

Inhalt

Vorwort — V

1	Actuarial Data Science – Business Cases — 1
2	Crashkurs in Data Mining Anwendungen — 5
3	Neue Versicherungsprodukte — 12
3.1	Innovative Produkte — 12
3.2	Produktentwicklung — 14
3.3	Kompetenz des Data Scientist — 15
3.4	Organisationsstruktur — 16
4	Tools, Sprachen, Frameworks — 19
4.1	WEKA und KNIME — 19
4.1.1	WEKA — 20
4.1.2	WEKA-Java-Library — 22
4.1.3	KNIME — 24
4.2	Python, R und Jupyter Notebooks — 25
4.2.1	Reproducible Research — 27
4.2.2	Weitere Aktuarielle Anwendungen — 28
5	Informationstechnologie — 35
5.1	Arbeiten mit Dateien — 35
5.1.1	Datenspeichermedien — 36
5.1.2	Datenformate — 36
5.1.3	Tabellarische Daten — 39
5.1.4	Excel - Nicht-tabellarische Daten — 41
5.1.5	Textdateien — 42
5.1.6	Datenverschlüsselung — 49
5.2	Arbeiten mit Daten aus dem Netz — 51
5.2.1	Objekt-Darstellungen in JSON — 52
5.2.2	Datenabfragen von Web-APIs — 55
5.2.3	Daten aus Webseiten — 57
5.2.4	Daten in Amazons S3-Object Storage — 58
5.3	Arbeiten mit Relationalen Datenbanken — 59
5.3.1	Der Nutzen von (relationalen) Datenbanken — 60
5.3.2	Objekte und Objekttypen in relationalen Datenbanken — 63
5.3.3	Datenbank-Modelle: Relationale Datenbanken — 68
5.3.4	SQL — 72

5.3.5	Umsetzung in einer relationalen Datenbank —	74
5.3.6	Fortgeschrittene Themen zu Relationalen Datenbanken —	80
5.4	Arbeiten mit No-SQL-Datenbanken —	87
5.4.1	Verteilte Datenbanksysteme —	88
5.4.2	CAP Theorem für verteilte Systeme —	89
5.4.3	Einsatz von No-SQL —	91
5.5	Datenwarenhäuser —	107
5.5.1	Operative und dispositive Datenhaltung —	107
5.5.2	Data Warehouses —	108
5.5.3	Sternschema —	110
5.5.4	Data Lakes —	111
5.6	Softwaretests —	117
5.7	Parallele Datenverarbeitung —	122
5.8	MapReduce, Hadoop und Spark —	132
5.9	Cloud Computing —	144
5.9.1	Begriffe und Konzepte —	144
5.9.2	Data Science Angebote in der Cloud —	146
5.9.3	Deployment von Machine Learning Algorithmen —	146
5.10	Informationsverarbeitung in Versicherungsunternehmen —	154
5.10.1	Geschäftsprozesse in Versicherungsunternehmen —	154
5.10.2	Typische Systemlandschaft —	157
6	Mathematische Verfahren —	160
6.1	Datenaufbereitung —	160
6.1.1	Schritte der Datenaufbereitung —	160
6.1.2	Fehlende Daten —	168
6.2	Datenverständnis und -visualisierung —	176
6.3	Klassifikations- und Regressionsmethoden —	184
6.3.1	Multiple Lineare Regression —	185
6.3.2	Solvenzkapital für Biometrische Risiken —	194
6.3.3	Binäre Regression —	200
6.3.4	Generalisierte Additive Modelle (GAM) —	205
6.3.5	Lineare Diskriminanzanalyse (LDA) —	212
6.3.6	k-Nearest-Neighbors (kNN) —	218
6.3.7	Naive Bayes —	222
6.3.8	Entscheidungsbäume —	223
6.3.9	Random Forests —	230
6.3.10	Boosting —	232
6.3.11	Support Vector Machines —	236
6.3.12	Künstliche Neuronale Netze —	240
6.4	Clustermethoden —	251
6.4.1	Clusterbasierte Bestandsverdichtung (k-Means) —	252

6.4.2	Hierarchische Clustermethoden — 258
6.4.3	Bestimmung optimaler Anzahl von Clustern — 266
6.5	Dimensionsreduktion — 271
6.5.1	Lineare Diskriminanzanalyse (LDA) — 272
6.5.2	Hauptkomponentenanalyse (PCA) — 273
6.6	Bewertung von Modellen — 278
6.6.1	Trainings-Sampling — 280
6.6.2	Kreuzvalidierung — 282
6.6.3	Hyperparameter-Tuning — 283
6.6.4	Umgang mit Sampling-Disbalancen — 284
6.6.5	Generische skalare Fehlermaße — 287
6.6.6	Konfusionsmatrix — 288
6.6.7	ROC-Kurven — 290
6.6.8	Marginalverteilungsanalyse — 292
6.6.9	Liftplot — 293
6.6.10	Lorenzkurve und Gini-Koeffizient — 294
6.6.11	Regularisierung — 297
7	Korrelation und kausale Inferenz — 305
8	Data Mining — 320
8.1	Verbreitete Prozessmodelle — 320
8.2	Geschäftsverständnis — 322
8.3	Datenverständnis — 323
8.4	Datenvorbereitung — 325
8.5	Modellierung — 327
8.6	Evaluierung — 328
8.7	Bereitstellung — 331
9	Gesellschaftliches Umfeld — 332
9.1	Künstliche Intelligenz — 332
9.2	Datenschutz — 336
9.3	Ethische Fragen — 342
A	Appendix — 346
A.1	Installationen — 346
A.1.1	Python — 346
A.1.2	R — 347
A.2	Datensätze — 348
A.2.1	AutoBi — 348
A.2.2	Simulierte Schadenfrequenzen in der Kfz-Versicherung — 349
A.2.3	Simulierte Verkaufsdaten — 350

A.2.4	Simulierte Verkaufsempfehlungsdaten —	352
A.2.5	Kfz-Betrugserkennung —	353
A.2.6	Kfz-Telematik —	353
A.3	Marginalverteilungsplots —	356
A.4	Liftplots —	358

Nachwort & Danksagungen —	361
--------------------------------------	------------

Literatur —	363
--------------------	------------

Stichwortverzeichnis —	369
-------------------------------	------------