

Inhaltsverzeichnis

Vorwort	v
Danksagung	vii
Inhaltsverzeichnis	ix
Autoren	xiv
Notation	xvi

Teil I Einführung, Konzept und Grundlagen **1**

1 Einführung

<i>Claus Weihs</i>	3
1.1 Motivation	3
1.2 Daten und Studientypen	4
1.3 Modelle und ihre Beurteilung	6
1.4 Inhalt	7
Literaturverzeichnis	8

2 CRISP-DM - Ein Konzept für die journalistische Datenanalyse?

<i>Anna Behrend</i>	9
2.1 Motivation und Methodik	9
2.2 Begriffsbildung	10
2.2.1 Datenjournalismus	10
2.2.2 Data Mining und KDD	11
2.2.3 Qualität im Datenjournalismus	12
2.3 Beschreibung der Arbeitsabläufe	13
2.3.1 Data-Mining-Prozess CRISP-DM	13
2.3.2 Typische datenjournalistische Arbeitsschritte	16
2.4 Wie ähnlich sind datenjournalistischer Arbeitsprozess und CRISP-DM?	18
2.4.1 Interviews als Informationsgrundlage	19
2.4.2 Die Ergebnisse	20

2.5	Kann CRISP-DM eine sinnvolle Leitlinie für den Datenjournalismus sein?	24
2.6	Zusammenfassung und Fazit	26
2.7	Diskussion	27
	Literaturverzeichnis	27
3	Data Literacy	
	<i>Katja Ickstadt, Henrik Müller, Henrike Weinert</i>	29
3.1	Einleitung	29
3.2	Was ist Data Literacy?	31
3.3	Wertschöpfung aus Daten	33
3.4	Daten und Unsicherheit	34
3.5	Operationalisierung: das DaCoNet-Konzept	38
3.6	Fazit: Data Literacy ist ein Prozess!	39
3.7	Diskussion	40
	Literaturverzeichnis	40
4	Datengrafiken zwischen Nutzwert und Design	
	<i>Christina Elmer</i>	43
4.1	Wozu wir Grafiken benötigen	43
4.2	Grafiken nutzerzentriert denken und entwickeln	44
4.3	Inhaltliche Qualität	46
4.4	Hochwertige Gestaltung	50
4.5	Angemessener Kontext	52
4.6	Diskussion	54
	Literaturverzeichnis	54
5	Algorithmen im Fokus	
	<i>Christina Elmer</i>	56
5.1	Einführung	56
5.2	Recherchen zu Algorithmen	57
5.3	Wie Algorithmen arbeiten und dazulernen	58
5.4	Die zentralen Fragen an Algorithmen	59
5.5	Diskussion	62
	Literaturverzeichnis	63
	Teil II Fallstudien	65
6	(Bedingte) Verteilung und statistische Tests	
	<i>Claus Weihs, Marcel Pauly</i>	67
6.1	Fallstudie 1: Altersstruktur von Parlamenten	67
6.2	Re-Analyse	68
6.3	Zusammenfassung	76
6.4	Diskussion	77
	Literaturverzeichnis	77

7 Zusammenhangsanalyse: Klassifikation

<i>Claus Weihs, Patrick Stotz</i>	78
7.1 Fallstudie 2: Abschneiden der AfD bei der Bundestagswahl 2017	78
7.2 Re-Analyse	79
7.3 Zusammenfassung	87
7.4 Diskussion	88
Literaturverzeichnis	88

8 Zusammenhangsanalyse: Regression

<i>Ana Moya, Marie-Louise Timcke, Claus Weihs</i>	89
8.1 Fallstudie 3: Wählerstruktur bei der Bundestagswahl 2017	89
8.2 Die Analyse von Zeit Online	90
8.3 Re-Analyse	96
8.4 Vergleich und Diskussion	101
8.5 Zusammenfassung	102
8.6 Diskussion	102

9 Zeitreihenanalyse: Modellentwicklung über die Zeit

<i>Claus Weihs</i>	104
9.1 Fallstudie 4: Corona (COVID-19) Pandemie	104
9.2 Generelles Vorgehen	105
9.3 Gesamtanzahl Infizierter: Obergrenze und Stagnation	109
9.4 Anzahl Neuinfizierte	115
9.5 Reproduktionszahl	121
9.6 Zusammenfassung und Schlussfolgerung	126
9.7 Nachher weiß man immer alles besser?	127
9.8 Kommentar von Marie-Louise Timcke: Das Corona-Datenchaos: „Netflix & Bug fixing“	128
Literaturverzeichnis	131

10 Gruppenbildung: Clusteranalyse

<i>Claus Weihs</i>	132
10.1 Fallstudie 5: Buchbestsellerlisten	132
10.2 Generelles Vorgehen	133
10.3 Analyse von Hardcover Belletristik	136
10.3.1 Längste und beste Karrieren	136
10.3.2 Glättung	137
10.3.3 Clusterbildung	138
10.3.4 Interpretation der Cluster	138
10.3.5 Einordnung von Buchtiteln in Cluster	142
10.4 Analyse von Hardcover Sachbuch	144
10.4.1 Längste und beste Karrieren	144
10.4.2 Glättung	145
10.4.3 Clusterbildung	146
10.4.4 Interpretation der Cluster	146
10.4.5 Einordnung von Buchtiteln in Cluster	149
10.5 Vergleich von Belletristik und Sachbuch	150

10.6 Diskussion	152
11 Sequentielle Daten: Analyse von Radverkehrsnetzen	
<i>Claus Weihs, Lilia Michailov</i>	155
11.1 Fallstudie 6: Radwege in Berlin	155
11.2 GPS-Koordinaten	156
11.3 Nutzung des Radnetzes	161
11.4 Identifikation der Lücken im Radnetz	164
11.5 Zusammenfassung	167
11.6 Kommentar von Hendrik Lehmann	167
Literaturverzeichnis	169
12 Datenerhebung: Verwendung von Vorwissen	
<i>Claus Weihs, Tanja Hernández Rodríguez</i>	170
12.1 Datenerhebung	170
12.2 Verwendung von Vorwissen	171
12.3 Fallstudie 7: Professorenfrage	172
12.4 Die E-Mail-Umfrage	172
12.4.1 Gesamtarbeitszeit	173
12.4.2 Summe von Teilarbeitszeiten	173
12.4.3 Ausschluss-Kriterien	175
12.5 Statistische Methoden	177
12.5.1 Prognoseintervalle frequentistisch	177
12.5.2 Relevantes Vorwissen	178
12.5.3 A-priori Verteilungen	180
12.5.4 A-posteriori Verteilungen	181
12.5.5 Prognoseintervalle Bayesianisch	181
12.5.6 Empirische Bayes-Methode am Beispiel	182
12.6 Ergebnisse	184
12.6.1 Gesamtschätzung der wöchentlichen Arbeitszeit	184
12.6.2 Summierter wöchentlicher Zeitaufwand	185
12.7 Weitere Ergebnisse	187
12.8 Zusammenfassung	187
12.9 Kommentar von Holger Wormer	188
Literaturverzeichnis	189
13 Geplante Studien	
<i>Claus Weihs, Gerret von Nordheim</i>	190
13.1 Motivation und Zielsetzung	190
13.2 Geplante Studien in der Wirtschaftsjournalistik: Wirkung unterschiedlicher Medien	190
13.3 Statistische Versuchsplanung	192
13.4 Fallstudie 8: Art der Präsentation	195
13.5 Diskussion	199
Literaturverzeichnis	200

Teil III Qualitätsstandards

201

14 Datenethik im Journalismus

<i>Detlef Steuer, Ursula Garczarek</i>	203
14.1 Berührungspunkte von Datenethik, Datenwissenschaften und Journalismus	203
14.2 Die Sonderrolle von Zahlen und Fakten	205
14.3 Relevante Ziffern des Pressekodex	206
14.4 Fazit	214
14.5 Diskussion	214
Literaturverzeichnis	215

15 Qualitätsstandards: Checklisten als Hilfsmittel

<i>Holger Wormer</i>	217
15.1 Einleitung	217
15.2 Arbeitstechniken der strukturierten journalistischen Recherche ...	218
15.3 Experten-Checkliste	220
15.4 Studien-Checkliste	224
15.5 Fazit	229
15.6 Diskussion	230
Literaturverzeichnis	232

Anhang: Daten und R-Programme

A.6 Kapitel 6: Fallstudie 1: Altersstruktur von Parlamenten	236
A.6.1 Daten	236
A.6.2 R-Programm	238
A.7 Kapitel 7: Fallstudie 2: AfD bei der Bundestagswahl 2017	244
A.7.1 Daten	244
A.7.2 R-Programm	249
A.8 Kapitel 8: Fallstudie 3: Wählerstruktur: Bundestagswahl 2017 ...	251
A.8.1 Daten	251
A.8.2 R-Programm	252
A.9 Kapitel 9: Fallstudie 4: Corona (COVID-19) Pandemie	253
A.9.1 Daten	253
A.9.2 R-Programm	254
A.10 Kapitel 10: Fallstudie 5: Buchbestsellerlisten	260
A.10.1 Daten	260
A.10.2 R-Programm	278
A.11 Kapitel 11: Fallstudie 6: Radwege in Berlin	283
A.11.1 Daten	283
A.11.2 R-Programm	300
A.12 Kapitel 12: Fallstudie 7: Professorenenumfrage	309
A.12.1 Daten	309
A.12.2 R-Programm	311

Sachverzeichnis

317