5 Modellierung und Betrieb . . . 258
6.5.1 Modellierung . . . 259
6.5.2 Modell-Wahl . . . 263
6.5.3 Modell-Evaluation . . . 290
6.5.4 Modell-Training . . . 292
6.5.5 Modell-Tuning . . . 302
Literatur . . . 307

7 Anwendung von Machine Learning . . . 309
7.1 Überwachtes Lernen . . . 310
7.1.1 Regression zur Preisschätzung . . . 310
7.1.2 Klassifikation zur Kreditanalyse . . . 321
7.2 Unüberwachtes Lernen . . . 329
7.2.1 Clustering zur Kundensegmentierung . . . 329
7.2.2 Empfehlungsmaschine für Cross-Selling . . . 336
7.3 Bestärkendes Lernen für Dynamic Pricing . . . 353
Literatur . . . 360

8 Ausblick . . . 361
Literatur . . . 366