

DISKUSSIONSBEITRÄGE DES FACHBEREICHS  
WIRTSCHAFTSWISSENSCHAFT  
DER GERHARD-MERCATOR-UNIVERSITÄT  
GESAMTHOCHSCHULE DUISBURG

Nr. 232

Methoden und Verfahren des Data Mining

P. Chamoni, C. Budde

Gerhard-Mercator-Universität  
Gesamthochschule Duisburg  
Fachbereich Wirtschaftswissenschaft  
Wirtschaftsinformatik und Operations Research  
Lotharstrasse 65  
D-47048 Duisburg  
Telefon: (+49)203-379-2627  
Telefax: (+49)203-379-4347  
Email: WI@uni-duisburg.de

Januar 1997

DEPARTMENT OF ECONOMICS DISCUSSION PAPERS  
UNIVERSITY OF DUISBURG, GERMANY

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>4</b>
<b>2</b>	<b>Data Warehouse und Data Mining</b>	<b>6</b>
2.1	Data Warehouse . . . . .	6
2.2	Potentiale eines Data Warehouse . . . . .	7
2.3	Multidimensionale Datenanalyse . . . . .	8
2.4	Data Mining . . . . .	9
2.4.1	Data Mining-Anwendungen . . . . .	9
2.4.2	Data-Mining-Verfahren . . . . .	10
<b>3</b>	<b>Darstellung von Objekten</b>	<b>14</b>
3.1	Merkmalstypen . . . . .	15
3.2	Normierung und Standardisierung . . . . .	15
3.2.1	Normierung . . . . .	16
3.2.2	Standardisierung . . . . .	17
3.3	Auswahl von Merkmalen . . . . .	17
<b>4</b>	<b>Beispiele</b>	<b>19</b>
4.1	Unüberwachtes Lernen . . . . .	19
4.2	Überwachtes Lernen . . . . .	20
<b>5</b>	<b>Data Mining durch Clusteranalyse</b>	<b>21</b>
5.1	Hierarchische Clusterung . . . . .	21
5.1.1	Voraussetzungen . . . . .	21
5.1.2	Agglomerative Methoden . . . . .	23
5.1.3	Divisive Methoden . . . . .	24
5.1.4	Eigenschaften Hierarchischer Methoden . . . . .	25
5.1.5	Beispiel . . . . .	25
5.2	Partitionierende/Nichthierarchische Clusterung . . . . .	29
5.2.1	Der $K$ -Means-Algorithmus . . . . .	29
5.2.2	Der FKM-Algorithmus . . . . .	31
5.2.3	Eigenschaften Partitionierender Methoden . . . . .	33
5.2.4	Gütekriterien . . . . .	33
5.2.5	Beispiel . . . . .	34
<b>6</b>	<b>Data Mining durch Statistik</b>	<b>37</b>
6.1	Bayes-Klassifikation . . . . .	37
6.2	Bestimmung der bedingten Wahrscheinlichkeiten . . . . .	38
6.2.1	Normalverteilung und Maximum-Likelihood-Schätzung . . . . .	38
6.2.2	Supervised Bayes-Learning . . . . .	39
6.2.3	Approximation der Dichtefunktion . . . . .	39
6.2.4	Stochastische Approximation . . . . .	40

6.3	Eigenschaften und Probleme . . . . .	41
6.4	Beispiel . . . . .	42
<b>7</b>	<b>Data Mining durch Induktives Lernen</b>	<b>46</b>
7.1	Konzeptionelles Clustern . . . . .	46
7.1.1	Erklärungen und Begriffe . . . . .	46
7.1.2	Cluster/2-Algorithmus (Partitionierendes Modul) . . . . .	48
7.1.3	Eigenschaften und Alternativen . . . . .	49
7.1.4	Beispiel . . . . .	51
7.2	Entscheidungsbäume und Entscheidungsregeln . . . . .	54
7.2.1	ID3-Algorithmus . . . . .	54
7.2.2	Eigenschaften . . . . .	57
7.2.3	Erweiterung zum C4.5-Algorithmus . . . . .	58
7.2.4	Beispiel . . . . .	60
<b>8</b>	<b>Data Mining durch Neuronale Netze</b>	<b>64</b>
8.1	Multilayer Perceptron . . . . .	64
8.1.1	Modell . . . . .	64
8.1.2	Verallgemeinerte Delta-Regel . . . . .	66
8.1.3	Algorithmus . . . . .	67
8.1.4	Eigenschaften, Probleme und Ergänzungen . . . . .	69
8.1.5	Beispiel . . . . .	70
8.2	Kohonen-Algorithmus . . . . .	73
8.2.1	Modell . . . . .	73
8.2.2	Algorithmus . . . . .	75
8.2.3	Fuzzy-Kohonen-Clusterung . . . . .	76
8.2.4	Eigenschaften und Probleme . . . . .	76
8.2.5	Beispiel . . . . .	77
<b>9</b>	<b>Anwendungen der Datenmustererkennung</b>	<b>82</b>
9.1	Datenmustererkennung in der Betriebsergebnisrechnung . . . . .	82
9.1.1	Problem . . . . .	82
9.1.2	Durchführung der Mustererkennung . . . . .	82
9.1.3	Ergebnisse . . . . .	82
9.2	Analyse von Kassenbons und Kundenkarten . . . . .	83
9.2.1	Problem . . . . .	83
9.2.2	Durchführung . . . . .	83
9.2.3	Ergebnisse . . . . .	83
9.3	Fuzzy Clusteranalyse zur Kundensegmentierung in Banken . . . . .	83
9.3.1	Problem . . . . .	83
9.3.2	Durchführung der Segmentierung . . . . .	84
9.3.3	Ergebnisse . . . . .	85
9.3.4	Weitere Anwendungen . . . . .	85
9.4	Eine Kombination von unüberwachtem und überwachtem Lernen . . . . .	86

9.4.1	Durchführung . . . . .	86
9.4.2	Ergebnisse . . . . .	87
9.5	Prognose von Aktienindizes . . . . .	87
9.5.1	Problem . . . . .	87
9.5.2	Durchführung . . . . .	88
9.5.3	Ergebnis . . . . .	88
9.5.4	Weitere Anwendungen . . . . .	88
9.6	Bonitätseinstufung von Anleihen . . . . .	88
9.6.1	Problem . . . . .	88
9.6.2	Durchführung . . . . .	89

**10 Zusammenfassung und Ausblick****92**