

Inhaltsverzeichnis

1. Problemüberblick und Zielrichtung.....	1
1.1 Begriffsdefinitionen.....	1
1.2 Problemstellung bei WEKA MEDIA.....	9
1.3 Zielrichtung und Aufbau der Arbeit.....	11
2. Informationstechnische und methodische Grundlagen.....	15
2.1 Informationstechnische Grundlagen.....	15
2.1.1 Das Datawarehouse Konzept.....	15
2.1.2 Knowledge Discovery in Databases.....	18
2.2 Methodische Grundlagen.....	23
2.2.3 Traditionelle Kundenbewertungsverfahren.....	23
2.2.2 Data Mining.....	30
2.2.3 OLAP.....	37
2.3 Zusammenhang der informationstechnischen und methodischen Grundlagen ...	38
3. Zweckmäßige Voranalysen.....	41
3.1 Datenerfassung.....	41
3.2 Datenvorverarbeitung.....	42
3.2.1 Variablenmodifikation.....	42
3.2.2 Fehlende Werte.....	47
3.2.3 Variablenreduktion.....	48
3.2.4 Ausreißer-Analyse.....	52
3.2.5 Aufteilung der Datenmatrix in Trainings-, Validierungs- und Testdaten..	53
3.3 Bewältigung niedriger Responsequoten.....	56
3.3.1 Stichprobenplanung.....	60
3.3.2 Clusteranalytische Verfahren zur Unterstützung der Stichprobenziehung	61
3.4 Zusammenfassung.....	74
4. Responseoptimierung mit Entscheidungsbaumverfahren.....	77
4.1 Entscheidungsbaumvarianten.....	77
4.2 Partitionierungskriterien.....	81
4.2.1 Gini-Index.....	83
4.2.2 Informationsgewinn.....	85
4.2.3 θ^2 -Unabhängigkeitstest.....	88
4.3 Pruning-Methoden.....	89
4.4 Spezielle Verfahren.....	95

4.5 Probleme bei Entscheidungsbaumverfahren.....	101
4.6 Empirische Ergebnisse.....	103
4.7 Zusammenfassung	113
5. Responseoptimierung mit der binären logistischen Regression	115
5.1 Das Logit-Modell.....	115
5.2 Parameterschätzung, Tests auf Signifikanz und Aufnahme von Variablen.....	118
5.3 Goodness-of-fit - Tests	123
5.4 Empirische Ergebnisse.....	124
5.5 Zusammenfassung	132
6. Responseoptimierung mit Künstlichen Neuronalen Netzen.....	135
6.1 Varianten und Architektur von KNN.....	136
6.2 Optimale Netzwerkstruktur.....	144
6.3 Empirische Ergebnisse.....	146
6.4 Zusammenfassung	153
7. Vergleich der verwendeten Variablen verschiedener Modellvarianten	155
7.1 Bildung einer Distanzmatrix	155
7.2 Beschreibung der Multidimensionalen Skalierung.....	157
7.3 Empirische Ergebnisse.....	159
7.4 Gesamtinterpretation der empirischen Ergebnisse in Verbindung mit der Repräsentation	166
7.5 Auswirkungen einer Reduzierung der Anzahl unabhängiger Variablen	168
7.6 Zusammenfassung	177
8. Zusammenfassung und Ausblick.....	179
Literaturverzeichnis	183
Anhang	205
A Umkodierungen.....	207
B Korrelation mit der Zielvariablen.....	209
C Korrelationen der verbleibenden Variablen untereinander	210
D Ausreißeranalyse.....	213
E CCC-Plots.....	214
F Datenmatrix für MDS.....	216
G Datenmatrix bei 8 Variablen für logistische Regression	218
H Datenmatrix bei 8 Variablen für Entscheidungsbäume	219
I Datenmatrix bei 3 Variablen für logistische Regression.....	220

J Datenmatrix bei 4 Variablen für Entscheidungsbäume.....	221
K Informationen zum SAS Institute	222